BIOFUNDAMENTALS 2E



Klymkowsky & Cooper
University of Colorado Boulder &
Michigan State University

LibreTexts™

University of Colorado Boulder & Michigan State University

Biofundamentals

Klymkowsky & Cooper

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 05/01/2026

TABLE OF CONTENTS

Licensing

1: Understanding Science and Thinking Scientifically

- 1.0: Introduction
- 1.1: The interconnectedness (self-consistency) of science
- 1.2: Models, Hypotheses, and Theories
- 1.3: Science is Social
- 1.4: Teaching and Learning Science
- 1.5: Understanding Scientific Ideas

2: Life’s Diversity and Origins

- 2.0: Introduction
- 2.1: What is life, exactly?
- 2.2: The cell theory and the continuity of life
- 2.3: The organization of organisms
- 2.4: Spontaneous generation and the origin of life
- 2.5: The Death of Vitalism
- 2.6: Thinking about life’s origins
- 2.7: Experimental studies on the origins of life
- 2.8: Mapping the history of life on earth
- 2.9: Fossil evidence for the history of life on earth
- 2.10: Life's impact on earth

3: Evolutionary Mechanisms and the Diversity of Life

- 3.0: Introduction
- 3.1: Organizing organisms (hierarchically)
- 3.2: Natural and un-natural groups
- 3.3: Fossils and family relationships- introducing cladistics
- 3.4: Evolution theory’s core concepts
- 3.5: Limits on populations
- 3.6: Darwin and Wallace's conceptual leap
- 3.7: Mutations and the origins of genotype-based variation
- 3.8: The origins of polymorphisms
- 3.9: Genotype-phenotype relationships- discrete and continuous traits
- 3.10: Variation, selection, and speciation
- 3.11: Types of simple selection
- 3.12: Considering stochastic processes
- 3.13: Genetic diversity, population size, founder effects and population bottlenecks
- 3.14: Population bottlenecks
- 3.15: Genetic drift
- 3.16: Gene linkage- one more complication
- 3.17: A brief reflection on the complexity of phenotypic traits
- 3.18: Speciation and extinction
- 3.19: Mechanisms of speciation
- 3.20: Mechanisms of reproductive isolation

- 3.21: Sympatric speciation
- 3.22: Signs of evolution- homology and convergence
- 3.23: Homologies provide evidence for a common ancestor
- 3.24: Anti-evolution arguments

4: Social Evolution and Sexual Selection

- 4.0: Introduction
- 4.1: Selecting social (cooperative) traits
- 4.2: Quorum Sensing
- 4.3: Active (altruistic) cell death
- 4.4: Inclusive fitness, kin and group selection, and social evolution
- 4.5: Group selection
- 4.6: Defense against social cheaters
- 4.7: The appearance of multicellular organisms
- 4.8: Origins and implications of sexual reproduction
- 4.9: Sexual dimorphism
- 4.10: Sexual Selection
- 4.11: Curbing runaway selection

5: Molecular Interactions, Thermodynamics and Reaction Coupling

- 5.0: Introduction
- 5.1: Approaching Thermodynamics
- 5.2: Coupling thermodynamically favorable and unfavorable reactions
- 5.3: Reaction rates
- 5.4: Coupling Reactions
- 5.5: Inter- and intra-molecular interactions
- 5.6: Covalent bonds
- 5.7: Bond Stability and Thermal Motion (a non-biological moment)
- 5.8: Bond polarity, inter- and intramolecular interactions
- 5.9: The implications of bond polarity
- 5.10: Interacting with Water

6: Membrane Boundaries and Capturing Energy

- 6.0: Introduction
- 6.1: Defining the Cell's Boundary
- 6.2: The origin of biological membranes
- 6.3: Transport across membranes
- 6.4: Channels and carriers
- 6.5: Generating gradients- using coupled reactions and pumps
- 6.6: Simple Phototrophs
- 6.7: Chemo-osmosis (an overview)
- 6.8: Oxygenic Photosynthesis
- 6.9: Chemotrophs
- 6.10: Using the energy stored in membrane gradients
- 6.11: Osmosis and living with and without a cell wall
- 6.12: An evolutionary scenario for the origin of eukaryotic cells
- 6.13: Making a complete eukaryote

7: The Molecular Nature of Heredity

- 7.0: Introduction
- 7.1: Discovering how nucleic acids store genetic information
- 7.2: Locating hereditary material within the cell
- 7.3: Identifying DNA as the genetic material
- 7.4: Unraveling Nucleic Acid Structure
- 7.5: Discovering the structure of DNA
- 7.6: DNA, sequences, and information
- 7.7: Discovering RNA- structure and some functions
- 7.8: DNA Replication
- 7.9: Replication machines
- 7.10: Further replication complexities
- 7.11: Mutations, Deletions, Duplications and Repair
- 7.12: Genes and alleles
- 7.13: The origin of new (de novo) genes
- 7.14: DNA repeat diseases and genetic anticipation

8: Peptide Bonds, Polypeptides and Proteins

- 8.0: Introduction
- 8.1: Specifying a polypeptide's sequence
- 8.2: The origin of the genetic code
- 8.3: Protein synthesis- transcription (DNA to RNA)
- 8.4: Protein synthesis- translation (RNA to polypeptide)
- 8.5: The translation (polypeptide synthesis) cycle
- 8.6: Effects of point mutations on polypeptides and proteins
- 8.7: Getting more complex- gene regulation in eukaryotes
- 8.8: Turning polypeptides into proteins
- 8.9: Regulating protein activity, concentrations, and stability (half-life)
- 8.10: Allosteric and post-translational regulation
- 8.11: Diseases of folding and misfolding
- 8.12: Molecular machines

9: Genomes, Genes and Regulatory networks

- 9.0: Introduction
- 9.1: Locating information within DNA
- 9.2: Enhancers and transcription start and stop sites
- 9.3: Interaction networks and model systems
- 9.4: E. coli as a model system
- 9.5: Adaptive behavior and gene networks- the lac response
- 9.6: Types of regulatory interactions
- 9.7: Final thoughts on (molecular) noise

10: Social Systems

- 10.0: Introduction
- 10.1: Targeting membrane proteins to where they need to be
- 10.2: Nuclear targeting and exclusion
- 10.3: Intercellular signaling- signals, receptors and responses
- 10.4: Signaling molecules and receptors

- 10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells

11: Cellular reproduction and horizontal gene transfer in prokaryotes

- 11.0: Introduction
- 11.1: What counts as sex in prokaryotes
- 11.2: Other naturally occurring horizontal gene transfer mechanisms

12: Asexual and Sexual Reproduction in Eukaryotes

- 12.0: Introduction
- 12.1: Ploidy during the cell cycle
- 12.2: Monitoring cellular processes- mitosis
- 12.3: Sex-determination and its chromosomal basis
- 12.4: Steps in meiosis- from diploid to haploid
- 12.5: Recombination and independent segregation
- 12.6: Linkage and haplotypes
- 12.7: X-inactivation and sex-linked traits
- 12.8: X-linked diseases and mono-allelic gene expression

13: How Alleles Arise - Mutations

- 13.0: Introduction
- 13.1: Mutations into alleles
- 13.2: Luria and Delbrück- Discovering the origin of mutations
- 13.3: Forward and reverse genetics
- 13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
- 13.5: Longer term mutation and evolution studies

14: Somatic Mutations and Genome Dynamics

- 14.0: Introduction
- 14.1: Rates and effects of somatic mutation
- 14.2: Non-disjunction- aberrant chromosome segregation
- 14.3: Genome dynamics
- 14.4: Gene duplications and deletions
- 14.5: Orthologs and paralogs
- 14.6: Transposons- moving DNA within a genome (and weird genetics)

15: Mendel and Weldon - Contexts and their Effects on Phenotypes

- 15.0: Introduction
- 15.1: How Mendel did what he did
- 15.2: Weldon's critique
- 15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
- 15.4: Dihybrid crosses- linkage and recombination
- 15.5: Genetic complementation
- 15.6: Interacting traits- synthetic lethality and co-dominance
- 15.7: Interacting traits- epistasis
- 15.8: Maternal and paternal effects
- 15.9: Mitochondrial inheritance
- 15.10: Imprinting- conflicts between mother, father, and fetus
- 15.11: Estimating the number of genes involved in a particular traits

- 15.12: On the nature of mutations (again)
- 15.13: Alleles, traits, and genetic diseases in humans.
- 15.14: Concordance between monozygotic twins and genetic influence on a trait
- 15.15: Measuring evolution’s impact on allele frequencies- Hardy-Weinberg
- 15.16: Genetic anticipation
- 15.17: The persistence of deleterious alleles

16: Tools for Studying Genes and Genomes

- 16.0: Introduction
- 16.1: Synteny examined using Genomicus
- 16.2: Where is a gene expressed?
- 16.3: Using web-based bioinformatic tools- gnomAD
- 16.4: Using web-based bioinformatic tools- BLAST
- 16.5: Genome-wide Association Studies (GWAS)

17: Supplemental Chapter - Fundamental concepts and developing systems

- 17.0: Introduction
- 17.1: How do systems change at the molecular level?
- 17.2: Steady state and changing molecular concentrations- synthesis and degradation
- 17.3: Direct and indirect cellular responses to signaling molecules
- 17.4: Modeling gene expression
- 17.5: Reversible, irreversible, and cascade effects

18: Supplemental Chapter - Social interactions between cells

- 18.0: Introduction
- 18.1: How do unicellular organisms generate phenotypic diversity?
- 18.2: Dying for others – social interactions between “unicellular” organisms
- 18.3: Quorum effects (somewhat redundant)
- 18.4: Transient and clonal ("true") metazoans
- 18.5: Unicellular to multicellularity to unicellularity
- 18.6: Evolutionary origins of clonal (permanent) multicellularity

19: Supplemental Chapter - The role of model systems in understanding metazoan development

- 19.0: Introduction
- 19.1: Model Systems
- 19.2: Some model systems
- 19.3: Frogs and fish
- 19.4: Chick and Quail
- 19.5: The fruit fly Drosophila melanogaster
- 19.6: The nematode Caenorhabditus elegans
- 19.7: The Mouse
- 19.8: ESC and iPSC derived organoids

20: Appendix- Muller’s Morphs

Index

Glossary

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Understanding Science and Thinking Scientifically

1.0: Introduction

1.1: The interconnectedness (self-consistency) of science

1.2: Models, Hypotheses, and Theories

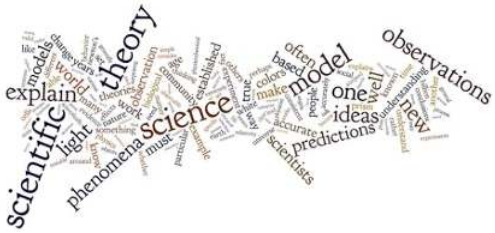
1.3: Science is Social

1.4: Teaching and Learning Science

1.5: Understanding Scientific Ideas

This page titled 1: Understanding Science and Thinking Scientifically is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

1.0: Introduction

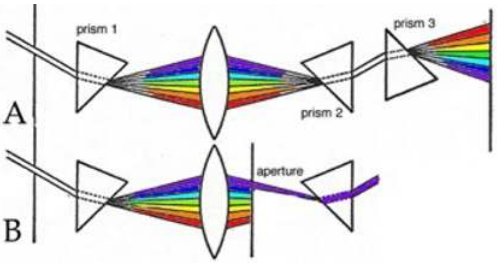


In which we consider what makes science a distinct, productive, and progressive way by which to understand how the universe works, using tools that enable us to distinguish what is possible and plausible from what is impossible or implausible. We consider the "rules" that distinguish a scientific from a non-scientific approach.

> *Revelation is necessarily limited to the first communication – after that it is only an account of something which that person says was a revelation made to him; and though he may find himself obliged to believe it, it can not be incumbent on me to believe it in the same manner; for it was not a revelation made to ME, and I have only his word for it that it was made to him. – Thomas Paine, The Age of Reason.*

A major feature of science that distinguishes it from many other human activities, is its reliance upon shareable experiences rather than personal or reported revelations. Thomas Paine (1737-1809), one of the intellectual parents of the American Revolution, made this point explicitly in "The Age of Reason" (↓).[20] In science, we do not accept that an observation or a conclusion is true because someone claims it to be true. We do not accept the validity of revelations. We require that given a detailed description, others with the resources and the opportunities, will be able to repeat the observation or the experiment. Science is based on social, shared, knowledge rather than revealed "truth".

As an example consider sunlight. It was originally held that white light was "pure" and that when light passed through a prism, the colors of the spectrum, the colors we see in a rainbow, were somehow created. In 1665, Isaac Newton (1642–1727) performed a series of experiments; white light, when passed through a prism emerged as many colors, a rainbow. His interpretation? That white light was not "pure", but composed of light of different colors.[21] To test his hypothesis he used a lens to focus the light from one prism onto a second prism; a beam of white light emerged from the second prism (Part A→). He found that the white light emerging from the prism–lens–prism combination behaved the same as the original beam of white light; when passed through a third prism it again produced a spectrum (right side of Part A→). In a second experiment Newton used a screen with a hole in it (Part B→). Light of a particular color was not altered when it passed through a second prism - no new colors emerged. Based on these observations, Newton concluded that white light was composed, unexpectedly, of many distinct colors and that the rainbow appeared because light of different colors was "bent" (refracted) by the prism to different degrees. Why this occurred was not clear nor was it clear what, exactly, light was. These questions were left unresolved. Scientific answers are often quite specific, describing a particular phenomenon, rather than providing universal explanations.



Two basic features make Newton's approach, observations, and conclusions scientific. The first was reproducibility. Based on his description others could, and did, reproduce, confirm, and extend his observations. You can too! Set up Newton's apparatus and you will observe the same phenomena that Newton did.[22] In 1800, William Herschel (1738-1822) did just that. He used Newton's

experimental approach and discovered infrared (beyond red) light. While infrared light is invisible to us, other organisms can see it. Its presence can be revealed because, when absorbed by an object, say by a thermometer or a human hand, there is an increase in the temperature of the object.[23] Inspired by Herschel's discovery, Johann Ritter (1776-1810) used the ability of light to initiate the chemical reaction:

silver chloride + light → silver + chlorine

to reveal the existence of another type of light that he called "chemical light". We refer to it as ultraviolet light.[24] Visible light accounts for a small portion of a continuous spectrum of "electromagnetic radiation", ranging from X-rays to radio waves. Understanding how light interacts with matter has led to a wide range of technologies and a coherent model of the history of the Universe. All these findings emerge, rather unexpectedly, from attempts to understand the rainbow.

The second scientific aspect of Newton's work was his clear articulation of the meaning and implications of his observations, the logic and limitations of his conclusions. These led to explicit predictions, such as that a particular color of light is homogenous, which he then confirmed. His view was that the different colors of light differ in the way they interact with matter, that this difference was revealed through the extent to which light rays of different colors are bent when they enter and then leave a prism. Newton used these ideas when he chose to use mirrors rather than lenses to build his reflecting (Newtonian) telescope. His design avoided the color distortions that arise when light passes through simple lenses.

Newton's approach illustrates the social and progressive nature of science. We can reproduce an observation or experiment by following the investigator's instructions. Our observations can lead us to identify factors that influence outcomes and identify unappreciated implications that may influence other scientific disciplines. Science rests on the premise that there is a world outside and independent of ourselves. That this world constrains what is possible and what is not – it rules out "magical thinking". Science is not about discovering over-arching and immutable truths (aside from the reality of the world), but rather about developing a working and accurate understanding of how objects in the world behave.

References

[20] The Age of Reason: https://www.ushistory.org/paine/reason/reason1.htm

[21] Newton's Prism Experiments & Khan Academy - Newton's prism experiment

[22] Infrared astronomy

[23] There are some animals that can see infrared light: see link & link

[24] Ritter discovers ultraviolet light

This page titled 1.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

1.1: The interconnectedness (self-consistency) of science



It was once thought that there were aspects of biological systems that transcended physics and chemistry, a presumption known as vitalism. If vitalism were correct, it would force a major revision of chemistry and physics. As it turns out vitalism is, apparently, incorrect. The world described by the sciences is like an extremely complex crossword puzzle (→) where the answer to one question must be compatible with the answers to all other questions.[25] Alternatively, certain questions, and their answers, once thought of as meaningful can come to be seen as irrelevant or meaningless and not part of the puzzle. For example, how many angels can dance on the head of a pin no longer impacts scientific explanations.

Over time, biological processes from the metabolic to the conscious have come to be seen as consistent with physicochemical principles. What makes them different is their complexity. They are the product of evolutionary processes, processes influenced by unpredictable events that influence subsequent events and stretch back in an uninterrupted “chain of being” over billions of years. Biological systems are composed of many types of molecules that interact in complex ways. All this means is that while biological systems obey physicochemical rules, their behavior cannot be predicted based on these rules.

It appears that all life we know of is related. All known organisms are modified (evolved) versions of a “last common universal ancestor”, known as LUCA. It may well be that life, as we know it, is unique in the Universe. If other kinds of life are possible, we have no evidence for them. We do not know the “general rules” governing life because we only know one type of life, that found on Earth.

References

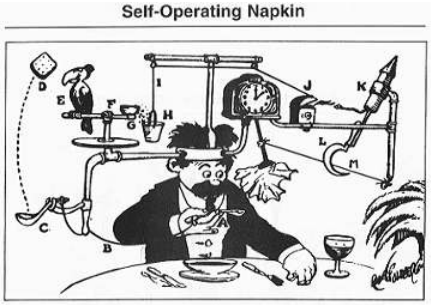
[25] This analogy is taken from a talk by Alan Sokal

This page titled 1.1: The interconnectedness (self-consistency) of science is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

1.2: Models, Hypotheses, and Theories

Scientific models are used in various ways. There are explanatory models that present an approach to a system and exploratory and predictive models used to test ideas. Predictive, mechanistic models are commonly known as hypotheses. Models serve to clearly articulate one's assumptions and their implications. They form the logical basis for generating testable predictions about the phenomena they purport to explain. As scientific models become more sophisticated, their predictions become increasingly accurate and/or apply to areas that previous versions of the model could not handle.

Let us assume that two models are equally good at explaining a particular observation. How do we decide between them? One way is the rule of thumb known as Occam's Razor (or the Principle of Parsimony), named after the medieval philosopher William of Occam (1287–1347).[26] Occam's Razor states that all other things being equal, the simplest explanation is the best. Classic examples of non-simple explanations are Rube Goldberg (1883-1970) machines (→).[27] Occam's Razor does not imply that an accurate scientific explanation will be simple, or that simple explanations will be correct, just that a scientific model should not be more complex than it needs to be. Consider two models for a particular phenomenon, one that involves angels and the other that does not. We need not seriously consider the model that invokes angels unless we can accurately monitor the presence of angels and their involvement in the process to be explained. Why? Because angels, if they exist, imply serious complexities. We would need to explain what angels are made of, their origins, and how they interact with the physical world, specifically how they make matter move. Do angels obey the laws of thermodynamics? What determines when and where they intervene? Are their interventions purposeful or capricious? If an angel-free model is as or more accurate at describing the phenomena and making verifiable predictions, the scientific choice would be the angel-free model. Parsimony, in science, is a preference for simplicity; it serves to constrain our thinking to the simplest model that explains the phenomena under study. The surprising result, illustrated in a talk by the physicist Murray Gell-Mann[28], is that simple, albeit often counter-intuitive rules can explain much of the Universe with remarkable precision. A model that fails to accurately describe and predict the observable world must be missing something and is either partially or completely wrong, no matter how "beautiful" or appealing.



Scientific models are continually being modified, expanded, or replaced in order to explain more and more phenomena more and more accurately. It is an implicit assumption of science that the Universe can be understood in scientific terms, and this presumption has been confirmed but by no means proven. A model that has been repeatedly confirmed is known as a theory.[29] It is worth noting that the word theory is often misused, even by scientists who should know better. If there are multiple "theories" to explain a particular phenomenon, it is more correct to say that i) these are not theories, but rather working models, hypotheses, or speculations, and that ii) one or more, and perhaps all are incorrect or incomplete. If you are curious, you might count how many times the word theory is misused, at least in the scientific sense, in the course of your day to day experiences.

That said, theories are not static. New or more accurate observations that a theory cannot explain will drive the theory's revision or replacement. The new theory explains the new observations as well as everything explained by the older theory. Consider gravity. Isaac Newton's law of gravity describes how objects behave; it is possible to make extremely accurate predictions using its rules. However, Newton did not have a theory of gravity; a naturalistic and mechanistic explanation for why gravity exists and behaves the way it does. He relied on a supernatural explanation.[30] Later on it was discovered that Newton's law of gravity failed in specific situations, such as when an object, such as the planet Mercury, is close to a massive object, the sun. New rules were needed. Albert Einstein's Theory of General Relativity more accurately predicts the behavior of these systems, and provides a naturalistic explanation for the origin of gravitational forces.[31] It makes predictions about future observations, such as gravity

waves, that have been confirmed.[32] So is general relativity "true"? Not necessarily. Scientists continue to test its predictions in increasingly extreme situations and to higher and higher degrees of accuracy.

> *Gravity explains the motions of the planets, but it cannot explain who sets the planets in motion.*
>
> *- Isaac Newton*

What do you know? How do you know you know it? How do you know its "true"?

How do we know what we know? This is a central question in philosophy and certainly relevant to teaching and learning. There is plenty of evidence that people consistently over-estimate their own knowledge and skills (including what they believe they have learned in a class).[33] There is, however, a well-established approach to evaluating one's knowledge, namely a "socratic" conversation with an engaged, critical, and knowledgeable person; such a conversation can reveal underlying assumptions. We use a socratic approach in beSocratic activities; we ask you to explain and illustrate your responses to questions. Is your use of scientific concepts and observations appropriate and logical? Have you left out important facts or considerations? Are unspoken (or unsupported) assumptions involved? Get ready to respond to the "questions to answer and ponder" in the book, in class, or asked in beSocratic. Unlike the citizens of Athens (who sentenced Socrates to death), these are not meant to be threatening, but provide opportunities to strengthen your understanding of biological processes and systems.

When answering a question, it helps to know exactly what the question wants you to explain. The ability to decode a question, and then compose a coherent and evidence-based response, is a learned skill that requires practice and knowledge.[34] This is exactly what working scientists do! We hope that talking with a socratic tutor (in the context of beSocratic) will help you get comfortable asking questions.[35] If you are confused, ask for clarification – what exactly is the question asking? what do you find confusing about it? Can you restate the question and answer it? Asking questions in class or to a classmate can clarify what a question is about. If they are equally confused ask the instructor. Typically we share questions and responses with the class, since it is very likely that you are not the only person who wants or needs clarification.

Once you understand what a question wants, take a breath and identify what facts and general principles you will need to construct an answer. Consider the question: "Based on the accumulation of an isotope known to be generated only by radioactive decay, a geologist claims that a particular rock is ~2 billion years old. A creationist claims that the rock is no more than 6000 years old. Why can't both be right?" To answer the question we begin by stating our assumptions. Geologists date rocks based on assumptions about the rock's formation and stability together with the observed rates of radioactive decay and the assumption that decay rates are constant. These assumptions enable one to deduce the age of the rock from the ratio of the specific isotopes present. As a rule, fossils are found in sedimentary rocks, but such rocks are difficult to date accurately, since they are derived, through processes of erosion and deposition of older rocks. The age of fossil containing rocks are based on the age of surrounding igneous (volcanic) rock layers. It is less clear what scientific ideas, if any, a creationist uses to date rocks and the fossils within them. Since there is no evidence that rates of radioactive decay have changed over the history of the Universe, and assuming no other natural processes are at play, a creationist is most likely to be scientifically incorrect – their assumptions implicitly contradict well established knowledge from physics, chemistry, and geology.

As you can see, answering a question can be complex – an answer often relies on a number of assumptions that need to be explicitly recognized. In dating a fossil, you need to consider the observed rate of radioactive decay, the method used to date sedimentary (and igneous) rocks, and the mechanism(s) by which fossils are generated. The complexity of explaining why correct answers are correct is one of the reasons that we often ask you to explain why wrong answers, such as those found in multiple-choice tests, are wrong or irrelevant. Typically a wrong answer is wrong for a single reason or, if correct, is irrelevant to the question at hand.

A similar situation applies when explaining something to someone, you need to identify what the person you are talking to needs to know to be able to understand your explanation. Consider the video interview with the physicist Richard Feynman (1918-1988) [video link→]; in it he describes what it takes to explain magnetic attraction. As you start answering or explaining, you need to be prepared to explain the underlying ideas you are using – the person you are talking with expects you to be able to justify your assumptions, clarify your logic, and defend your conclusions. The same applies when you listen to an explanation by an instructor; do their assumptions make sense to you? Are they telling you all you need to know to be able to understand their explanation?



Similarly, when listening to an explanation, consider whether the evidence they are using is correct, relevant, and complete, do their conclusions follow logically? In a scientific discussion, are the methods they used capable of generating the data upon which their argument rests? Through these interactions, you are taking part in a Socratic dialog.

It can be helpful to study with a group of people who are comfortable questioning and explaining to each other, but beware, groups do not always arrive at coherent or reasonable conclusions. Check the group's conclusions by presenting them to a knowledgeable expert (hopefully your instructor). When working on your own, it is helpful to develop your own "inner Socrates", the habits of mind that help challenge and refine your thinking. Ask yourself "am I answering the question I am being asked? have I identified the key ideas and observations needed to answer the question? Are there other observations or concepts that need to be considered? Are other, simpler explanations possible?" This is one area in which talking out loud to yourself can be useful!

Questions to answer:

1. How can Occam's Razor be applied when two models appear to be equally accurate? What are the benefits of using this principle? (see footnote 41)
2. Can you describe the process you use to determine if an idea, model, or hypothesis qualifies as scientific? Include considerations such as testability, falsifiability, and the use of empirical evidence.

References

[26] William of Ockham

[27] Wikipedia – Rube Goldberg

[28] Murray Gell-Mann: Beauty, truth and ... physics?

[29] Ideas are cheap, theories are hard

[30] Want to read an interesting biography of Newton, check out "Isaac Newton" by James Gleick

[31] A good video on General Relativity [here]

[32] Physicists find another gravitational wave to suggest that Einstein was right

[33] The Kruger & Dunning effect: Unskilled and Unaware

[34] Norris & Phillips. 2003. How literacy in its fundamental sense is central to scientific literacy

[35] The answers can often be surprising. see McClymers & Knowles.Ersatz Learning, Inauthentic Testing

---

This page titled 1.2: Models, Hypotheses, and Theories is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

1.3: Science is Social

Science is often portrayed as an activity carried out by isolated (and often crazy, "brilliant", or otherwise deranged) individuals.[36] The image of the mad scientist comes to mind (→). Reality is different, science is a fundamentally social activity. It works because it depends upon an interactive community who keep each other, in the long run, honest and anchored in objective, observable and reproducible reality. Scientists present their observations, hypotheses, and conclusions in the form of scientific papers, where their relevance and accuracy can be evaluated, more or less dispassionately, by others with a working knowledge of the topic under study.



Over the long term, these socratic interactions that leads to a logical, evidence-based consensus. Certain ideas and observations are so well established that they appear to be universally valid, whereas others are extremely unlikely to be true. We can be confident that perpetual motion machines and zero-waste processes (versions of the same idea) are impossible. As we see it, modern biology is based on a small set of assumptions that include the Physicochemical and Cell Theories of Life, and the Theory of Evolution.[37] That said, as scientists we keep our minds open to novel observations and work to understand them and their implications. The openness of science means that one person, taking a new observation or idea seriously, can challenge and change accepted scientific understanding. That is not to say that it is easy to change the way scientists think about the world. Most theories are based on large bodies of evidence and have been confirmed on multiple occasions using multiple methods. It turns out that most “revolutionary” observations are either mistaken, misinterpreted, or can be explained within the context of established theories. It is, however, worth keeping in mind that it is not at all clear that all phenomena can be put into a single “theory of everything.” It has certainly proven difficult to reconcile quantum mechanics with general relativity (or so have been told)[38].

A final point, mentioned before, is that the sciences are not independent of one another. Ideas about the behavior of biological systems cannot contradict well established observations and theories in chemistry or physics. If they did, one or the other would need to be modified. For example, there is substantial evidence for the dating of rocks based on the behavior of radioactive isotopes. There are well established patterns of where rock layers of specific ages are found. When we consider the dating of fossils, we use rules and evidence established by geologists. We cannot change the age we assign to a fossil, making it inconsistent with the rocks that surround it, without challenging our understanding of the atomic nature of matter, the quantum mechanical principles involved in isotope stability, or a range of geological mechanisms and established patterns of Earth history.

A classic example involves the physicist William Thompson (1824-1907), also known as Lord Kelvin. Based on thermodynamic principles, he estimated the age of the Earth to be between ~20 to ~100 million years. He assumed that the Earth was once completely molten together with the known rate of heat dissipation of such a massive molten object.[39] This was a time-span that seemed too short for a number of geological and evolutionary processes, and greatly troubled Charles Darwin. Somebody was wrong, or better put, current scientific understanding was incomplete or incorrect. The answer in this case was with the fact that Thompson ignored the effects of radioactive decay, not surprising since radioactivity had not yet been discovered. Including the heat released by radioactive decay in such calculations led to an increase in the estimated age of the Earth to ~4.54 billion years, an age compatible with both evolutionary and geological processes.

References

[36] A good introduction of how science can be perverted is “The Undergrowth of Science” by Walter Gatzer. You might also want to watch the “The Centrifuge Brain Project” | A Short Film by Till Nowak and consider whether it is scientific or not.

[37] Thinking about the conceptual foundations of the biological sciences

[38] Why Quantum Physics Messes With Reality - Sabine Hossenfelder

[39] An interesting book on this topic is “Discarded Science: Ideas That Seemed Good at the Time” by Paul Barnett

This page titled 1.3: Science is Social is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

1.4: Teaching and Learning Science

Science involves the reconciliation of new observations and ideas with those previously established. This process can lead to conclusions that appear strange, counterintuitive, and sometimes disconcerting but that are nevertheless logically unavoidable. It is now accepted that the Earth rotates around its axis and travels around the sun, which itself is moving around the center of the Milky Way galaxy, and that the Universe as a whole is expanding at what appears to be an ever increasing rate. At the same time, none of these facts are immediately obvious. Relatively few people can explain exactly how we have come to know that these ideas accurately reflect the way the universe works (or at least how it appears to work). At the same time, when these conclusions were first developed they conflicted with the assumption that the Earth is stationary (which it appears to be) and sits at the center of a static Universe. New ideas about the Earth's actual position in the Universe could be seen as a threat to the sociopolitical and religious order. In fact, a number of people were persecuted for holding "heretical" views on the topic. Most famously, the mystic Giordano Bruno (1548-1600) was burnt at the stake for promoting these and other ideas. Galileo Galilei (1564-1642) one of the founders of modern physics, was arrested in 1633, tried by the Roman Catholic Inquisition, forced to publicly recant his views on the relative position of the Sun and Earth, and spent the rest of his life under house arrest.[40] In 1616 the Church placed Galileo's book, which held that the sun was the center of the solar system, on the list of forbidden books – it remained there until 1835.

The idea that we are standing on the surface of a planet that is rotating at ~1000 miles an hour and flying through space at ~67,000 miles per hour is, understandably, difficult to reconcile with our everyday experience, yet science continues to generate, and provide confirmatory evidence for much weirder ideas. Based on various observations, it appears that the Universe arose from "nothing" ~13.8 billion years ago[41] and will continue to expand forever. Einstein's theory of general relativity implies that matter distorts space-time, which is really one rather than two discrete entities, and that this distortion produces gravity and black holes. A range of observations indicate that all organisms are derived from a single type of ancestral uni-cellular organism (LUCA) that arose from non-living material between ~3.5 to 4 billion years ago. There appears to be an uninterrupted link between LUCA and every cell in your body, and to the cells within every other living organism, including microbes that live in your gut. You yourself are a staggeringly complex collection of cells. Your brain and its associated sensory organs generate consciousness and self-consciousness. Your nervous system contains ~86 billion neurons as well as a similar number of non-neuronal (glial) cells. These cells interact with one another through ~1.5 x 10^{14} connections, known as synapses.[42] How exactly such a system produces thoughts, ideas, dreams, feelings, and self-awareness remains obscure, but it appears that these "emergent" behaviors arise from this staggeringly complex natural system. Scientific ideas, however weird, arise from the interactions between the physical world, our brains, and the social system of science that rigorously tests ideas based on their ability to explain and predict the behaviors of the observable universe.

References

[40] The History, Philosophy, and Impact of the Index of Prohibited Books

[41] The Origin Of The Universe: From Nothing Everything?

[42] Are There Really as Many Neurons in the Human Brain as Stars in the Milky Way? & Shapson-Coe et al, 2024.A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution

This page titled 1.4: Teaching and Learning Science is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

1.5: Understanding Scientific Ideas

A difficulty in understanding scientific ideas and their implications is that these ideas build upon a wide range of interdependent observations and empirically established rules. Understanding biological systems depends upon an understanding of chemical reaction systems that in turn depends on an understanding of molecules that rests upon an understanding of atoms, energy and their interactions. Our working premise is that to understand a topic it is necessary to know the key observations and common principles upon which basic conclusions and working concepts are based. To test one's understanding of a system, you need to be able to construct plausible claims for how, and why the system behaves the way it does, and how various perturbations might influence it. Your analysis needs to be based on facts, observations, or explicit presumptions that combine, logically, to support your claim. You need to be able to present your model to others, knowledgeable in the topic, in a clear way to get their feedback, to answer rather than ignore or disparage their questions, and to address their criticisms and concerns.[43] Sometimes you will be wrong because your knowledge of the facts is incomplete or inaccurate, your understanding or application of general principles is incorrect, or your logic is faulty. It is important to appreciate that generating coherent scientific explanations and arguments can be difficult. Through useful coaching and practice, we hope to help you learn how to understand biological systems and processes. We, your fellow students, and your inner Socrates will help you identify when you produce a coherent critique, explanation or prediction, and where you fall short.

Distinguishing the scientific from the non-scientific, religious, and ideological

When we consider various personal and public policy decisions we are often told that science has reached a consensus, but what exactly does that mean? By consensus, we mean the common conclusions accepted by scientists working in the field, conclusions supported by available evidence. But evidence is rarely complete. When approaching a system scientifically, it is often necessary to make simplifying assumptions. These assumptions make the system tractable but also somewhat "unreal"; the real system can behave differently. Given the uniqueness of each organism (a topic we will discuss), how any particular person responds to a particular drug is influenced by many, often interacting, factors, not perfectly defined or definable in our model. Interventions can have individual-specific side-effects, both desirable and undesirable. There are risks in taking (or not taking) a drug, getting vaccinated, undergoing a surgery, opening or closing nuclear (or coal-based) power plants, but knowing **exactly** what the costs and benefits are may be difficult to predict. Any cost-benefit analysis often involves non-scientific factors as well.

Questions to answer:

3. Given a news story that claims that spirit forces influence the weather, develop a set of questions to evaluate the scientific plausibility of this claim. Consider aspects such as empirical evidence, testability, and consistency with established scientific principles.*
4. Imagine a scenario where scientific research concludes that free will is an illusion. Develop a set of questions to explore the implications of this conclusion on human behavior and decision-making. How might this influence ethical, social, and personal perspectives on autonomy and responsibility?*

Questions to ponder

- Is attaining "truth" and developing a theory of everything the actual goal of science?
- How much money should we spent to better understand the natural world.
- How should we, as a society, deal with the tentative nature of scientific knowledge?



Short chapter summary

This page titled 1.5: Understanding Scientific Ideas is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

2: Life’s Diversity and Origins

2.0: Introduction
2.1: What is life, exactly?
2.2: The cell theory and the continuity of life
2.3: The organization of organisms
2.4: Spontaneous generation and the origin of life
2.5: The Death of Vitalism
2.6: Thinking about life’s origins
2.7: Experimental studies on the origins of life
2.8: Mapping the history of life on earth
2.9: Fossil evidence for the history of life on earth
2.10: Life's impact on earth

This page titled 2: Life’s Diversity and Origins is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.0: Introduction



In which we consider what biology is all about, namely the study of life, that is organisms and how they work. We discover that organisms are built of one or more, to millions to billions of cells. Social processes are involved in multicellular organisms and when single-celled organisms act in a coordinated manner. We consider plausible models for the origins of organisms and their relationships to one another.

Biology is the science of organisms, how they function, behave, interact, and vary genetically from one another. How individual organisms adapt and how populations of organisms evolve over time. Organisms are discrete, highly organized, bounded but open, non-equilibrium physicochemical systems. That is a lot of words! what do they all mean? How is a mushroom that looks like a rock different from a rock? What is genetic variation and how does it influence the properties and behaviors of an organism? How, exactly, is one mushroom different from another? What exactly is a bounded, non-equilibrium system? The answers are not simple; they assume a working knowledge of core principles from physics and chemistry and experimental observations. For example, to understand what it means to be a “bounded, non-equilibrium system” you need to understand basic thermodynamics (which we consider in Chapter 5). Non-equilibrium systems can do various forms of work, and work is an outcome or behavior that requires energy to achieve. Work includes generating movement and molecular gradients, and driving a range of unfavorable reactions that include the synthesis of nucleic acids, proteins, lipids, carbohydrates, and other types of molecules.

We will focus on what is known as Gibbs free energy; the energy available to do work. The free energy of a system at equilibrium is zero. Organisms maintain their non-equilibrium state by importing free energy in the form of light and chemically unstable molecules from the external world. Organisms contain information that can be replicated, changed, and passed from parent to offspring. Other types of non-equilibrium systems, such as hurricanes and tornados, occur – they differ from organisms in that they are transient and arise *de novo*, that is, they do not have “parents”. When they disappear they leave no offspring, no baby hurricanes or tornados. In contrast, each organism alive today arose from one or more pre-existing organisms, its parent(s) and that each and every organism, past, present, and future has an uninterrupted history stretching back billions of years. A remarkable conclusion, given the fragility of life.

Biology is based on a few over arching theories. The Cell Theory explains the historic continuity of organisms, while Evolution Theory explains the observed fossil record, the diversity of organisms, how populations of organisms change over time, and extinction. The Physicochemical Theory of Life explains the remarkable properties of (and behaviors) of organisms without violating the laws that govern all physical and chemical systems.[44]

References

[44] Thinking about the conceptual foundations of the biological sciences

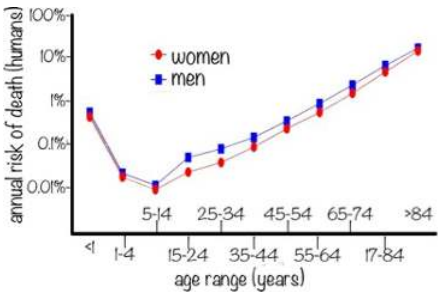
This page titled 2.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.1: What is life, exactly?

When talking about biology, we need to define what we mean by life. While some talk about the possibility of various "forms" of life, in fact we know of only one fundamental form of life - derived from LUCA. These share a range of common structures and molecular processes, best illustrated by the fact that they encode hereditary information in molecules of deoxyribonucleic acid (DNA) using the same "genetic code" and mechanisms to access and use that information. Human genetic information "makes sense" to bacteria, an amazing observation!

We do not know whether the origin of life on Earth was a likely and predictable or whether life's origin(s) is rare and unlikely. The discovery of extraterrestrial life, or organisms unrelated to LUCA, would provide valuable data.[45] An alternative would be to assemble "new" living systems in the laboratory. Until someone manages to create or identify non-standard forms of life, it seems reasonable to focus on the characteristics of life as we know them.

So, what do we mean by life. First, the core units of life are cells. From a structural and thermodynamic perspective, each cell is a bounded, non-equilibrium system that persists over time and has the potential to produce one or more copies of itself. Cells can exist as genetically distinct individuals or genetically-related communities displaying various levels of communication and interdependence. Sometimes organisms form closely integrated, mutualistic relationships and can be difficult to grow in isolation from one another.[46]



Why the requirement for, and emphasis on reproduction? Assume that a non-reproducing form of life was possible. Any such system runs the risk of death, or perhaps better put, accidental extinction. Over time, the probability of death for any individual approaches certainty (→).[47] In contrast, a system that reproduces makes multiple copies of itself and so minimizes, although by no means eliminates, the chance that its offspring will go extinct. While there have been a number of mass extinction events during the history of life on Earth, LUCA's descendants continue to survive, flourish, and evolve.[48]

Living systems are a unique family of non-equilibrium physicochemical systems. Coming to equilibrium means death, an irreversible process. The non-equilibrium nature of living systems drives a range of processes and behaviors, including growth, reproduction, directed physical movements, and thought. All living systems we know of appear to be related, sharing a unique origin on Earth. To maintain their non-equilibrium state, living systems must import "free" energy. Living systems capture energy directly by absorbing photons (in plants) and by ingesting thermodynamically unstable molecules (non-plants). Biological systems have distinct boundaries that separate them from the external, non-living world.[49] The cellular barrier layer retains molecules generated by the living cell while allowing waste products to leave. We review the relevant laws of thermodynamics in Chapter 5.

Questions to answer:

5. How might you determine if a specific object is alive?
6. Consider the risk of death graph (↑). Provide a plausible explanation for the graph's shape; what factors might be expected to influence the various regions of the curve?
7. Should the points in the graph be connected or is a smooth “best fit” curve a more accurate way to describe the system? Justify your reasoning.

References

[45] The possibility of alternative microbial life on Earth Signatures of a shadow biosphere Life on Earth but not as we know it

[46] Cultured Asgard archaea shed light on eukaryogenesis by Lopez-Garcıa & Moreir 2020.

[47] Image modified from “risk of death” graph: http://www.medicine.ox.ac.uk/bandolier/booth/Risk/dyingage.html

[48] Mass extinction events

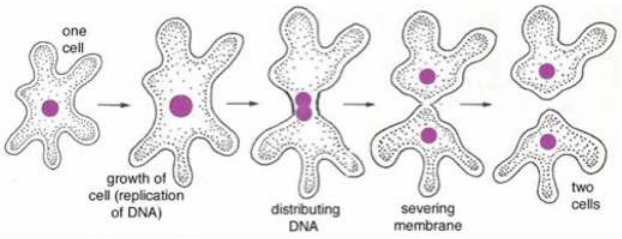
[49] Suggestions that every thing is conscious (see link) seem quite silly.

This page titled 2.1: What is life, exactly? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.2: The cell theory and the continuity of life

Toward the end of the 1800's, observations using microscopes suggested that all organisms contained structurally similar units, termed "cells." Based on such observations, a rather sweeping conclusion, the Cell Theory, was formulated. The Cell Theory has two distinct parts. The first is the prediction that every organism is composed of one or more, and in some cases millions to billions, of cells together with secreted products, such as bone, hair, scales, and slime. The Cell Theory postulates that cells are membrane-bounded, open (in a controlled way), non-equilibrium systems. Up to the present day) there has been no evidence that modern cells can form from non-cellular materials. Therefore the second part of the Cell Theory is that cells arise only from pre-existing cells. The implication is that organisms, and the cells that form them, arise from pre-existing cells. Cell Theory says nothing about how the first cell or how life on Earth appeared.

Cells contain hereditary information, stored in a physical and relatively stable form, in molecules of double-stranded deoxyribonucleic acid (DNA). The Cell Theory implies that all organisms that exist now and in the past are related through an unbroken series of DNA replication and cell division (reproductive) events. Based on the information present in DNA together with, molecular level comparisons of how cells are constructed suggests that they are derived from a single common ancestor. This organism, LUCA, lived between ~3.5 to ~4 billion years ago. It implies that each cell in every organism, including all of the cells that make you up, has an uninterrupted history going back billions of years. Exactly how the first cells formed and what they looked like is unknown and unknowable.



The "birth" of a new cell is a continuous process by which one cell becomes two. A new cell is formed when the original surface (or plasma) membrane pinches off to form two distinct cells (→). Note, there is no sharp discontinuity. The new cell does not spring into existence *de novo* but emerges from the preexisting cell. We often define the start of a "new" life with the completion of cell division. At the organismic level, we mark the origin of a new organism with the fusion of gametes (e.g. egg and sperm cells). There is no sharp discontinuity, both gametes are derived from existing cells and their fusion results in the formation of a new hybrid cell.

The information stored in DNA molecules, an organism's genotype or genome, is more stable than the organism itself but it is dead. It can survive, for a time, the death of the organism. As we will see, it can be transferred from cell to cell, a process known as horizontal gene transfer. That said, DNA means nothing outside of a system that can interpret the information stored within it.[50]

References

[50] A DNA-Based Archival Storage System

This page titled 2.2: The cell theory and the continuity of life is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.3: The organization of organisms

Some organisms consist of a single cell, some of many, often of distinct "types". Cells can be highly specialized, yet they all appear related, sharing many molecular and structural details. So why do we consider the organism rather than the cell to be the basic unit of life? The distinction is a matter of reality versus abstraction. It is organisms, whether single- or multi-cellular, that produce new organisms. A single cell within a multicellular organism normally cannot survive outside the organism nor can it produce a new organism – reproduction depends upon cooperation between the cells of the organism. A multicellular organism is a cooperative, highly integrated social system.

In a typical multicellular organism most cells have given up their ability to reproduce a new organism; their future depends upon the reproductive success of the organism as a whole. Within the organism the cells that give rise to the next generation are known as germ cells; cells that do not give rise to new organisms are known as somatic cells. All organisms arise from a pre-existing organism or through the cooperation of two sexually reproducing organisms. We will also see that breakdowns in social systems can lead to the death of the organism or the disruption of a social system. Cancer is the most obvious example of an anti-social cellular behavior. In the short term, cancerous behavior maybe "rewarded" (more copies of the cancer cell are produced) but ultimately it leads to the death of the organism and the extinction of the cancer cells.[51] Evolutionary mechanisms involve cost-benefit “calculations” based on in terms of reproductive success.

References

[51] Cancer cells as sociopaths: cancer's cheating ways Recently the situation has gotten more complex with the recognition of transmissible cancers in dogs and Tasmanian devils.

This page titled 2.3: The organization of organisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.4: Spontaneous generation and the origin of life

The diversity of organisms, estimated to be in the millions of distinct species[52] raises two obvious questions: how did life start and why are there so many different types of organisms? At one point, people believed that these questions had a single answer. We now recognize that they are distinct questions and their scientific answers involve distinct mechanisms. Originally supernatural processes were thought to be necessary to "create" the different "forms" of life (including human beings). In the light of Cell and Evolution theories, and the accumulation of molecular (gene/genome sequence) it seems a certainty that life on Earth had a single successful origin.

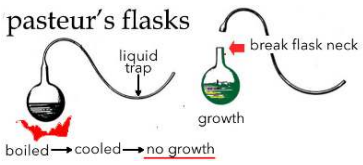
How then did life originate? It was once believed by some that organisms, such as flies, frogs, and mice, arose spontaneously from non-living matter.[53] If true, spontaneous generation would have profound implications for our understanding of biological systems. It would, presumably, involve a rather simple process. All bets are off if the origin of life process is supernatural. If each organism arose independently, we might expect that, at the molecular level, details of each would be unique, since they presumably arose independently from different stuff and under different conditions compared to other organisms. This does not appear to be the case; all known organisms use similar molecular mechanisms and are composed of structurally similar cells.

> *He who experiments increases knowledge. He who merely speculates piles error upon error.*
>
> *- Arabic epigraph quoted by Francisco Redi.*

A key event in the development of modern biology was the publication of Francesco Redi's (1626-1697) paper "Experiments on the Generation of Insects". His hypothesis (an observation-informed guess) was that spontaneous generation did not occur.[54] He thought that the organisms that appeared had developed from "seeds" deposited by adults. One prediction of his hypothesis was that if adult flies were kept away from rotting meat maggots, the larval form of flies, would not appear no matter how long one waited and that the type of organism that appeared would not depend on the type of rotting meat, but rather on the type of adult fly that had access to the meat. To test his hypothesis Redi set up two sets of flasks both of which contained fresh meat. One set of flasks were sealed with paper or cloth so that no flies could enter, the "control" condition. The experimental condition was exposed to the air and so to flies. Maggots appeared only in the flasks open to the air (and flies). Redi concluded that organisms as complex as insects, and too large to pass through the cloth, could arise only from other insects, or rather from the eggs laid by those insects. Life came from life.

The invention and use of the light microscope by Antony van Leeuwenhoek (1632-1723) and Robert Hooke (1635-1703) led to the discovery of a completely new and unexpected world of organisms, known as microbes.[55] While it was relatively easy to generate compelling evidence that macroscopic (that is, big) organisms, such as flies, mice, and people did not arise spontaneously, it seemed plausible that microscopic, and presumably much simpler, organisms could. Some began to study such "microorganisms". Lazzaro Spallanzani (1729-1799) found that when a broth was boiled and isolated from contact with the air, it remained sterile. He concluded that microscopic organisms, like larger macroscopic organisms, did not arise spontaneously but were descended from other microbes, many of which were floating in the air.

A key aspect of scientific thinking is to think carefully about the limitations of a particular observation or experiment. Are there aspects of Spallanzani's experiment that would leave open the possibility of the spontaneous origin of microbes? One possibility was that boiling the broth destroyed a key component that were necessary for the spontaneous generation, or perhaps exposure to fresh air was the "vital" ingredient. Boiling and isolation could produce an unintended effect that obscured rather than revealed the process. Louis Pasteur (1822-1895) carried out a particularly convincing set of experiments to address both concerns. He sterilized broths by boiling them in special "swan-necked" flasks. What was unique about his experimental design was the shape of the neck.



Microbes in the air were trapped in the bent region of the flask's neck (←); the result was that air but not air-borne microorganisms to reach the broth. This design enabled Pasteur to address these concerns. He found that the boiled broths, even with access to air, remained sterile for months, but when the flask neck was broken (red arrow) the broth was quickly overrun with microbial growth. He conclusion? Air, by itself, was not necessary for spontaneous generation, but rather contained microbes. The fact that the broth could support microbial growth after the neck was broken served as a "positive control"; it indicated that the heating of the broth had not destroyed some vital element needed to

support growth. Positive control experiments to test whether specific assumptions are correct. For example, if we are using a drug in a study, we need to establish (rather than take someone's word for it) that the sample of the drug we are using is active. In Pasteur's experiment if, after the neck was broken, the boiled broth could not support growth we would not expect it to support spontaneous generation, and so the experiment would be meaningless. We will return to the description of "negative control" experiments later.[56]

No experiment is perfect, nor does it have to be for science to work. How would you argue against the idea that spontaneous generation takes thousands to millions of years to occur? A direct experiment would have its own practical issues (can you identify some?). So far, many studies have led to the consistent conclusion that neither microscopic nor macroscopic organisms can arise spontaneously in nordern world.

> *It is often said that all the conditions for the first production of living organisms are now present. But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc. present, that a proteine compound was formed, ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed.*
>
> *- Charles Darwin (1887).*

But we assume that spontaneous generation occurred at least once. What explains its absence of in the modern world? Living systems involve complex chemical reaction networks. In the modern world, there are many organisms around, essentially everywhere; these organisms are eating complex molecules to maintain their living state. Given the tendency of organisms to eat one another, one might argue, as did Darwin →, that once organisms had appeared in a particular environment they would suppress subsequent events – they would have eaten the molecules needed for spontaneous generation to occur. There are no welcoming and sterile, that is, life-less places left for new forms of life to arise.

Here we see the importance of history. It appears that a new form of life could arise only in the absence of life. Studies of Earth's history put some limits on the time it took for life to appear, sometime after the Earth's surface solidified from its early molten state to the first fossil evidence for life, about 100 to 500 million years. Once life arose conditions changed.Once organisms were present, only there descendants could survive. In such a system, history matters.

References

[52] Ritchie (2022) "How many species are there?" Retrieved from: 'https://ourworldindata.org/how-many-species-are-there'

[53] Farley. The spontaneous generation controversy (1700-1860): The origin of parasitic worms. and The spontaneous generation controversy(1859-1880): British and German reactions to the problem of abiogenesis.

[54] see Richard Feynman's description of the role of guessing in the scientific process

[55] see the wikipedia article on protists

[56] Wikipedia on control experiments and observations

This page titled 2.4: Spontaneous generation and the origin of life is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.5: The Death of Vitalism

Naturalists originally thought that life was a supernatural process, too complex to obey or be understood using the laws of chemistry and physics.[57] In this vitalistic view, organisms were thought to obey laws different from those acting in the non-living world. It was assumed that "organic" molecules. those thought to exist only in living organisms could not be synthesized outside of an organism; they had to be made by a living organism. Friedrich Wöhler (1800–1882) challenged this view by synthesizing urea in the laboratory. Urea ($O=C(NH_2)_2$) is a simple organic molecule found in the waste derived from living organisms. Urine contains lots of urea. Wöhler's *in vitro* or "in glass" synthesis of urea was simple. Attempting to synthesize ammonium cyanate (NH_4NCO), he mixed the inorganic compounds ammonium chloride (NH_4Cl) and silver cyanate (AgNCO), urea was found in the products of this reaction. What happened was this reaction:

$AgNCO + NH_4Cl \rightarrow NH_4NCO + AgCl \rightarrow O=C(NH_2)_2 + AgCl$.

Please do not memorize this reaction, but recognize that this reaction involves two inorganic compounds. The urea synthesized in Wöhler's reaction is identical to the "natural" urea found in urine.

While simple, the *in vitro* synthesis of urea had a profound impact on the way scientists viewed living processes. It implied that there was nothing supernatural in the way organisms worked. This is not to say that such syntheses are simple. Over time, organic chemistry has "evolved" from the study of molecules found in organisms to the study of molecules containing carbon.

Questions to answer:

8. Why did the discovery of bacteria reopen the debate on spontaneous generation?
9. In Pasteur's experiment would you expect to see microbial growth in the flask's bent loop? Explain your thinking.
10. Why is a positive control experiment important, what does it tell you?

Questions to ponder

- Is the assumption of spontaneous generation unscientific? What types of (scientific) evidence would support the view that the origin of life (or consciousness) requires supernatural intervention?

References

[57] In a sense this is true since many physicists at least do not seem to understand biology.

This page titled 2.5: The Death of Vitalism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.6: Thinking about life's origins

A religious (i.e., non-scientific) approach would likely postulate that life has a supernatural origin. Since the process is supernatural it cannot, by definition, be studied scientifically. Intelligent design creationists claim that they can identify those aspects of life that are beyond natural processes. Would these claims, if true, force us to abandon science in general? Given the interconnectedness of the sciences, does a supernatural biology call into question the validity of all scientific disciplines? The dating of fossils is based on geological and astrophysical (cosmological) evidence for the age of the Earth and the Universe, which are themselves based on physical and chemical observations and principles. A truly non-scientific biology would be incompatible with physics and chemistry.

It is worth remembering that predictions as to what is beyond the ability of science to explain have routinely been found to be wrong, often only a few years after being made! This speaks to the power of science and science-based technologies. For example, would an intelligent design creationist be tempted to synthesize human proteins in bacteria or plants, something now done routinely to make a range of drugs, such as insulin?[58] Would they predict that a few genetic modifications could make it possible to transplant pig hearts (and other organs) into people, so as to reduce rejection by the immune systems?[59]

An alternative explanation for the appearance of life on Earth, termed panspermia, assumes that advanced aliens brought (or left) life on Earth. Perhaps we owe our origins to casually discarded alien litter. Unfortunately, the principles of general relativity, one of the best confirmed of all scientific theories, limit the speed of travel. Given the size of the Universe, travelers from beyond the solar system are unlikely. There seems to be little or no chance that StarWars-type storm troopers will be arriving. More to the point, panspermia does not resolve the question of how life began. Our alien visitors must have come from somewhere; panspermia does not explain their origin(s). Understanding the origin of alien life is no simpler than understanding the origin of life on Earth. On the other hand, if life is discovered on other planets or the moons in our solar system, its structural and molecular details will be informative – it would transform "astrobiology" into a real scientific discipline.[60]

References

[58] Making human insulin in bacteria

[59] New life for pig-to-human transplants

[60] Top 5 Bets for Extraterrestrial Life in the Solar System

This page titled 2.6: Thinking about life's origins is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.7: Experimental studies on the origins of life

An approach to understanding the origin of life is to generate plausible precursors of living systems. Stanley Miller (1930-2007) and Harold Urey (1893-1981) carried out an early and influential example of this approach.[61]

They made an educated guess as to the composition of Earth's early (before life) atmosphere. They assumed the presence of oceans and lightning. They set up an apparatus to mimic these conditions; they passed electrical sparks through their experimental atmosphere. When analyzed, after a few days they found a complex mix of compounds; included were many of the amino acids found in modern organisms. Similar experiments, using combinations of starting compounds more likely to represent the environment of the early Earth, have produced similar results.[62] Quite complex organic molecules have been detected in interstellar dust clouds and certain types of meteorites. Similarly, the chemistry occurring in deep sea hydrothermal vents can produce complex mixtures of biomolecules abiogenically.[63] Around 4 billion years ago, a time known as the period of heavy bombardment, meteorite impacts with the Earth could have supplied substantial amounts of organic molecules.[64]

Given that the potential building blocks for life were present, the question becomes what conditions and what steps led to the formation of the first living systems? Assuming that these early "pre-LUCA" systems were simpler than modern organisms, they were likely to be molecular communities of chemical reactions able to exchange molecules with the outside world. A selective boundary was necessary to keep the molecules of these systems from diffusing away from one another. One possible model is that such systems were originally tightly associated with the surface of specific minerals and that these mineral surfaces served as catalysts, speeding up important reactions. Over time, these pre-living systems acquired more sophisticated boundary structures (membranes) and were able to exist free of the mineral surface, perhaps taking small pieces of the mineral with them.[65]

The generation of an isolated but open system, something we might term a protocell, was a critical step in the origin of life. Such a system has properties that facilitated its evolution. For example, because of its membrane boundary, changes that occurred within one such structure would not be shared with neighboring systems. Rather, they would accumulate in, and could favor the survival of, the system over its neighbors. Such systems might reproduce by mechanical fragmentation. If changes within a system improved its stability, its ability to accumulate resources, avoid competition, or its ability to survive, grow, and reproduce, that system, and its progeny, would be likely to become more common. As these changes accumulate and were passed from parent to offspring, such systems might evolve into cells and organisms.

As in living systems today, the earliest steps in the formation of the first organisms required a source of energy to establish and maintain the non-equilibrium living (or pre-living) state. There are really only two sources for this energy, light (electromagnetic radiation from the sun) and thermodynamically unstable molecules present in the environment. A number of plausible scenarios for the steps leading to life have been suggested. For example, a recent study based on the analysis of the genes found in modern organisms, and the proteins that they encode, suggests that LUCA arose in association with hydrothermal vents and derived energy from thermodynamically favorable chemical reactions.[66] But whether this reflects LUCA or an ancestor of LUCA that became adapted to living in association with hydrothermal vents is difficult, and perhaps impossible to resolve unambiguously, given that LUCA lived ~3.5 to 4 billion years ago and cannot be studied directly.

References

[61] The Miller-Urey experiment & wikipedia: http://en.wikipedia.org/wiki/Miller–Urey_experiment

[62] A reassessment of prebiotic organic synthesis in neutral planetary atmospheres:

[63] The last universal common ancestor between ancient Earth chemistry and the onset of genetics

[64] A time-line of life's evolution: http://exploringorigins.org/timeline.html

[65] Mineral Surfaces, Geochemical Complexities, and the Origins of Life

[66] Meet LUCA, the Ancestor of All Living Things:

This page titled 2.7: Experimental studies on the origins of life is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.8: Mapping the history of life on earth

Assuming that life arose spontaneously, we can look at the fossil record to better understand how life diversified and it's impact on the Earth. Consider what we know about where the Universe and Earth came from. The current scientific model for the origin of the universe is known as the “Big Bang”, the “primeval atom”, or the “cosmic egg”. This model is based on an idea originally proposed by the priest, physicist, and astronomer Georges Lemaître (1894-1966).[67] The Big Bang model arose from efforts to answer the question where are the fuzzy nebulae (patches of light in the night sky) located? are they within or outside of our galaxy? Answering this question required a way to determine how far these nebulae were from Earth. Edwin Hubble (1889-1953) and his co-workers were the first to provide compelling evidence that nebulae were actually galaxies in their own right, each very much like our own Milky Way galaxy. Each is composed of many billions of stars. This was surprising. It made that the Earth, sitting on the edge of the Milky Way seem less important – a change in cosmological perspective similar to that associated with the idea that the Sun, rather than the Earth, was the center of the solar system.

Hubble and colleagues combined two types of observations to measure the movement of galaxies with respect to the Earth. The first allowed them to estimate the distance from the Earth to a galaxy. The second used Doppler shift. For an object moving toward an observer, the observed wavelength of emitted light will be shortened, that is, shifted toward the blue end of the spectrum. When an object is moving away from the observer the wavelength will be lengthened, shifted to the red end of the spectrum. Based on their observations, Hubble concluded the further a galaxy appears to be from Earth, the greater that shift is toward the red. Galaxies, outside of our local group, were all moving away from one another. Running time backward, he concluded that at one point in the past, all of the matter and energy in the Universe must have been concentrated in a single point.[68] The "Big Bang" model concludes that the Universe is ~13.8 +/- 0.2 billion (10^9) years old. This is a length of time well beyond human comprehension; it is sometimes referred to as deep time. The age of the Earth and the other planets in the solar system has been estimated as ~4.5 to 5 x 10^9 years. You can get some perspective on deep time using the “Here is Today” website.

After its formation, the Earth was bombarded by extraterrestrial materials. This bombardment began to subside around ~3.9 billion years ago, reaching its current level by ~3.5 billion years ago.[69] It is not clear whether life arose multiple times and was repeatedly destroyed early on or if the origin of life was a one-time event, taking hundreds of millions of years before it succeeded, after which it has survived and expanded to the present day.

References

[67] Georges Lemaître: http://www.physicsoftheuniverse.com/scientists_lemaitre.html

[68] The origin of the universe and the primeval atom

[69] The violent environment of the origin of life

This page titled 2.8: Mapping the history of life on earth is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.9: Fossil evidence for the history of life on earth

The earliest period in Earth's history is known as the Hadean, after Hades, the Greek god of the dead. The Hadean is defined as the period from the origin of the Earth to the first appearance of life. Fossils provide our only direct evidence for when life first appeared. Such fossils are found in sedimentary rock; rock formed when fine particles of mud, sand, or dust entomb an organism. Hunters of fossils (paleontologists) do not search for fossils randomly; they use geological information to identify outcroppings of sedimentary rocks of the specific ages they are interested in.[70]

Early in the history of geology, before Darwin and Wallace proposed their theory of evolution, geologists recognized that specific types of fossils were associated with rocks of specific ages. This correlation was so robust that rocks could be accurately dated based on the types of fossils they contained. Given the claims by theology-based young earth creationists that Earth is less than ~10,000 years old, it is worth reflecting on the interconnectedness of the sciences. Geologists do not rely solely on fossils to date rocks; many types of rocks do not contain fossils. Rocks can be dated based on the physics of isotope stability and the chemistry of atomic interactions. It uses the radioactive decay of elements with isotopes with long half-lives, such as ^{235}Ur (uranium), which decays into ^{207}Pb (lead), with a half-life of ~704 million years and ^{238}Ur which decays into ^{206}Pb with a half-life of ~4.47 billion years. Since these two Pb isotopes appear to be formed only through the decay of uranium, the ratios of uranium and lead isotopes can be used to estimate the age of a rock. In order to use isotope abundance to accurately date rocks, it is critical that all of the atoms in a mineral sample originated there and stayed there; none were washed into or out of the rock. Since uranium and lead have different chemical properties, this can be difficult to establish in some mineral types. That said, with care, and using rocks that contain chemically inert minerals, like zircons, the isotope ratio method can estimate the age of rocks to an accuracy of ~1% or better. Such age estimates support James Hutton's (1726-1797) dictum that the Earth is ancient, with "no vestige of a beginning, no prospect of an end."[71] We know now, however, that this statement is not true; while very old, the Earth had a beginning, it coalesced around ~4.54 billion years ago, and it will disappear when the sun expands and engulfs it in about ~5.5 billion years from now.[72]

Back to fossils. There are many types of fossils. Chemical fossils are molecules that, as far as we know, are naturally produced only through biological processes.[73] Their presence in an ancient rock implies that organisms were present at the time the rock formed. Chemical fossils first appear in rocks that are between ~3.8 to ~3.5 x 10^9 years old. What makes chemical fossils problematic is that there may be non-biological but currently undiscovered or unrecognized mechanisms that could produce them, so we need to be cautious in our conclusions.

Moving from the molecular to the physical, there are what are known as trace fossils. These can be subtle or obvious. Leaves and immotile organisms can settle on mud or sand and leave impressions. Burrowing and slithering animals make tunnels or disrupt surface layers. Walking animals can leave footprints in sand, mud, or ash. How does this occur? If the ground is covered, compressed, and converted to rock, these various impressions can become fossils. Later erosion can then reveal them. For example, if you live near Morrison, Colorado you can visit the rock outcrop known as Dinosaur Ridge and see fossil dinosaur footprints; there may be similar examples where you live!

We can learn a lot from trace fossils, they can reveal the general shape of an organism, its ability to move, or to move in a particular way. To move, an organism must have some kind of muscle or alternative mobility system and probably a nervous system that can integrate internal and external information and produce coordinated movements. Movement also suggests that the organisms that made the trace had something like a head and a tail. Tunneling organisms are likely to have had a mouth to ingest sediment, much like today's earthworms - they were predators, eating the microbes they found in mud.

In addition to trace fossils, there are also the type of fossils that most people think about, which are known as structural fossils; the mineralized remains of the hard parts of organisms such as teeth, scales, shells, or bones. As organisms developed hard parts fossilization, particularly of organisms living in environments where they could be buried before being dismembered and destroyed by predators, scavengers, or microbes, became more likely.

Unfortunately for us (as scientists), many and perhaps most types of organisms leave no trace when they die. In part this may be because they live in places where fossilization is unlikely. Animals that live in woodlands, for example, rarely leave fossils. The absence of fossils for a particular type of organism does not imply that these types of organisms do not have a long history, but rather that the conditions where they lived and died was not conducive to fossilization. Many types of living organisms have no

fossil record at all, even though, as we will see, there is molecular evidence that they arose tens to hundreds of millions of years ago.

References

[70] A process described in some detail by Neil Shubin in The Evolution of Limbs from Fins

[71] Changing Views of the History of the Earth

[72] How the sun will die

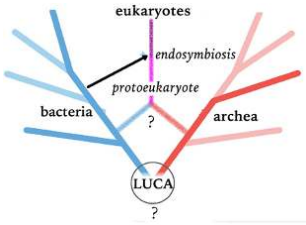
[73] Although as Wohler pointed out, they can be generated in the laboratory.

This page titled 2.9: Fossil evidence for the history of life on earth is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

2.10: Life's impact on earth

Today, there are three distinct, although related families of organisms: the bacteria, the archaea, and the eukaryotes. While the exact nature of LUCA is unclear, it is likely that it was single celled and relatively simple in general organization (→) consisting of a boundary membrane that controlled the movement of molecules into and out of the cell, a cytoplasm in which various biosynthetic reactions took place, and molecules of the genetic material, DNA, located within the cytoplasm. Both bacteria and archaea share this type of cellular organization and basic molecular mechanisms, although they differ in a number of molecular details.[74]

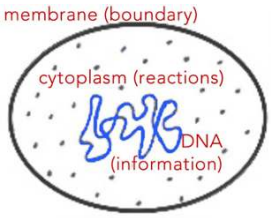
Eukaryotes are structurally more complex; they contain internal membrane systems. Their genetic material is located within a double membrane-bounded compartment, the nucleus, located within the cytoplasm. Movement between nuclear interior and cytoplasm is facilitated by complex molecular machines, known as nuclear pores. How the nucleus came to be remains, not surprisingly, unclear. We do not have direct evidence, just informed guesses. What is now clear, however, is that the formation of eukaryotes involved a symbiotic event in which a α-proteobacterium (a type of bacteria) was engulfed by a proto-, pre-nuclear eukaryote (→).[75] This "endogenous bacterium" was not digested (and destroyed) but became the eukaryotic mitochondrion. Essentially all eukaryotes (the protozoa, fungi, animals, and plants) have mitochondria, descended from this event. Later on in the history of life, a second endosymbiotic event occurred in which a mitochondria-containing eukaryote engulfed but did not digest a second type of bacteria, a photosynthetic cyanobacterium-like organism, leading to the algae and the plants.

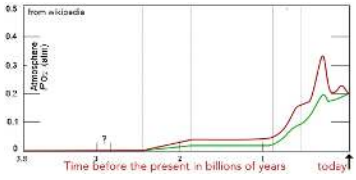


The first "life-derived pollutant" molecular oxygen

While the earliest organisms likely used energy released in the course of thermodynamically favorable chemical reactions to maintain their structural integrity and to grow, relatively soon bacterial-type organisms appeared that could capture the energy in light and use it to drive various thermodynamically unfavorable chemical reactions. A major class of such reactions involves combining CO_2 (carbon dioxide), H_2O (water), and other small molecules to form carbohydrates (sugars) and biologically important molecules, such as lipids, amino acids, and nucleotides (from which proteins, and nucleic acids are built). At some point light-eating organisms began to release molecular oxygen (O_2), a waste product of the oxygenic photosynthesis process. These oxygen-releasing organisms became so numerous that they began to change Earth's surface chemistry - they represent the first life-driven ecological catastrophe or opportunity, depending on your perspective.

The level of O_2 in the atmosphere represents a balance between its production, primarily through oxygenic photosynthesis, and its breakdown through various chemical reactions. As O_2 first appeared, it reacted with iron to form deposits of water-insoluble Fe(III) oxide (Fe_2O_3) – that is, rust. This rust reaction removed large amounts of O_2 from the atmosphere, keeping levels of free O_2 low for a long time (↓). The rusting of iron in the oceans is thought to be responsible for the massive banded iron deposits found around the world.[76] O_2 also reacts with organic matter, as in the burning of wood. When large amounts of organic matter are buried before they can react with O_2, as occurs with the formation of coal, more O_2 accumulates in the atmosphere. Although O_2 was probably being generated and released earlier, by ~2 billion years ago, atmospheric O_2 had appeared in detectable amounts and by ~850 million years ago O_2 had risen to significant levels (←). Atmospheric O_2 levels have changed significantly over time since then, based on the relative rates of its synthesis and breakdown. Around ~300 million years ago, atmospheric O_2 levels reached ~35%, almost twice the current level. It has been suggested that these high levels of atmospheric O_2 made the evolution of giant insects possible.[77]





Although we tend to think of O_2 as a natural and benign substance, it is in fact highly reactive and potentially toxic; its production and accumulation posed serious challenges and unique opportunities to organisms. O_2 can be "detoxified" through reactions that lead to the formation of water; this type of thermodynamically favorable reaction has been co-opted for a wide range of biological purposes. For example, through coupled reactions O_2 can be used to capture the maximum amount of energy from the breakdown of complex molecules (food), leading to the generation of CO_2 and H_2O, both of which are stable.

Around the time that O_2 levels were first rising, that is ~10^9 years ago, the first trace fossil burrows appeared in the fossil record.



These were likely to have been produced by simple worm-like, macroscopic (that is, large enough to see without magnification) multicellular organisms. These were animals capable of moving along and through the mud on the ocean floor. Around a billion years ago, new and more complex structural fossils (←) began to appear.[78] The first of these are the Ediacaran organisms, named after the geological formation in which their fossils were first found.[79] Current hypotheses suggest they were immotile, like modern sponges but flatter; it remains unclear how or if they are related to later animals.

Since the fossil record does not contain all organisms, we are left to speculate on what earlier metazoans looked like. By the beginning of the Cambrian age (~545 x 10^6 years ago), a wide variety of organisms had appeared in the fossil record, many clearly related to modern animals. Molecular level data suggests that their ancestors originated more than ~30 million years earlier. These Cambrian organisms show a range of body types. Most significantly, many were armored. Building armor involves capturing and expending the energy needed to synthesize these structures. The presence of armor suggests the presence of predators, and a need for a defensive response.

This page titled 2.10: Life's impact on earth is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

3: Evolutionary Mechanisms and the Diversity of Life

3.0: Introduction
3.1: Organizing organisms (hierarchically)
3.2: Natural and un-natural groups
3.3: Fossils and family relationships- introducing cladistics
3.4: Evolution theory’s core concepts
3.5: Limits on populations
3.6: Darwin and Wallace's conceptual leap
3.7: Mutations and the origins of genotype-based variation
3.8: The origins of polymorphisms
3.9: Genotype-phenotype relationships- discrete and continuous traits
3.10: Variation, selection, and speciation
3.11: Types of simple selection
3.12: Considering stochastic processes
3.13: Genetic diversity, population size, founder effects and population bottlenecks
3.14: Population bottlenecks
3.15: Genetic drift
3.16: Gene linkage- one more complication
3.17: A brief reflection on the complexity of phenotypic traits
3.18: Speciation and extinction
3.19: Mechanisms of speciation
3.20: Mechanisms of reproductive isolation
3.21: Sympatric speciation
3.22: Signs of evolution- homology and convergence
3.23: Homologies provide evidence for a common ancestor
3.24: Anti-evolution arguments

This page titled 3: Evolutionary Mechanisms and the Diversity of Life is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.0: Introduction

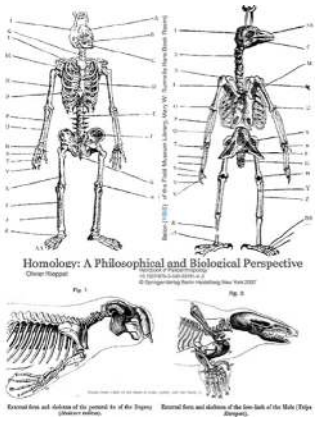


In which we consider the exuberant diversity of organisms and how they came to be. To understand these processes requires an introduction to core evolutionary mechanisms, both adaptive (natural, sexual, and social selection) and non-adaptive (genetic drift and bottlenecks). As part of our discussion we consider the history of how people considered the diversity and meaning of life.

In medieval Europe there was a tradition of books known as bestiaries, illustrated catalogs of real and imagined organisms. Often each organism was associated with a moral lesson. "*Male lions were seen as worthy reflections of God the Father ... while the dragon was understood as a representative of Satan.*"[82] Bestiaries are an example of a natural theology, an attempt to gain an understanding of the supernatural through the study of natural objects.[83] The presumption was that each type of organism was created for a particular purpose, and that this purpose was to provide a moral lesson. Natural theology grew problematic as more and more different types of organisms were discovered, many with no obvious significance to humans. Scientists have identified approximately ~1,500,000 different types of plants, animals, and microbes. The actual number of different types of organisms (species) may be much higher.[84] These numbers refer to species that currently exist, but as we know from the fossil record many species that once existed are now extinct. So why are there so many different types of organisms?[85] Given how different types of organisms look and behave, it seemed possible that trees, mushrooms, spiders, whales, and humans represent distinct lineages and separate creation events.



As the number and diversity of organisms was recognized, a number of observations undermined the assumption that organisms were created to serve or instruct humanity. The first was the fact that many organisms had little obvious relevance to humans. While obvious in the case of extinct organisms, this extended to newly discovered (by Europeans) organisms; panda bears, potatoes, and maize (corn). At the same time students of nature, known as naturalists, discovered different types of upsetting and "cruel" behaviors in the natural world. Consider the fungus *Ophiocordyceps unilateralis;* it infects the ant *Camponotus leonardi* and takes control of the ant's behavior, causing it to migrate to environments that favor fungal growth before killing the infected ant. Similarly, the nematode worm *Myrmeconema neotropicum* infects the ant *Cephalotes atratus*, leading to dramatic changes in the infected ant's morphology and behavior. The infected ant's abdomen turns red and is held raised up, making it resemble a fruit and increases the likelihood that it will be eaten by a bird (→). The bird then transports the worms, which survive in its digestive system; excreted worms can be eaten by, and so infect new ants thus completing the worm's life cycle.[86] Perhaps the most famous example of "natural cruelty" involves wasps of the family *Ichneumonidae*. Female wasps deposit fertilized eggs into the bodies of various types of caterpillars. The wasp's eggs hatch and produce larvae that feed on the living caterpillar, consuming it from the inside out. Charles Darwin, in a letter to the American naturalist Asa Gray, remarked "*There seems to me too much misery in the world. I cannot persuade myself that a beneficent and omnipotent God would have designedly created the Ichneumonidae with the express intention of their feeding within the living bodies of caterpillars, or that a cat should play with mice*." Rather than presume that a supernatural creator was responsible for such "cruel" behaviors, Darwin and others sought alternative, morally neutral "natural" processes that could generate biological diversity and explain biological behaviors.



As the amazing diversity of organisms became apparent and difficult to ignore, another broad and inescapable conclusion emerged. Different organisms displayed detailed structural similarities. Comparative studies revealed similarities between different types of organisms. A classic work, published in 1555, compared the skeletons of a human and a bird, both vertebrates.[87] While many bones have different shapes and

relative sizes, many bones were at least superficially similar to one another (top →). In another example, the skeleton of the dugong, a large aquatic mammal, appears similar to that of the European mole (bottom →), a small terrestrial mammal that tunnels underground. There are general skeletal similarities between all vertebrates. The closer we look, the more similarities we find. These similarities run deeper than the anatomical, they extend to the cellular and molecular levels as well and apply to both vertebrates and invertebrates. So the question remains, what explains these similarities? Why build an organism that walks, runs, and climbs, such as a human, with a skeleton similar to that of a organism that flies (birds), swims (dugongs), or tunnels (moles). Are these anatomical similarities just flukes or do they imply something deeper about how organisms were formed?

References

[82] Northumberland Bestiary And as a general note, we focus on the European scientific tradition here, but others are similar.

[83] What Is Natural Theology?

[84] How many species are there on Earth and in the ocean?

[85] As a technical point, which we will return to, we will refer to each distinct type of organism as a species.

[86] The Life of a Dead Ant: The Expression of an Adaptive Extended Phenotype

[87] Belon (1555) L'Histoire de la Nature des Oyseaux. Paris, Guillaume Cavellat

This page titled 3.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.1: Organizing organisms (hierarchically)

Carl Linnaeus (1707-1778) was a pioneer in taking the similarities between different types of organisms seriously. Based on similarities and differences he developed a system to classify organisms in a coherent and hierarchical manner. Each organism had a unique place in this scheme, a unique set of "coordinates".[88] What was, and occasionally still is, the controversial aspect of such a classification system is in how to decide which traits are considered significant and which superficial or unimportant, at least for the purposes of classification. Linnaeus proposed no explanation for why organisms could be classified in such a hierarchical manner.

This is a good place to consider again the importance of guesses, hypotheses, models, and theories in biology, and science in general. Linnaeus noticed apparent similarities between organisms; he used it to generate his classification scheme, but he had no coherent explanation for why such similarities existed. Newton's law of gravitation explains how objects (planets and apples) behave but not why. So what are the features of a scientific, predictive model? Such models have to suggest observations or predict outcomes that have yet to be observed. It is the validity of these predictions that enables us to identify useful models. A model that fails to make empirically testable predictions is not useful scientifically, although it may be psychologically comforting. Linnaeus's scheme was descriptive. The value of scientific models, even when proven wrong, is that they enable us to refine, or force us to abandon, the model–something quite different from theological or ideological models. As a scientific model expands what it explains, and its predictions are confirmed experimentally, the model becomes a theory, while "competing" models are abandoned. We assume that the model reflects the way the world works. This enables us to distinguish between a law and a theory. A law describes what we see but not why we see it. A theory provides the explanation for why the law exists.[89]

The Linnaean classification system placed organisms of a particular type together into a species. Species were then grouped into larger and larger groups. This raises questions such as, how different do two organisms have to be in order to fall into different species? How do we make such a decision? If we look at organisms that appear similar, do we place larger individuals (of the same age) into a different species than smaller ones? The situation is complex when we think about reproduction. Some organisms can produce offspring by themselves; they can be asexual or self-fertilizing, often called hermaphroditic. Other types of organisms are sexual, individuals need to cooperate with another of a different "type" to produce offspring. Organisms, such as yeasts, have multiple "mating types". In most multicellular organisms there are two distinct mating types, or sexes - male and female. Often, but not always, individuals of the two sexes appear different, often dramatically so - a situation known as sexual dimorphism.[90] Different sexes of the same type of organisms, at different developmental stages, or growing under different conditions can differ in their appearance and/or behavior (have different phenotypes). It requires careful study to recognize a particular type of organism.

What originally counted as an organism of a particular species was based on Linnaeus's or other naturalists' judgement. It depended on which particular traits were assumed to be significant and useful to distinguish organisms of one species from those of another, perhaps quite similar appearing species. The choice of these key traits is subject to debate. Naturalists can argue whether similar organisms should be split into two or more species or reclassified into a single species.




The individuals that make up a species are not identical but share key traits. As noted there are sometimes dramatic differences between males and females of the same species (→ left ♂ & right ♀ spiders and ducks). These differences can be so dramatic that without further evidence, it can be difficult to tell whether two animals are members of the same or different species. In this light the primary criteria for determining whether sexually reproducing organisms are members of the same or different species is whether they can and do successfully interbreed with one another in the wild. Reproductive compatibility is not relevant with asexual species, such as most microbes. An asexual organism is essentially a clone and species distinctions are based on other criteria. Within a species, there can be regional differences that are distinct enough to be recognizable. Where this is the case, these groups are known as populations or subspecies.[91] While distinguishable, the organisms in these groups retain the ability to interbreed and so are considered members of a single species. As an example tigers are *Panthera tigris*, while Siberian tigers are known as *Panthera tigris sumatrae*; sumatrae is the subpecies name. Sometimes, reproductive barriers between species break down; this appears to have occurred between modern humans (*Homo sapiens*) and Neanderthals (*Homo neanderthalensis*).[92]



HILLIS CIRCLE OF LIFE TATTOO

Linnaeus grouped species that displayed similar traits into more inclusive groups, known as genera. Each organism has a unique Linnaean name, its genus and species names written in italics with the genus name capitalized. Following on this pattern, organisms can be placed into larger, more inclusive groups which we will ignore here. The end result is the rather surprising conclusion that all organisms fall into a small number of "supergroups" known as phyla. Even more surprising, perhaps, is that all organisms can be placed into a single phylogenetic tree (or circle)(→); they are all connected.[93] That this should be the case is by no means obvious. There could be multiple, disconnected groups, but no – they form a single connected network composed of discrete but related groups. Considering both the fossil record and living organisms, there appears to be an intrinsic continuity between all forms of life on Earth.

References

[88] Each organism can be identified by a species, within a genus, within a family, within an order, within a class, within a phylum, within a Kingdom - but please to not memorize this as it is not important us..

[89] If we go back, Newton's law of gravity explained how objects behaved gravitationally, but it not why. In contrast, Einstein's theory of general relativity explained why there was gravity, and predicted behaviors that were not predicted by Newton's law.

[90] Sexual dimorphism & sexual dimorphism in spiders & unexpected sexual dimorphisms

[91] The term race, a social construct, as no real value in biology: see Taking race out of human genetics

[92] Consider the latest results: Nature - Neanderthal–human baby-making was recent — and brief

[93] David Hillis. Tree of LIfe Graphics page

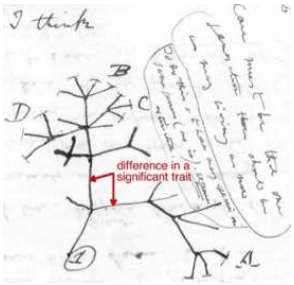
This page titled 3.1: Organizing organisms (hierarchically) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.2: Natural and un-natural groups

Species can be seen as natural groups. Genera and higher-level classification groups are based on decisions about what traits are "matter". The genus *Canis* includes wolves and coyotes while the genus *Vulpes* includes foxes. The distinction between these two groups is based on skull shape and size. Now consider the genus *Felis*, the common house cat and the genus *Panthera*, which includes tigers, lions, jaguars, and leopards. These are distinguished by both cranial features and the fact that *Panthera* but not *Felix*, can roar. So what do we make of these distinctions? are they sufficient to justify distinct groups or should *Canis* and *Vulpes* (and *Felix* and *Panthera*) be merged together into a single genus? Such a question recognizes the fact that genus and higher levels of classifications can be changed. A species originally placed in one genus be moved to another, or placed in its own genus. Consider the eight types of bear-like animals; they are placed into five genera.[94] The genus *Ursus* contained the brown bear (*Ursus arctos*), the Asiatic black bear (*Ursus thibetanus*), the American bear (*Ursus americanus*), and the polar bear (*Ursus maritimus*). The other bear-like organisms, the spectacled bear (*Tremarctos ornatus*), the sloth bear (*Melurus ursinus*), the sun bear (*Helarctos mayalanus*), and the giant panda (*Ailuropoda melanoleuca*) are each placed in their own distinct genus. Why? becausescientists consider them "significantly" different from members of the genera Ursus, *and each other.* The question of how such differences come warrant the generation of a new genus is something for a subsequent course.

Evolution: why Linnaean classification makes sense

So where are we? Given the idea that life comes from pre-existing life a logical assumption is that the more similar two species are,



the more traits they share at the macroscopic and molecular levels, the more likely they are derived from a common ancestral population. Traits used to identify the descendants of ancestral populations are known as synapomorphies. The various species derived from that ancestral population share, or once had and then lost that synapomorphy. Other populations, those that give rise to distinct groups of species, lack that synapomorphy. The presence or absence of a particular synapomorphy defines a branch point in the family tree. Organisms on a branch represent an evolutionary lineage and form a “natural” group, more closely related to one another than to organisms on the other branches (→). Different synapomorphies define different levels of relationship. A obvious question then is, how do we determine ancestry when the ancestors lived, thousands, millions, or billions of years in the past. Since we cannot travel back in time, we have to deduce relationships from comparative studies of living and fossilized organisms. Here the biologist Willi Hennig (1913-1976) played a key role.[95] He established rules for using shared, measurable traits to reconstruct ancestral relationships, such that each group should have a single common ancestor, or more accurately, a single ancestral population. As we will discover, one of the traits commonly used in modern studies are gene (DNA) sequences and genomic organization, although given the time scales involved ambiguities can remain.

References

[94] http://en.wikipedia.org/wiki/List_of_bears

[95] A description of Willi Hennig’s impact on taxonomy

This page titled 3.2: Natural and un-natural groups is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.3: Fossils and family relationships- introducing cladistics

Fossils provide evidence for the existence, in the past, of both familiar and previously unknown "types" of organisms. An important point to keep in mind is that organisms differ in their fossilizability, which depends upon the environments in which they lived. Most types of organisms have left no fossils whatsoever.[96] That said, fossils provide critical evidence for understanding the history of life. An obvious question, answered through the study of fossils, is whether Ediacaran organisms[97], Burgess shale creatures (→), and dinosaurs fit into the same classification scheme as living organisms or do they form separate family trees? Do they provide evidence for multiple independent origins of life. The result of many detailed studies lead to the conclusion that essentially all fossilized organisms fall into single classification scheme. A classic example are dinosaurs which are clearly related to a specific type of reptile and gave rise to modern birds, while mammals are more closely related to a second, now extinct, group of "mammal-like reptiles." If we had samples of Ediacaran organisms we could quickly (through DNA analyses) resolve how they are related to modern organisms.[98] It has been possible to extract and analyze DNA from the bones of Neanderthals and Denisovian-type humanoids; both species went extinct ~30,000 years ago. Analysis of this DNA has clarified the relationship between Neanderthals, Denisovians, and modern humans [99] providing compelling evidence for inbreeding between these groups. The result is to reclassify Neanderthals and Denisovians as subspecies of *Homo sapiens*.[100]

Questions to answer:

14. Explain why extinct species fit into the same classification scheme as used for living (observable) organisms.
15. What factors would influence your decision on whether a trait found in two different organisms was present in their common ancestor?

Questions to ponder:

- You discover life on a planet orbiting another star in another galaxy; would you expect such organisms to fit into the Linnaean classification system?
- Would sex with a Neanderthal be immoral?

The theory of evolution and the organization of life

Why is it that pterosaurs, birds, whales, and humans share common features, such as the organization of their skeletons, similarities that lead them to be classified together as vertebrates? Why are there extinct organisms, known only from their fossils, but which nevertheless share many common features with living organisms? And most importantly, why are there so many different types of organisms? Charles Darwin (1809-1882) and Alfred Wallace (1823-1913) proposed a model, described in detail in Darwin's book "*On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*" originally published in 1858, and more succinctly by Wallace, that answered these and a number of other questions.

> *The main unifying idea in biology in Darwin's theory of evolution through natural selection*
>
> *- John Maynard Smith*

Evolutionary theory is based on observations of the natural world and their logical implications. Organisms are similar because they are related to one another – they share common ancestors.[101] We can infer that the more characters two species share the more recently they shared a common ancestor. We can make plausible and testable deductions about what those common ancestors looked like. As an example, we predict that the common ancestor of all terrestrial vertebrates resembles a fish with leg-like limbs and we can predict the number and shape of the bones found in those limbs.[102] Scientists have discovered fossils of such an organism (← *Tiktaalik roseae*).[103]



Tiktaalik roseae, an extinct organism that lived ~375 million years ago, is likely to be similar to the common ancestor of all terrestrial vertebrates (from “Your inner fish” by Neil Shubin).

The discovery of *Tiktaalik* and other "missing" links is an example of the fact,well before the molecular, cellular, and developmental mechanisms were uncovered evolutionary theory "made sense" of what was observed, made testable predictions about what would be found, and where to look, and has been supported by what has, in fact, been found. In the case of particularly fast growing organisms, and very strong selection pressures, such as the presence of an antibiotic, we can observe evolutionary processes taking place over the course of days, weeks, and months – that is, in real time.[104]

References

[96] The incompleteness of the fossil record & Absolute measures of the completeness of the fossil record

[97] Ediacaran organisms

[98] On the eve of animal radiation: phylogeny, ecology and evolution of the Ediacara biota

[99] :Paleogenomics of archaic hominins

[100] Humans mated with Neandertals much earlier and more frequently than thought & The downside of sex with Neanderthals

[101] As we will discover, there are organisms can appear similar that are not closely related; this is due to what is known as convergent evolution. That said, such organisms share a common ancestor, although it existed further back in time.

[102] Your inner fish video

[103] Meet Tiktaalik roseae: An Extraordinary Fossil Fish A similar situation applies to the terrestrial ancestors of whales

[104] Visualizing evolution as it happens see also Phagotrophy by a flagellate selects of colonial prey: a possible origin of multicellularity - Boraas et al 1998

This page titled 3.3: Fossils and family relationships- introducing cladistics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.4: Evolution theory's core concepts

On what facts and inferences is the Theory of Evolution based? Two foundational observations are deeply interrelated and based on the observations of plant and animal breeders and the characteristics of natural populations. The first is the fact that whatever type of organism we examine, if we look carefully enough, making accurate measurements of visible and behavioral traits, which is known as the organisms phenotype, we find that individuals vary with respect to one another. Plant and animal breeders recognized that the offspring of controlled matings often displayed phenotypes similar to those of their parents, indicating that the (invisible) factors responsible for phenotypic (observable) traits are inheritable. Animal and plant breeders used what is known as artificial selection to generate a range of domesticated plants and animals with specific phenotypes. For example, beginning ~10,000 years ago plant breeders in Mesoamerica developed modern corn (maize) by the selective breeding of variants of the grass teosinte (→).[105] Evidence supports the idea that all of the various breeds of dogs, from the tiny to the rather gigantic (←), were derived from common ancestors that lived around 30,000 years ago.[106] Human controlled breeding has been used to shape all of the different types of domesticated organisms.





In artificial, that is, human-driven selection, organisms with desirable (or desired) traits are selected for breeding with one another. Organisms without these traits are not permitted to breed. This process of artificial selection, carried out over many generations, produces organisms that display the desired form of the selected trait. A key point, this strategy only works because desired versions of the trait are present in the original population and are due, at least inpart, to genetic, that is stable and heritable factors. The implication is that different organisms differ with regards to these factors, but what these differences entailed was not known, and in fact did not matter in terms of results.



Traditional plant and animal breeders recognized that offspring tended to display traits similar to those of their parents. A classic example is the Hapsburg lip (→), a trait passed through this European ruling family for generations.[107] An important point is that artificial selection can produce quite exaggerated and, in the absence of human intervention, highly deleterious traits. Just look at domesticated chickens and turkeys, which, while not completely flightless, can fly only short distances and so are extremely vulnerable to predators. Neither modern corn (*Zea mays*) or chihuahuas, one of the smallest breeds of dog, developed by Mesoamerican breeders, would be expected to survive for long on their own in the wild.[108]

References

[105] Molecular Evidence and the Evolution of Maize

[106] From wild animals to domestic pets, an evolutionary view of domestication

[107] 'Imperial Stigmata!' The Habsburg Lip, A Grotesque 'Mark' Of Royalty Through The Centuries!: & Genes and Queens

[108] How DNA sequence divides chihuahua and great dane

This page titled 3.4: Evolution theory's core concepts is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

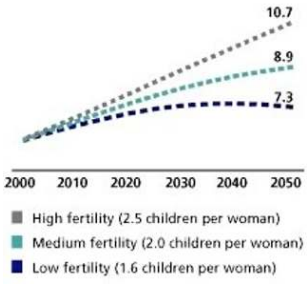
3.5: Limits on populations

It is a demonstrable fact that all types of organisms (as opposed to specific individuals) can produce more, sometimes many more than one copy of themselves. Consider, as an example, a breeding pair of elephants or a single asexually reproducing bacterium. Assume that there are no physical limits to their reproduction, that is, that once born, the offspring reproduce in the same way over the course of their lifespans. Allowed to continue unchecked, within a few thousand years a single pair of elephants could produce a mass of elephants larger than the mass of the Earth, an absurd, impossible outcome. We must have left something out of our calculations! As another example, let us consider a solitary, asexual photosynthetic bacterium that needs no mate to reproduce. A bacterium is much smaller than an elephant but it can produce a new bacterium at a much faster rate. Under optimal conditions our bacterium might divide once every ~20 minutes, or even faster, and would, within a few days produce a mass of bacteria greater than that of Earth as a whole. Again, we are clearly leaving something out from our thinking.

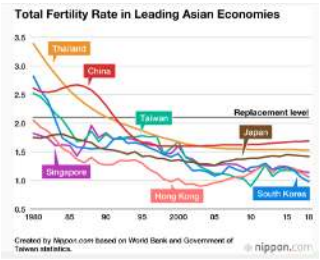
> *A single cell of the bacterium E. coli would, under ideal circumstances, divide every twenty minutes. That is not particularly disturbing until you think about it, but the fact is that bacteria multiply geometrically: one becomes two, two become four, four become eight, and so on. In this way it can be shown that in a single day, one cell of E. coli could produce a super-colony equal in size and weight to the entire planet Earth.… A single cell of the bacterium E. coli would, under ideal circumstances, divide every twenty minutes. That is not particularly disturbing until you think about it, but the fact is that bacteria multiply geometrically: one becomes two, two become four, four become eight, and so on. In this way it can be shown that in a single day, one cell of E. coli could produce a super-colony equal in size and weight to the entire planet Earth.*
>
> *- Michael Crichton (1969) The Andromeda Strain*

The ability to produce more, sometimes many more offspring than are needed to replace themselves, is a trait known as superfecundity. But unlimited growth cannot continue for very long. So what factors act to constrain superfecundity, what limits population growth? One obvious factor involves the resources needed for growth, which are limited. Thomas Malthus (1766-1834) articulated the role of limited resources as a constraint on population. His was a purely logical argument. Competition between increasing numbers of organisms for a limited supply of resources necessarily limits the number of organisms. Malthus painted a gloomy picture of organisms struggling with one another for access to these resources, with many living in a version of extreme poverty, starving to death because they could not out-compete others for the food or spaces they needed to survive and reproduce. One point that Malthus ignored, or more likely was ignorant of, is that organisms rarely behave in this way. It is common to find behaviors that limit direct struggles between organisms for resources. In some organisms, an adult has to establish, and defend, a territory before it can successfully reproduce.[109] The end result of these types of behaviors is to stabilize the population around a steady state level, which is a function of both environmental and behavioral constraints.

An organism's environment includes all factors that influence the organism. Environmental factors include changes in climate, as well as changes in the presence or absence of other organisms. For example, if one organism depends in important ways upon another, the extinction of the first will necessarily influence the survival of the second.[110] Similarly, the introduction of a new type of organism or a new trait, such as O_2-generating photosynthesis, into an established environment may disrupt existing interactions and conditions. When the environment changes, existing steady state population levels may be unsustainable or some of the different types of organisms present may not be viable. If the climate gets drier or wetter, colder or hotter, if yearly temperatures reach greater extremes, or if new organisms, including for example, disease-causing pathogens, enter an area, the average population density of particular organisms may change or in some cases, if the environmental change is drastic enough, they could drop to zero, in other words populations could go extinct. Changing environmental conditions will influence the sustainable steady-state population level of an organism (something to think about in the context of global warming and the destruction or disruption of natural environments).



An obvious example of this type of behavior involves humans (→). Once constrained by disease, war, and periodic famine, the



introduction of better public health and sanitation measures such as clean water, a more secure food supply, and vaccination, have led to reductions in infant mortality that resulted in explosive growth in the population. Now, in many countries, populations appear to be heading to a new steady state level, although exactly what that final population total level will be is unclear.[111] In a number of countries, the birth rate has fallen into the low fertility domain (→), although that is no guarantee that it will stay there![112] In this low fertility domain (ignoring immigration), a country's population decreases over time, since the number of children born is less than the number of people dying. This can generate social stresses.[113] Decreases in birth rate per woman correlate with socioeconomic factors, including educational level and the reproductive self-determination, that is, the emancipation of women. Where women have the right to control their reproductive behavior, the birth rate tends to be lower. Clearly changes in the environment, and here we include the sociopolitical environment, can dramatically influence behavior and impact reproductive rates and population levels.

References

[109] Territorial Defense, Territory Size, and Population Regulation

[110] Why the Avocado Should Have Gone the Way of the Dodo & Neotropical Anachronisms: The Fruits the Gomphotheres Ate

[111] Global population growth & The Joy of Stats

[112] Hans Rosling: Religions and babies

[113] Global fertility has collapsed, with profound economic consequences - from the Economist - 2023

This page titled 3.5: Limits on populations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.6: Darwin and Wallace's conceptual leap

Charles Darwin and Alfred Wallace recognized the implications and significance of these key biological facts: the hereditable nature of variation between organisms, the ability of organisms to reproduce many more offspring than are needed to replace themselves, and the various constraints on population. Based on these facts, they drew a logical implication, namely that individuals would differ in their reproductive success – that is, the number of viable descendants. We can then expect that phenotypic variations associated with greater reproductive success, and the genotypes underlying these phenotypic differences, in a population will increase in frequency over time. Darwin termed this process natural selection, in analogy to the artificial selection practiced by plant and animal breeders. Natural selection is one of the major, but not the only driver of biological evolution.

To be clear reproductive success is more subtle than the phrase "survival of the fittest" implies. First and foremost surviving alone does not matter much if the organism fails to produce fertile offspring. An organism's impact on future generations depend not on how long it lives but on how many fertile offspring it generates. An organism that can produce many reproductively successful offspring at an early age will have more of an impact on subsequent generations than an organism that lives an extremely long time but has few offspring. Again, there is a subtle point here. It is not simply the number of offspring that matter but the relative number of reproductively successful offspring produced.

We can identify a number of factors that influence reproductive success. For example, organisms that reproduce sexually need to find a mate and to deal with the stresses associated with reproduction. These include the ability to obtain adequate nutrition and to avoid premature death from predators and pathogens. Similarly, organisms can cooperate (help) each other, and through cooperation increase the odds that their offspring will survive, compared to solitary organisms. Both individual and social traits are part of the organism's phenotype, which is what natural selection acts on. It is also the case that organisms can have impacts on their environment that can influence selection. Another point that we will return to is that traits can be interdependent. The mechanisms (and genotype) involved in producing one trait may influence others. There are also non-genetic sources of variation. For example, molecular fluctuations occur at the cellular level; these can lead genotypically identical cells that display different behaviors, different phenotypes.[114] Environmental factors and stresses can also influence the growth, health, and behavior of organisms. These are generally termed physiological adaptations. While an organism's genotype influences how it responds to environmental factors, the responses that emerge can be complex.[115]

References

[114] Something that has been studied in nine-banded armadillos that produce "identical" quadruplets.

[115] The global influence of genome on traits: An Expanded View of Complex Traits: From Polygenic to Omnigenic

This page titled 3.6: Darwin and Wallace's conceptual leap is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.7: Mutations and the origins of genotype-based variation

Where does inheritable variation come from? How do genotypes change? As a simple analogy, we can think of the genotype as the book that contains the instructions to build the tools and parts needed to form and maintain a specific, individual cell/organism. This book, known as the genome. is complex - it contains both instructions but also codes that determine when particular instructions should be read. The complexity comes from how the genome "works". In humans, this involves DNA molecules that taken together are approximately 3.2 billion base pairs long. Within these DNA molecules there are regions that appear to be meaningless (but may not be) and those that clearly encode specific instructions. To continue our analogy, a few changes to the DNA sequence may have no effect on meaning, while others may dramatically change the narrative or make no sense at all. Do not worry, much more on DNA later.

At this point we will define the meaningful regions, the words and sentences, as corresponding to genes and the other sequences as intragenic regions, that is, spaces between and within genes. It has been estimated that humans have ~25,000 genes. As we continue to learn more molecular biology, our understanding of genes and intragenic regions has become more nuanced. Regions that originally appeared meaningless have been found to have meaning. Many regions of the genome are unique, they occur only once within the string of characters. Others are repeated, sometimes hundreds to thousands of times. When we compare the genotypes of individuals of the same type of organism, we find that they differ at a number of places. For example, over ~55,000,000 variations have been found between all human genomes examined to date, and more are likely to be identified. These genotypic differences are known as polymorphisms, from the Latin meaning multiple forms. Polymorphisms are the basis for DNA-based forensic identification tests. One thing to note, however, is that only a small number of these variations are present within any one individual, and considering the size of the human genome, most people differ from one another at less than 1 to 4 letters out of every 1000. That amounts to between 3 to 12 million letter differences between two unrelated individuals. Most of these differences are single characters, but there can be changes that involve moving regions from one place to another, or the deletion or duplication of specific regions.

In sexually reproducing organisms, like humans, there are typically two copies of this book in most types of cells of the body, one derived from each of the organism’s parents. Organisms (and cells) with two genomic “books” are known as diploid. When a sexual organism reproduces, it produces cells, known as gametes: sometimes these are the same size. When gametes differ in size, the smaller one is known as a sperm and the larger is known as an egg. Each gamete contains one copy of its own unique version of the genomic book and is said to be haploid. This haploid genome is produced through a complex process known as meiosis (considered in Chapter 11). Meiosis leads to a shuffling of the organism’s original parental genomes. When a haploid sperm and a haploid egg cell fuse, a new diploid organism is formed with its own unique pair of genomic books..

This page titled 3.7: Mutations and the origins of genotype-based variation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.8: The origins of polymorphisms

So what produces the genomic variations found in different individuals in a population? Are these processes continuing to produce genotypic and phenotypic variations or have they ended? First, as we have alluded to, and will return to again and again, the sequence of letters in an organism's genome corresponds to the sequence of in DNA molecules. A DNA molecule in water (and over ~70% of a typical cell is water) is thermodynamically unstable and can undergo various types of reactions, such as deamination, depurination, and depyrimidination that can, if not repaired, lead to changes in the sequence of characters within the molecule (discussed later).[116] In addition, we are continually bombarded by radiation that can damage DNA.[117] Mutagenic radiation, that is, the types of radiation capable of damaging a DNA molecule, arises from various sources, including cosmic rays that originate from outside of the solar system, UV light from the sun, the decay of naturally occurring radioactive isotopes found in rocks and soil (e.g. radon), and the ingestion of naturally occurring isotopes, such as potassium-40. When a DNA molecule absorbs such radiation it can lead to chemical changes in its DNA, that is, mutations. Many but not all of these changes can be identified and repaired by cellular repair systems (to be discussed later).

The second, and major source of change to the genome involves the process of DNA replication. DNA replication happens every time a cell divides. While remarkably accurate it is not perfect; copying can lead to mistakes. In humans, it appears that DNA replication creates one error for every ~1,000,000,000 (10^9) characters copied. Cells have systems that can correct ~99% of these errors, leading to an overall error rate during replication of about 1 in 10^{10} bases replicated. Since a single human cell contains ~6,400,000,000 bases of DNA sequence, that means that generally less than one new mutation is introduced per cell division cycle. Given the number of generations (cell division cycles) from fertilized egg to sexually active adult, that results in about ~100-200 new mutations (changes) added to an individual's genome per generation.[118] These mutations can have a wide range of effects, complicated by the fact that essentially all of the various aspects of an organism's phenotype are determined by the action of hundreds to thousands of genes working in a complex network that is influence by intrinsic and environmental factors. When a mutation leads to a change in a gene, it the new version is known as an allele of the gene. When a mutation changes the DNA's sequence, whether or not it is part of a gene, it creates what is known as a sequence polymorphism or simply a polymorphism, a different DNA sequence. Once an allele or polymorphism has been generated, it is as stable as the original molecule - it can be inherited from a parent and passed on to an offspring. Through the various processes associated with reproduction, which we will consider in detail later on, each organism carries its own distinctive set of alleles and its own unique set of polymorphisms. Taken together these genotypic differences, that is, differences in alleles and polymorphisms, produce different phenotypes. The DNA tests used to determine paternity and forensic identity work because they use the unique polymorphisms and alleles present within an individual's genome as a type of bar code for that person.

Two points are worth noting about genomic changes or mutations. First, whether produced by mistakes in replication or chemical or photochemical reactions, it appears that these changes occur randomly within the genome. With a few notable and highly specific exceptions, for example, in the immune system, there are no known mechanisms by which the environment (or the organism) can influence where a mutation will occur. The second point is that a mutation may or may not influence an organism's phenotype. The effects of a mutation will depend on a number of factors, including exactly where the mutation is in the genome, its specific nature, the role of the mutated gene, the rest of the genome–known as the genetic background–and the environment in which the organism finds itself. We will consider the factors that influence gene and genome dynamics when we return to the behavior of DNA in cells.

Questions to answer:

16. Explain why superfecundity is required for evolution to occur.
17. Why is the presence of genetically inheritable variation essential for any evolutionary model?

Questions to ponder:

- What advantages might be associated with self-imposed controls on mating?
- How might behaviors that limit an individual's ability to reproduce arise?
- How might dam-building beavers influence their environment and various selective effects?
- Is the predicted slowing down and decrease in human population numbers due to genetic or environmental factors? How could your tell?

References

[116] Instability and decay of the primary structure of DNA & DNA has a 521-year half-life:

[117] Although not not to worry, the radiation energy associated with cell phones, bluetooth, and various wifi devices is too low to damage DNA. But no matter what you might hear, it is a mistake to swallow a lamp that emits ultraviolet light.

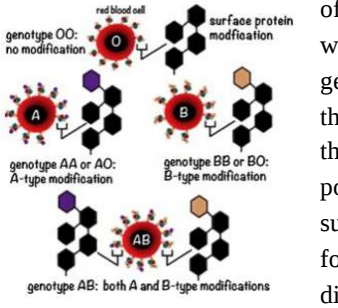
[118] Human mutation rate revealed

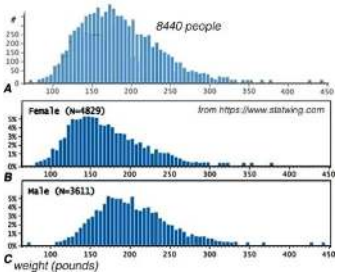
This page titled 3.8: The origins of polymorphisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.9: Genotype-phenotype relationships- discrete and continuous traits

When we think about genetic polymorphisms and alleles (different versions of the same gene), it is tempting to assume simple relationships. In some ways, this is an unfortunate residue from the way many are introduced to genetics. Perhaps you remember Gregor Mendel (1822-1884) and his peas. He identified distinct versions (alleles) of particular "factors", now called genes, that he linked to distinct phenotypes - yellow versus green peas, wrinkled versus smooth peas, tall versus short plants, etc. In the real, wild type world peas, for example display a range of colors, ranging from green to yellow and including intermediate shades. How allelic variations in genes influence traits (phenotypes) is more complex and involves both internal and external environmental factors, stochastic events, and the individual's genetic background (more in chapter 15).

For now we will consider one "simple" example of a gene-trait connection, namely the determination of blood type. Which alleles of the ABO gene you inherit determines whether you have an O, A, B or AB blood type. While we will consider how genes work in greater detail later on, for now it is enough to know that the ABO gene encodes the sequence of a polypeptide that acts as a glycotransferase, a catalyst (an enzyme) that facilitates the addition of a specific chemical group, a carbohydrate, to a protein. Differences in the DNA sequences of the A, B, and O alleles result in differences in the encoded polypeptides. The polypeptides encoded by the A and B alleles differ in the reactions that they catalyze – different sugar groups are added by the A and B versions of the glycotransferase. The polypeptide encoded for by the O allele is inactive, it does not function as a glycotransferase. Remember your cells are diploid; each cell has two copies of each gene (with the exception of the sex chromosomes - in humans, known as X and Y). Each cell involved in synthesizing blood has two copies of the ABO gene, one inherited from your mom and one from your dad. The two ABO alleles you inherited may be the same or different.[119] If they are A and B, the proteins on your red blood cells will have both the A and B modifications, resulting in an AB blood type. If they are A and O or A and A, your red blood cells have only the A modification, if they are B and O or B and B, your red blood cells have only the B modification, and if you have O and O, no modifications (of this type) occur and you have an O blood type (→). These are examples of what are known as discrete traits; you have either an A, B, AB, or O blood type – there are no intermediates. You cannot be 90% A and 10% B.[120] The situation when the presence of a particular allele uniquely determines a trait, as in the case of the ABO gene, is rare – most traits are more complex, they are known as polygenic (many genes) and influenced by non-genetic factors.





Most traits are continuous rather than discrete, they involve hundreds to thousands of genes and their alleles. For example, people come in a continuous range of heights, rather than in discrete sizes. If we look at the values of a trait within a population, that is, if we can associate a discrete number to a trait (which is not always possible), we find that each population can be characterized graphically by a distribution. For example, consider the weights in a group of 8440 adults in the USA (←). The top panel (A) presents a graph of the weights, along the horizontal or X-axis, versus the number of people with that weight along the vertical or Y-axis. We can define the “mean” or average of the population ($\bar{x}$) as the sum of the individual values of a trait (in this case each person’s weight) divided by the number of individuals measured, as defined by the equation:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

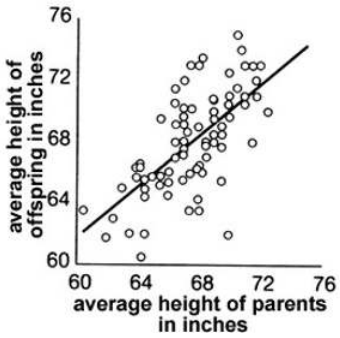
In this particular data set, the mean weight of the population is ~180 pounds. It is common to recognize another characteristic of the population, the median. The median value is the point at which half of the individuals have a smaller value of the trait and half have a larger value. In this case, the median is ~176. Because the mean does not equal the median, we say that the distribution is asymmetric, that is there are more people who are lighter than the mean value compared to those who are heavier. Another way to characterize the shape of the distribution is by what is known as its standard deviation, indicated by the Greek letter sigma (σ). There are different ways to calculate the standard deviation that reflect the shape of the population distribution; we will use a simple one, the so-called uncorrected sample standard deviation (→).[121] To calculate this value subtract the mean value for the population ($\bar{x}$) from the value for each individual (x_i); since x_i can be larger or smaller than the mean, this difference can be a positive or a negative number. We take the square of the difference, which makes all values positive (hopefully this makes sense).

We add together these squared differences, divide that sum by the number of individuals in the population (N), and take the square root, which reverses the effects of our squaring x_i, to arrive at the standard deviation of the population. The smaller the standard deviation, the narrower the distribution - the more organisms in the population have a value near to the mean. The larger σ is, the greater is the extent of the variation in the trait in the population.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

How do we determine whether a complex trait, like height or weight, is genetically determined or influenced? A standard approach is to ask whether there is a correlation between the phenotype in the parents (e.g. their heights) and the phenotypes of their offspring (their heights). Such a correlation between parents and offspring exists for height is suggested by this graph (→), but notice we are seeing a trend, parental height is not a perfect predictor of offspring height - other factors must be involved.

One thing that we cannot determine from such data, however, is how many genes and how allelic interactions influence the "strength" of the trait or how these effects are effected by intrinsic noise (stochasticity), environment and the offspring's specific history. As an example, "*human height has been increasing during the 19th century when comprehensive records began to be kept. The mean height of Dutchmen, for example, increased from 165cm in 1860 to a current average height of 184cm, a spectacular increase that reflects improvements in health care and diet, rather than changes in genes.*"[122] Geneticists currently estimate that allelic differences at more than ~50 genetic loci (positions in the genome) make significant contributions to the determination of height, while allelic differences at hundreds of other genes have smaller effects.[123] At the same time, specific alleles of certain genes can lead to extreme shortness or tallness. For example, mutations that inactivate or over-activate genes encoding factors required for growth can lead to dwarfism or gigantism.



On a didaskalogenic note you may have learned that alleles are often described as either dominant or recessive (a topic to which we will return). But the extent to which an allele is dominant or recessive often depends upon how well we define a particular trait and the extent to which it is influenced by other genetic, developmental, and environmental factors. These effects reveal themselves through the fact that people carrying the same alleles of a particular gene can display (or not) the associated trait, which is known as penetrance, and they can vary in the "strength" of the trait, which is known as expressivity.[124] Both the penetrance and expressivity of a trait can be influenced by genetic background, the presence or absence of particular alleles of other genes. Environmental factors can also have significant effects on the phenotypes associated with a particular allele or genotype. In his studies, Mendel used extensive inbreeding and controlled environments to minimize these effects. In normal (wild type) populations of peas, pea color ranges from green to yellow rather than being green or yellow. Through controlled breeding, he removed the genetic variation that lead to intermediate colors. Another British scientist Walter F.R. Weldon (1860 – 1906) pointed out these complicating factors, but died before he published a complete account of his observations.[125]

References

[119] There are a number of the ABO gene alleles present in the human population, the most common (by far) are the A, B, and O alleles: http://omim.org/entry/110300

[120] Unless you are a chimera, which we will consider anon; see Human blood types have deep evolutionary roots

[121] wikipedia: standard deviation & http://www.mathsisfun.com/data/standard-deviation.html

[122] From Galton to GWAS: quantitative genetics of human height: http://www.ncbi.nlm.nih.gov/pubmed/21429269

[123] Genetics of human height: http://www.ncbi.nlm.nih.gov/pubmed/19818695

[124] Where genotype is not predictive of phenotype: understanding reduced penetrance in human inherited disease

[125] Weldon, W.F.R. (1902). Mendel's laws of alternative inheritance in peas. *Biometrika*, *1*, 228–254. & G. Radick (2024). Your genes do not define you

This page titled 3.9: Genotype-phenotype relationships- discrete and continuous traits is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.10: Variation, selection, and speciation

Combining genetic and associated phenotypic variation, superfecundity, and stable population size, Darwin and Wallace's breakthrough conclusion was that different members of the population would display differences in reproductive success. Some genotypes, and the alleles they contain, would become more common within subsequent generations because the individuals that contained them would reproduce more successfully; their progeny were present as a larger percentage of the next generation. Other genotypes became less common, or disappeared altogether. The effects of specific alleles on an organism's reproductive success will, of course, be influenced by the rest of the organism's genotype and its environment. While some alleles can have a strong positive or negative impact on reproductive success, the effects of most alleles are subtle, assuming they produce any noticeable phenotypic effects at all. A strong positive effect will tend to increase the frequency of the allele (and genotype) associated with it in future generations, while a strong negative effect may lead to the allele disappearing altogether. An allele that increases the probability of death before reproductive age is likely to be strongly selected against, whereas an allele that has only modest effects on the number of offspring an organism produces will be selected for or against more weakly.

What Darwin and Wallace did not know was that genetic information is stored in molecules of DNA, and that that information can be altered through a variety of mechanisms (mutations) that include sequence duplication, deletion, and recombination (shuffling). Moreover, because DNA molecules are relatively stable they can survive the death of the organism (for at least a little while), be released into the environment, and (under certain very specific conditions) be transferred into another organism and become part of their genetic material. These are all features of the molecular nature of genetic information (genes) and how DNA is replicated, repaired, and used to express information within cells. Recognizing these facts led to what is known as the Modern Synthesis of evolutionary biology.[126] While the basic Darwinian rules are the same, the possible molecular complexities make evolutionary processes even more powerful. We will consider various molecular processes as we proceed.

Questions to answer:

18. Generally the products of artificial selection are not competitive with "native" organisms, how come?*
19. What does the word correlation mean to you? what does it mean mathematically or practically?
20. If an individual's height were "determined" by the genotypes of their parents, shouldn't the height measurements of their offspring line on a straight line? Where might the scatter come from?

References

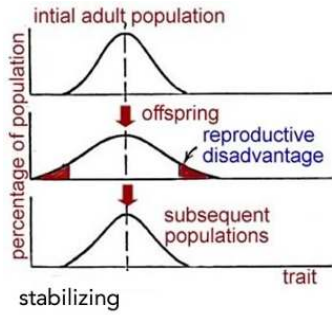
[126] Modern synthesis in evolutionary biology

This page titled 3.10: Variation, selection, and speciation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.11: Types of simple selection

While something of an oversimplification, we begin with three basic types of selection: stabilizing (or conservative), directed, and disruptive. We then introduce the complexities associated with the random aspects of reproduction and the linked nature of genes. We start with a population composed of individuals displaying genetic variation in a particular trait. The ongoing processes of mutation continually introduces new genotypes with varying effects on phenotype. The effects of mutations that produce new alleles can range from the lethal, the organism that carries the mutation either dies or produces no offspring, to apparently neutral – an organism that carries the new allele displays no obvious change in phenotype or reproductive success. A complicating factor is that the phenotypic effects of a particular new allele often depend upon the rest of the genome - due to genetic background effects. At the same time, changes in the population and the general environment influence the predominant types of selection that occur over time, and different types of selection may well (and most certainly are) occurring for different traits.

For each type of selection, we will illustrate the effects as if they were acting along a single dimension, for example smaller to larger, stronger to weaker, lighter to darker, or slower to faster. In fact, most traits vary along a number of dimensions. For example, consider the trait of ear, paw, heart, or big toe shape. An appropriate type of graph would be a multi-dimensional surface, but that is harder to draw clearly. It is also often the case that a genotype that influences one trait also influences other, apparently independent, trait(s). For simplicity's sake, we will start with populations whose distribution for a particular trait can be described by a simple and symmetrical curve, that is the mean and the median are the same. New variants, based on new mutations (new alleles and combinations of alleles), generally fall more or less randomly within this distribution. Under these conditions, for selection NOT to occur we would have to make an seriously unrealistic assumption, namely that all organisms (or a pair of organisms a sexually reproducing species) are equally successful at surviving and producing offspring, something that is observably not the case. Selection occurs when genetic variation influences reproductive success, although the strength of selection (the difference in the average number of viable offspring produced) can vary dramatically between traits.

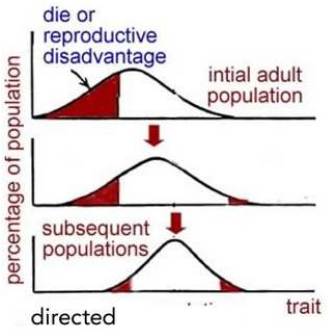


Stabilizing selection: Sometimes a population of organisms appears static, that is, the mean and standard deviation of a trait are not changing over time. Does that mean that selection has stopped? Obviously we can turn this question around: if we assume that there is a population with a certain stable mean and standard deviation of a trait – what would happen over time if selection disappeared?

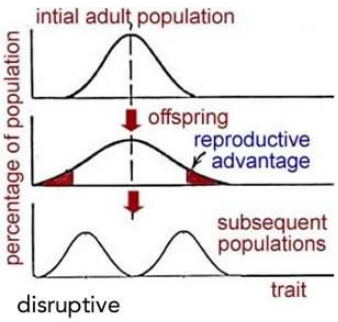
Let us assume we are dealing with an established population living in a stable environment. In this population organisms are capable of reproducing more organisms than are needed to replace them. Now consider the factors that lead to the original population distribution: why is the mean value of the trait the value that it is? What factors influence the observed standard deviation? Assuming that natural selection is active, it must be that organisms that display a value of the trait far from the mean are (on average) at a reproductive disadvantage compare to those with the mean value of the trait (→); there is “negative” selection against extreme (far from the mean) values of the trait. These individuals tend to produce fewer fertile offspring than those with a value of the trait near the mean. We can generate a crude estimate of the "strength" of negative selection by looking at the shape of the trait's population distribution curve. The narrower the distribution (the smaller the standard deviation), the stronger is the negative selection effect. Similarly, where a trait’s distribution is broad we predict that the impact of the trait on reproductive success will be relatively weak. The variation in the trait will be characterized only by maximum and minimum values, reflecting the limits of what the system can produce and remain viable.[127] We do not know why this is the case and don’t really care at the moment.

Directed selection: Imagine that the population’s environment changes. It may now be the case that the phenotype of the mean is no longer the optimal phenotype in terms of reproductive success, the only factor that matters, evolutionarily. A different value of the trait may now be more favorable. Under these conditions we would expect that, over time, the mean of the distribution will shift toward the phenotypic value associated with maximum reproductive success (→). Once reached, and assuming the environment stays constant, stabilizing selection again becomes the predominant process. One outcome of directed selection is that, as the selected population’s mean moves, it may well alter the environment of other organisms.

For directed selection to work, the environment must change at a rate and to an extent compatible with the changing mean phenotype of the population. Too big and/or too rapid an environmental change and the reproductive success of all members of the population may be dramatically reduced. The ability of the population to change will depend upon both the genetic variation



present within the original population and the rate at which new mutations are produced, generally a relatively slow and constant process.[128] In some cases, the change in the environment may be so fast or so drastic, and the associated impact on reproduction so severe, that selection will fail to move the population and extinction will occur.



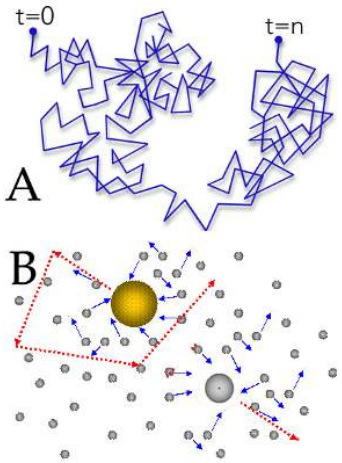
Disruptive selection: A third possibility is that a population of organisms find themselves in an environment in which traits at the extremes of the population's phenotypic distribution have a reproductive advantage over those around the mean. If we think about the trait distribution as a multidimensional surface, it is possible that in a particular environment (which may correspond to multiple geographic regions), there will be multiple distinct strategies that lead to greater reproductive success compared to others. This leads to what is known as disruptive selection (→). In an asexually reproducing population, various lineages will be subject to selective pressures based on the environments (regions) they come to inhabit, and the likelihood that individuals move from environment to environment.

The effect of disruptive selection in a sexually reproducing population will be opposed by the random mating between members of the population. But is random mating really a good assumption? It could be that the different environments, referred to as ecological niches, are physically distant from one another and that organisms do not travel far to find a mate. In the process of adapting to the two different niches, the population may split into subpopulations. Over time, two species may emerge, since when and with whom one chooses to mate with and the productivity of such matings are selectable traits. Disruptive selection can lead to reproductive isolation, a hallmark of speciation. Over time, millions of species can be expected to appear. This was the observation (many species) that Darwin and Wallace set out to explain.

This page titled 3.11: Types of simple selection is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.12: Considering stochastic processes

Stochastic processes play an important role in evolutionary mechanisms (population bottlenecks, founder effects, genetic drift, meiotic recombination) as well as molecular processes within cells and tissues (expanded upon later on). While we have already used it, you may not be familiar with the word stochastic, a word whose meaning is often confused with random. What distinguishes a stochastic from a random process? A truly random process has no underlying natural cause and is completely unpredictable. A miracle could be considered a random event. From a scientific perspective, one could argue that there are no truly random natural processes or events, no miracles. Our working hypothesis is that all natural events have identifiable and measurable causes. That said, that does not mean that individual events are predictable. Natural events can be unpredictable for one of two basic reasons: the event may be determined by theoretically unknowable or currently unknown factors, as in the case of the radioactive decay of atoms. Alternatively, the event may be the result of a large number of theoretically knowable events that are, for practical reasons impossible to measure accurately without perturbing the system in unknowable ways. Such events are analogous to, or versions of, Brownian motion, a phenomena named after the Scottish botanist Robert Brown (1773-1858). In Brownian motion small but visible particles, suspended in a solution (air or water), are found to move in a jerky and irregular manner (A→ path of particle over time). Brownian motion is due to a visible particle colliding with many invisible objects (molecules) present in the environment (air/water: B→ collisions).[129] The average energy transferred through these collisions reflects the temperature of the system. At higher temperatures the molecules have a higher average (mean) kinetic energy ($^1/_2\ mv^2$). During a particular time interval, the sum of all collisions can lead to an unbalanced force on the particle that causes it to move. A short time later the sum of these collision forces is likely to point in a different direction and the particle will move in that direction. Collisions between molecules supply the energy to drive the dissociation of molecules from one another; they provide the activation energy required for chemical reactions to proceed (see Chapter 5). At the individual event level, the system is unpredictable. Yet because a typical system involves many molecules and collision events. In water there are ~3×10^{22} water molecules per cubic centimeter, with the average water molecule traveling ~$2.5\text{x}10^{-8}$ centimeters between collisions.[130] The end result is that the system's bulk behavior (e.g. rates of diffusion) are highly predictable while individual particle movements are unpredictable.



In classical, that is pre-quantum mechanical physics, it was assumed that if we knew the velocity (speed and direction) of every molecule in the system, as well as the dynamics of the collisions, we could predict the future behavior of the system and the paths of Brownian particle movements.[131] But it turns out that the world does not behave that way. We cannot, even theoretically, achieve the necessary level of accurate measurement. We are limited by what is known as the Heisenberg Uncertainty principle, which arises from the fact that matter appears to be composed of objects with both wave- and particle-like properties, rather than simple billiard ball-like particles.[132] So how is it possible to understand Brownian motion scientifically? The answer is that when we look at a large enough population of objects, the population's behavior becomes predictable – this predictability implies an underlying cause. For example, consider measurements of a large number of particles undergoing Brownian movement. If we measure the distance between where they start (time (t) t=0) and where they end up at t=n (see A↑ above), we find that the average distance of all particles travelled but not the direction of travel or extent of travel of any particular particle is predictable and reflects particle size, the nature of the system (water, air, etc), and temperature.



The situation is similar to that of rolling dice. While it is impossible to accurately predict the outcome of a single dice roll, as we increase the number of rolls (the population of rolls), we find that the overall behavior becomes increasingly predictable, each of the six numbers (assuming that this is a fair cube dice) will appear $1/6^{th}$ of the time. The larger the number of rolls, the more closely the number of each possible outcome will approach $1/6^{th}$ of the total. This is known as the law of large numbers. A similar situation occurs with radioactive atoms; while it is impossible to predict when any particular atom will decay, when we consider a large enough population we can accurately predict when any particular percentage of the original population will have decayed. The time it takes for 50% of the atoms present originally to decay is known as the “half-life” of the isotope and can be determined to high precision.

In the case of rolling dice, and other similar (simple) stochastic processes, it is important, but hard to remember, that each individual event is independent, what happened in the past does not influence what will happen next. Forgetting this rule leads to what is known as the Gambler's Fallacy.[133] As an example, you roll a die eight times and get 2, 2, 5, 2, 2, 6, 2, 2. Assuming of course that this is a fair die, what is the probability that the next roll will come up 2? No matter how many times a 2 came up in the past, the chance of rolling a 2 on the next roll remains the same, 1/6.

A complexity that occurs within biological systems is that while a particular event can be stochastic, individually unpredictable but well behaved in a large enough population, in a cell or an organism, a single inherently unpredictable event, such as the activation or mutation of a particular gene, can change the system so as to produce different behaviors and outcomes. A mutation can, for example, initiate the process by which a cell becomes cancerous. It is therefore possible, and perhaps likely, that if the history of the organism (or life) were to be "rerun" (an impossible situation), outcomes would be different.

Questions to answer:

24. How can you distinguish stochastic from random events? What types of everyday stochastic events are you familiar with.
25. How might you decide whether a pattern in data was due to an underlying process or "just" to chance?

Question to ponder:

- Is it possible to study random events scientifically?

References

[129] Albert Einstein: The Size and Existence of Atoms & Einstein and Brownian Motion

[130] The properties of water: http://galileo.phys.virginia.edu/classes/304/h2o.pdf

[131] see Laplace's demon: https://en.wikipedia.org/wiki/Laplace's_demon

[132] Want to know more? check out: What is the Heisenberg Uncertainty Principle? and How Heisenberg Became Uncertain(https://youtu.be/UFYnsxLuFdQ)

[133] Gambler's Fallacy: https://en.wikipedia.org/wiki/Gambler's_fallacy

This page titled 3.12: Considering stochastic processes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.13: Genetic diversity, population size, founder effects and population bottlenecks

We can determine the exact frequency of each allele present in a population by examining each individual BUT for a large population this is expensive and likely impractical. Instead, we turn to "sampling". We examine a subset of the population. If the number "sampled" is small with respect to total population size, we can expect significant differences in allele frequencies in the measured (sampled) and actual (total) populations. These differences become smaller as the sample size increases. To provide a concrete example, consider a large population in which each individual carries one (and only one) of six alleles of a particular gene and the six alleles are present in the same frequencies. The selection of any one individual from the total population is like a throw of a fair die; there is an equal $1/6^{th}$ chance of selecting an individual with one particular allele. Since the parental population is large, the removal of one individual does not appreciably change the distribution of alleles remaining, so the selection of a second individual produces a result that is independent of the first. Just like individual rolls of the die, there will be a $1/6^{th}$ chance to select any one of the six alleles. But the odds that a small subpopulation will be made of individuals with equal numbers of the six alleles will be small. The more alleles present in the parent population, the less likely that a small sample will reflect the allele frequencies present in the parental population.

A founder population is a sample of the parent population. The smaller the founder population, the greater the chance that it will be missing some alleles and over-represent others; it will be genetically distinct from the original population. So when a small group from a parent population invades or migrates into a new environment, it is likely to have a genotypic (allelic) profile different from its parent population. This difference is not due to natural selection but to chance alone. Nevertheless, it will influence subsequent evolutionary events; the smaller founder population will likely respond in different ways to new mutations and environmental pressures based on which alleles are present. The situation will be further influenced if genetic factors impact migratory behavior or reproductive success in the new environment.

When we think about evolutionary processes it is important to remember founder effects. Something similar happens when a large population is dramatically reduced in size for any of a number of reasons, a situation known as a population bottleneck (discussed further below). Both are stochastic events that can lead to populations with allele frequencies different from that of the original “parental” population. Small populations are also susceptible to a stochastic process known as genetic drift. Founder effects, bottlenecks, and genetic drift can produce populations with unique traits that are not directly due to the effects of natural selection. Since founder effects and population bottlenecks can occur a number of times during the course of a population's evolution, it is a mistake to assume that all observed traits have positive effects on reproductive success. If we think of evolutionary change as reflecting the movement of a population through a fitness landscape–the combination of the various factors that influence reproductive success–over time, then the isolation of small populations, and allelic changes within them, can cause a jump from one place in the landscape to another. As the population grows larger, new mutations leading to new alleles and adaptations can occur – selection again becomes the main driver of evolutionary change. Deleterious effects, that become frequent due to non-adaptive processes, can be ameliorated.

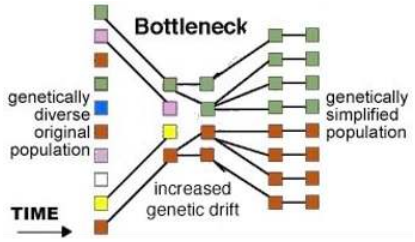
A population invading a new environment will encounter a new set of organisms to compete and/or cooperate with. A catastrophic environmental change can change the selective landscape, removing or introducing competitors, predators, pathogens, and possible cooperators, favoring new adaptations and selecting against others that might have once been beneficial, in terms of reproductive success. One effect of the major extinction events that have occurred during the history of life on Earth is that they provide new adaptive contexts, a different and less densely populated playing field with fewer direct competitors.[134] The expansion of various species of birds and mammals that followed the extinction of the dinosaurs is an example of one such opportunity, associated with changes in selective pressures.

References

[134] Big Five mass extinction events and How life blossomed after the dinosaurs died

This page titled 3.13: Genetic diversity, population size, founder effects and population bottlenecks is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.14: Population bottlenecks



Population bottlenecks occur when an environmental change leads to the dramatic reduction in the size of a population. Catastrophic environmental changes, such as asteroid impacts, massive and prolonged volcanic eruptions associated with continental drift, or the introduction of a particularly deadly pathogen that kills a high percentage of the organisms that it infects, can all create a population bottleneck (←). Who survives the bottleneck can be due to "luck" or may be based on genetic and other factors, for example, alleles associated with disease resistance.

There is compelling evidence that such drastic environmental events are lead to population bottlenecks so severe that they led to mass extinctions. The most catastrophic of these events was the Permian extinction that occurred ~250 million years ago; during this event it appears that ~95% of all marine species and ~75% of land species went extinct.[135] If most species were affected, we would not be surprised if the surviving populations experienced serious bottlenecks. The subsequent diversification of the surviving organisms, such as the *Dinosauria* that includes both the extinct dinosaurs and modern birds, and the *Cynodontia*, which includes the ancestors of modern mammals, including us, could be due in part to bottleneck-associated effects. An asteroid impact, known as the Cretaceous-Tertiary event, occurred ~65 million years ago; it contributed to the extinction of the dinosaurs and led to the rapid expansion and diversification of mammals, which first appeared in the fossil record ~100 million years earlier.

Surviving an asteroid impact or dramatic changes in climate may be "random". Consider the effects of a severe drought or highly virulent bacterial or viral infection. The organisms that survive may have specific phenotypes, and associated genotypes, that influence their chance of survival. In such cases, the effect of the bottleneck event would lead to genotypic changes in the post-bottleneck population – effects that could influence the population in various ways. For example, a trait positively associated with pathogen resistance may have negative (or other) phenotypic effects. After the pathogen-driven bottleneck, mutations that mitigate negative effects associated with the pathogen resistance trait may have a selective advantage. The end result is that traits that would not be selected in the absence of the pathogen, are selected and become common.

We can identify extreme population reduction events (founder effects and bottlenecks) by looking at what are known as neutral polymorphisms, variation in regions of the DNA not known to influence selectable phenotypes. These changes are expected to accumulate at a constant rate over time (can you suggest why?) The rate of the accumulation of neutral polymorphisms serves as a type of population-based clock. Its rate can be estimated, at least roughly, by comparing the genotypes of individuals of different populations whose time of separation can be accurately estimated, assuming of course that there has been no significant migration between the populations. A bottleneck (or founder effect) will typically lead to a dramatic reduction in the number of neutral polymorphisms present.

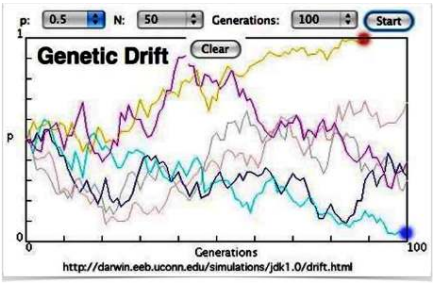
References

[135] The Permian extinction and the evolution of endothermy

This page titled 3.14: Population bottlenecks is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.15: Genetic drift

Genetic drift is a stochastic process that becomes important in small populations or over long periods of time. It can lead to non-adaptive evolutionary phenomenon that explain a number of observations. In asexual populations genetic drift is due to random effects on organismic survival that can, in practice be difficult to distinguish from selective effects. In contrast, drift is unavoidable in small populations of sexually reproducing organisms. This is because cells known as gametes are produced during the process of sexual reproduction (Chapter 4). While the cell that generates these gametes contains two copies of each gene, and each gene reflects one of the alleles present within the population, any particular gamete contains only one allele of each gene. Two gametes fuse to produce a new diploid organism. This process combines two chance events: which allele is present in a particular gamete and which particular gametes fuse to produce a new organism. In a small population, over a reasonably small number of generations, one of multiple alleles at a particular genetic locus may be lost simply by chance. In this figure (←), six experimental outcomes (each line) were analyzed over the course of 100 generations. The population originally contained two different alleles of a particular gene, originally present in equal numbers in a population of 50 individuals. While we are tracking only one genetic locus, the same behavior impacts every gene for which multiple alleles are present. In two of these six populations, one allele (blue dot) has been lost or is close to being lost while another allele (red dot) has become the only allele present in the population; it is said to have been fixed. Assume that the two alleles convey no selective advantage with respect to one another, can you predict what will happen if we let the experiment run through 10,000 generations? What assumptions are you making? For the mathematically inclined, it is possible to estimate the effects of mild to moderate positive or negative selective pressures on allele frequencies and the probability that a particular allele will be lost or fixed through genetic drift.



A historical example of the effects of drift: Consider the observation that many primates are strictly dependent on the presence of ascorbic acid (vitamin C) in their diet. Primates are divided into two suborders, the *Haplorhini*, from the Greek meaning “dry noses”, and the *Strepsirrhini*, meaning “wet noses”. The *Strepsirrhini* include the lemurs and lorices, while the *Haplorhini* include the tarsiers and the anthropoids, monkeys, apes, and humans. The *Haplorhini*, but not the *Strepsirrhini*, require vitamin C in their diet. In vertebrates vitamin C plays an essential role in the synthesis of collagen, a protein involved in the structural integrity of a wide range of tissues. In vitamin C-dependent organisms the absence of dietary vitamin C leads to the disease scurvy that according to Wikipedia, “often presents itself initially as symptoms of malaise and lethargy, followed by formation of spots on the skin, spongy gums, and bleeding from mucous membranes. Spots are most abundant on the thighs and legs, and a person with the ailment looks pale, feels depressed, and is partially immobilized. As scurvy advances, there can be open, suppurating wounds, loss of teeth, jaundice, fever, neuropathy, and death.”[136]

The *Haplorhini* require dietary vitamin C due to a mutation in the *GULO1* gene. The GULO1 protein is an enzyme (1-gulono-gamma-lactone oxidase) required for the synthesis of vitamin C. One can show that the absence of a functional *GULO1* gene is the root cause of vitamin C dependence by putting a working copy of the gene, for example derived from a mouse, into human cells. The mouse-derived *GULO1* allele encodes a functional form of the Gulo1 enzyme and “cures” the human cells’ of their need for exogenous vitamin C. But, no matter how advantageous a working *GULO1* allele might be, particularly for British sailors, who died in large numbers before a preventative treatment for scurvy was discovered[137], no new, functional *GULO1* allele has appeared in the lineage leading to humans or the other Haplorhini. It is easier to break something than to fix it through random changes.

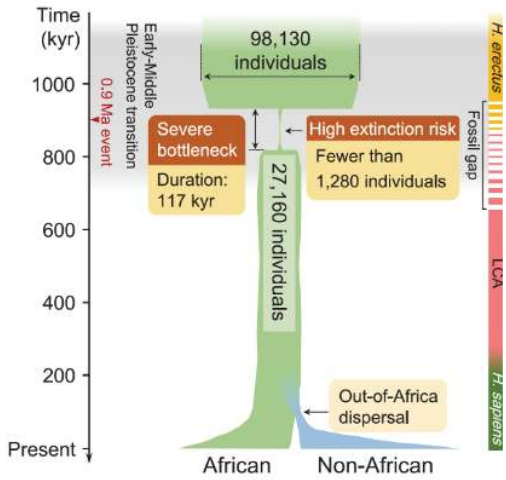
Alleles are selected from alleles already present in the population or that appear through *de novo* (new) mutations. Mutation is a stochastic process; organisms do not necessarily produce the genes or alleles they "need" or that might be beneficial. In some cases there may be no plausible molecular pathway that can generate a "needed" allele (or gene). The mutant (non-functional) *GULO1* allele appears to have become "fixed"; it is the only *GULO1* allele present in the population that gave rise to the Haplorrhini, around ~40 million years ago. So the question is, how did our ancestors come to survive the loss of a functional version of an important gene? A plausible scenario is that the loss of a functional allele occurred when vitamin C was present in the organisms' diet.

The rest of the organism’s genotype can influence the phenotype associated with a particular allele. The presence or absence of various alleles within the population (and an organism) produces genetic background effects that can influence the phenotypes

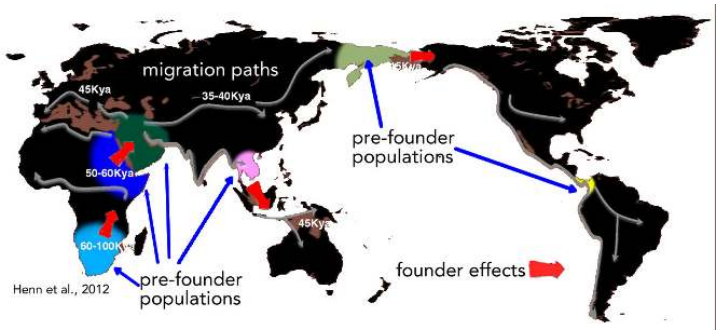
observed (chapter 12). If an allele disappears because of genetic drift, future evolutionary changes may be constrained, or perhaps better put, redirected. At each point, the future directions open to evolutionary mechanisms depend in large measure on the alleles already present in the population. Of course new alleles continue to arise by mutation, but they are originally infrequent, just sporadic effects in individuals, so unless they are strongly selected for (and even if they are selected for) they may be lost from the population by genetic drift.[138] Drift can lead to some weird outcomes. For example, what happens if drift leads to the fixation of a mildly deleterious allele, let us call this allele BBY. Now the presence of BBY will change the selective landscape: mutations and or alleles that ameliorate BBY's negative effects will increase reproductive success, selection pressures will favor those alleles. This can lead to evolution changing direction, even if only subtly. With similar effects going on across the genome, one quickly begins to understand why evolution is something like a drunken walk across a selective landscape, with genetic drift, founder and bottleneck effects resulting in periodic stochastic staggers in new directions. In fact this can be beneficial, these phenotypic variants enable the population to sample the range of accessible variations, and "select" those that work best (at least in terms of short term reproductive advantage).

The use of pre-existing variation, rather than the idea that an organism invents genetic variations as they are required, was a key point in Darwin's view of evolutionary processes. There is no known mechanism by which organisms can create the alleles they need or "want" on demand, no simple link between a particular genetic variant (allele) and a specific phenotype. Rather, the allelic variation generated by mutation, selection, and drift are what evolutionary processes have to work with.[139] Only a very rare mutation that recreates (resurrects or fixes) a lost allele can bring an allele back into the population once it has been lost. Founder and bottleneck effects, together with genetic drift combine to produce what are known as non-adaptive processes and make the history of a population a critical determinant of its future evolution.

Considering human evolution



The ancestral population of *Homo sapiens* appears to have emerged in Southern Africa roughly ~1 million years ago.[140] Hu et al used comparative genomic analysis to conclude that there was a "reduction in the population size of our ancestors from about 100,000 to about 1000 individuals, which persisted for about 100,000 years" and that "estimated effective (ancestral human) population size during the bottleneck period was only 1280 breeding individuals" (figure from Hu et al →).[141] The population then increased and small populations of humans migrated out of southern Africa about ~65,000 years



ago into the Horn of Africa and the Arabian peninsula, and from there into Europe, Asia, Oceania, and finally into North America and throughout central and South America (from Henn et al 2012 ↓).[142] These migrations involve multiple founder effects. The result is that African populations display a greater overall genotypic (genetic) complexity than do groups derived from it. The migrating *Home sapiens* populations had a period of inbreeding with Neanderthals and the Denisovians (see earlier). Neanderthals and the Denisovians appear to have diverged from the *Homo sapiens* lineage ~1.2 million years ago.[143] Comparing genotypes, that is, neutral polymorphisms, between isolated populations enables us to estimate that humans reached Australia ~45,000 years ago and entered the Americas in multiple waves beginning ~20,000 years ago (and perhaps a earlier).[144] The presence of humans changed the environmental pressures on such organisms around the world. For example, arrival of humans has been linked to the extinction of a group of mammals known as the megafauna in those environments.[145]

Questions to answer:

26. How might the extinction of one type of organism influence the evolution of others?
27. What factors can influence the different effects of a bottleneck and a founder effect on a population's future evolution?
28. How is it that genetic drift can be accurately quantified, but still be unpredictable in any particular population?

Questions to ponder:

- How is determining allele frequency in a population similar to and different from political polling?
- Does passing through a bottleneck improve or hamper a population's chances for evolutionary success?

References

[136] An amazing fact is that it took the deaths of thousands of sailors to understand the nutritional role of vitamin C.

[137] http://mentalfloss.com/article/24149/how-scurvy-was-cured-then-cure-was-lost

[138] A short video on the genetic drift app (which appears to have disappeared). There are examples where an allele almost disappears and then becomes fixed; it does happen.

[139] An exception involves the process known as horizontal gene transfer. Viruses also contain genes that they can transfer from organism to organism.

[140] Dating origins depends upon finding fossils: see The great human expansion and Oldest Homo sapiens fossil claim rewrites our species' historyand Mobile elements reveal small population size in the ancient ancestors of Homo sapiens

[141] Hu et al, (2023). Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. Science, 381: 979-984.

[142] Henn et al. (2012). The great human expansion. Proceedings of the National Academy of Sciences, 109: 17758-17764.

[143] Genetic Data and Fossil Evidence Tell Differing Tales of Human Origins

[144] New age constraints for human entry into the Americas on the north Pacific coast

[145] Megafauna extinction effects and an interesting video

This page titled 3.15: Genetic drift is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.16: Gene linkage- one more complication

So far, we have not worried much about the organization of genes in an organism. We also have not considered what, exactly a gene is. For now, let us just say that a gene is information encoded within a region of a DNA molecule and that multiple genes are present in a single DNA molecule – we consider genes in detail later (Chapter 7). As you might conclude from genetic background effects, genes and the products they encode interact. We bring this up because the way genes are organized can influence evolutionary processes. As mentioned above, in his studies, Gregor Mendel spent a fair amount of time looking for “well behaved” genes and alleles, those that display simple recessive and dominant behaviors and that act as if they were independent of one another.[146] As noted by Kampourakis, "*Weldon’s (1902) studies of varieties of pea hybrids led him to conclude that there was a continuum of colors from greenish yellow to yellowish green, as well as a continuum of shapes from smooth to wrinkled. It thus appeared that in obtaining purebred plants for his experiments, Mendel had actually eliminated all natural variation in peas, and that characteristics were not as discontinuous as he had assumed*". The situation complex for most traits and the genes that influence them. Traits are rarely dichotomous (one or the other), and often influenced by multiple genes. Genes often act as if they are linked together because they often are. Gene linkage arises from the organization of genes within chromosomes; each chromosome is in individual DNA molecules. So what happens to linked genes when a particular allele of a particular gene is strongly selected for or against? That allele, together with the alleles of linked genes, are also "selected". We can think of this as a "by-stander" effect. Their frequency in a population increases or decreases not because of their direct effects on reproductive success, but because of their location within the genome, their “linkage” to an allele that strongly influences selection.

In sexually reproducing organisms the linkage between alleles (genes) is not permanent. There are processes (meiotic recombination) that can shuffle the alleles on a chromosome. The end result of such recombination events is that the further away two genes are from one another on a DNA molecule in a chromosome, the more likely it is that alleles of those genes will become unlinked, that is, have independent effects on reproductive success. Over time, the effects of linkage will be lost, but not necessarily before particular alleles have been fixed and other alleles lost within the population. For example, extremely strong selection for a particular allele of one gene can lead to the fixation of mildly deleterious alleles in closely linked, neighboring genes.

Let us clarify some terms related to genes. These terms arise from the history of biology in general, and genetics in particular. We now know that genetic information is stored in the sequence of double-stranded DNA molecules. A gene is the region of a DNA molecule that encodes a particular “gene product”, either an RNA molecule or a polypeptide, together with regions required for controlling when and where the gene's "product" is “expressed”. Where and when a gene is expressed is regulated by networks of interacting molecules. All of the DNA molecules present in a cell are known collectively as the cell’s genome. We refer to the position of a particular gene within the genome as its genetic locus. In Latin locus means ‘place’; think location – a word derived from the same root. A particular genetic locus can be occupied by any of a number of distinct alleles (DNA sequences). There are various mechanisms in cells that can duplicate, delete, insert, or move a region of DNA within the genome, creating or eliminating new genetic loci. The phenotype associated with an allele can be influenced by its position within the genome.

A trait that increases reproductive success, by increasing the number of surviving offspring, may be associated with other not-so-beneficial, and sometime detrimental (to individuals) effects. The key is to remember that evolutionary mechanisms do not necessarily result in what is best for an individual but what in the end enhances the net (short term) reproductive success of a population. Evolutionary processes do not select for particular genes or new versions of genes but rather for those combinations of alleles that optimize reproductive success. The situation gets more complicated when evolutionary mechanisms generate organisms, like humans, who think and feel and can actively object to the outcomes of evolutionary processes. From the point of view of self-conscious organisms, evolution can appear harsh, or at the very least totally uninterested in the desires and happiness of individuals. This was one reason that Darwin preferred impersonal (naturalistic) mechanisms over the idea of a God responsible for what can appear to be the gratuitously cruel aspects of its creation.

Questions to answer:

29. What, exactly, is the difference between a gene and an allele? a gene and a chromosome?
30. How might the linkage of genes along a chromosome influence evolutionary processes?
31. Consider this quote from Charles Darwin, “Natural selection will never produce in a being any structure more injurious than beneficial to that being, for natural selection acts solely by and for the good of each.” How would you modify it in light of our understanding of genes and chromosomes?

Question to ponder:

- Given evolution's focus on reproductive success and cost-benefit analysis, rather than individual well-being impact your thinking on whether the natural is inherently good (or is it irrelevant)?

References

[146] Mendelian controversies: a botanical and historical review

---

This page titled 3.16: Gene linkage- one more complication is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.17: A brief reflection on the complexity of phenotypic traits

We can classify traits into three general types: adaptive, non-adaptive, and deleterious. Adaptive traits are those that, when present increase the organism's reproductive success. These are the traits we normally think of when we think about evolutionary processes. Non-adaptive traits are those generated by stochastic processes, like drift, founder effects, and bottlenecks. These traits become established not because they improve reproductive success but simply because they happened to have become fixed within the population. If an allele is deleterious independent of its environment, it will be expected to rapidly disappear from the population, unless other factors are in play. Rare, strongly deleterious alleles are, most likely, the result of new mutations, or they may lead to a selective advantage in specific situations, which are no longer relevant.

When we consider a deleterious allele we are referring to its effects on reproduction. An allele can "harm" the individual carrying it yet persist in the population because it leads to an increase in the number of viable offspring produced. Similarly, there are traits that can be seen as actively maladaptive, but which occur within the population because they are linked mechanistically to some other positively selected trait. Many genes are involved in a number of distinct processes and their alleles can lead to multiple phenotypic effects. Such alleles are said to be pleiotropic. Not all of the pleiotropic effects of an allele are necessarily of the same type; some can be beneficial, others deleterious. As an example, a trait that dramatically increases the survival of the young but leads to earlier senility and sudden death in older adults could be positively selected for. In this scenario, the maladaptive senility/death trait would persist because of its association with a highly adaptive juvenile survival trait. What is happening is a form of population-level cost-benefit analysis. If the net evolutionary benefits of an allele exceeds its costs, the allele and the trait associated with it can be subject to positive selection. If the costs exceed the benefits, it will be selected against. It is worth noting that a trait that is advantageous in one environment may be disadvantageous in another. Consider the effects of the *GULO1* mutation. All of which is to say that when thinking about evolutionary mechanisms, it is best not to assume that a particular trait exists independent of other traits, that it acts independently of the environment, or that the presence of a trait is evidence that it is beneficial.

This page titled 3.17: A brief reflection on the complexity of phenotypic traits is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.18: Speciation and extinction

As noted, an important observation that needed a scientific explanation was why, exactly, are there so many different types of organisms. The Theory of Evolution explains this observation through the process of speciation. The basic idea is that populations of organisms can split into distinct groups. Over time evolutionary mechanisms acting on these populations produce and maintain distinct types of organisms, different species. We also know, from the fossil record and from modern experiences, that species can disappear – they can go extinct. What leads to the formation of new species or the disappearance of existing ones?

To answer these questions, we have to consider how populations behave. A population of a particular type of organism will typically inhabit a particular geographical region. The size of these regions can range from over an entire continent or more to very limited regions, such as a isolated lake or cave. When we consider organisms that reproduce sexually, a process that involves cooperation between individuals, we have to consider how far a particular organism (or its gametes) can travel. The reproductive range of some organisms is quite limited, whereas others can travel significant distances. Another key factor is how an organism makes its living - where and how do they get the matter and energy (that is, food) and the space needed to reproduce successfully? Together these are referred to as a specie's (population's) ecological niche.

An organism's ecological niche is the result of its past evolutionary history, the selection pressures acting within a particular environment, and its current range of behaviors. In a stable environment, and a large enough population, reproductive success will reflect how effectively individuals exploit their ecological niche. Stabilizing selection will optimize adaptation to the niche. It is possible that different types of organisms will compete for similar resources, for a similar niche. Interspecies competition leads to a new form of selective pressure. If individuals of one population can exploit a different set of resources or the same resources differently, these organisms can minimize competition and become more reproductively successful compared to individuals that compete directly with individuals of other species. The result is what is known as the competitive exclusion principle or Gause's Law that states that two species cannot stably occupy the same ecological niche – over time either one will leave (or rather be forced out) of the niche, or will evolve to fill a different, often subtly different niche.[147] What can be hard to appreciate is how specific a viable ecological niche can be. For example, consider the species described by the evolutionary biologist Theodosius Dobzhansky (1900-1975): "*Some organisms are amazingly specialized. Perhaps the narrowest ecologic niche of all is that of a species of the fungus family Laboulbeniaceae, which grows exclusively on the rear portion of the elytra (the wing cover) of the beetle Aphenops cronei, which is found only in some limestone caves in southern France. Larvae of the fly Psilopa petrolei develop in seepages of crude oil in California oilfields; as far as is known they occur nowhere else.*"

> *So, naturalists observe, a flea has smaller fleas that on him prey; and these have smaller still to bite 'em; and so proceed ad infinitum.*
>
> *- Jonathan Swift*



While it is tempting to think of ecological niches in broad terms, the fact is that subtle environmental differences can favor specific traits and specific organisms. If an organism's range is large enough and each individual's range is limited, distinct traits can be prominent in different regions of the species' range. These different subpopulations reflect local adaptations.[148] For example, it is thought that as human populations migrated out of the equatorial regions of Africa, they were subject to differential selection based on exposure to sunlight, due in part to the role of sunlight in the synthesis of vitamin D and its ability to induce cancer-causing mutations and skin damage (sun burn).[149] In their original ecological niche, the ancestors of humans were thought to hunt in the open savannah (rather than within forests), and so developed adaptations to control body temperature. Our general lack of body hair and ability to sweat compared to other mammals are thought to be such adaptations.

The absence of a thick coat of hair also allowed direct exposure to UV-light from the sun. While UV exposure is necessary for the synthesis of vitamin D, too much exposure can lead to skin cancer. Dark skin pigmentation is thought to be an adaptive compromise. As human populations moved away from the equator, the dangers of UV exposure decreased while the need for vitamin D production remained. Under such conditions, allelic variations that favored lighter skin pigmentation, but retained the ability to tan to some extent appears to have been selected (←). Genetic analyses of different populations has begun to reveal exactly which alleles in which genes emerged in different human populations as they migrated out

of Africa and across the Earth. Of course, with humans the situation has an added level of complexity. For example, the relatively recent trait of wearing clothing directly impacts the pressure of “solar selection.” And some pinker folk favor darker (tanned) skin. A number of different phenotypic variations can occur over the geographical range of a species. Differences in climatic conditions, pathogens, predators, and prey can all lead to multiple local adaptations, like those associated with human skin color.

References

[147] Competitive exclusion principle - reminiscent of the Pauli exclusion principle in Quantum Mechanics

[148] Sometimes sub or local populations are termed subspecies or races. One can (and we will) argue that the term race is obsolete and used to justify group prejudices. Consider: Avoiding unrecognized racist implications arising from teaching genetics.

[149] Genetics of skin color: image sources: http://hmg.oxfordjournals.org/content/18/R1/R9.full

---

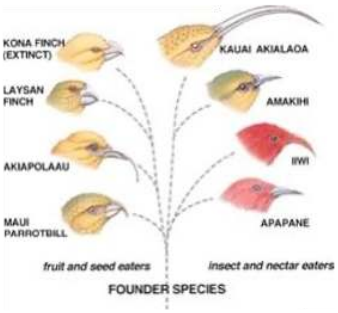
This page titled 3.18: Speciation and extinction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.19: Mechanisms of speciation

Various mechanisms that can lead a species to give rise to one or more new species. Remembering that species, at least species that reproduce sexually, are defined by the fact that they can and do interbreed to produce fertile offspring. An important point is that the process of speciation is continuous, there is no magic moment when one species becomes another, rather a new species emerges over time from a pre-existing species, after which the two populations evolve independently.[150] The origin of species through evolutionary mechanisms is formally analogous to the Cell Theory; each cell is derived from a pre-existing cell. The difference is that cell division provides an unambiguous benchmark in the history of a cell. Species are populations of organisms at a moment in time, they are connected to past species, can produce new species, or can go extinct in the future.

Allopatric speciation is the simplest way to form a new species. It occurs when the original population is physically divided into isolated subpopulations that can no longer breed with one another. If the environments inhabited are distinct and different in terms of available ecological niches, climate, geographical features, predators, prey, and pathogens, the isolated subpopulations will be subject to different selection pressures leading to different phenotypes. If the physical separation between the populations persists over a sufficient period of time, the populations will diverge. Divergence will be influenced by the genetic variations (alleles) present in the founding populations and the new mutations that arise. The end result will be populations adapted to ecological niches that may well differ from the niche occupied by the parental population. For example, if the parental population was a generalist, occupying a broad range of niches, the subpopulations may occupy more specialized niches.

Consider the situation with finches (honeycreepers) found in the Hawai'ian islands.[151] Derived from an ancestral founder population, these birds have adapted to a number of highly specialized niches. Their specializations give them a competitive edge with respect to one another when feeding off a particular type of flower. As they specialized, however, they become more dependent upon the continued existence of their host flower type (→). It is a little like the fungus that can only grow on one place on a particular beetle. The drive to occupy a particular ecological niche can lead to ecological vulnerability. If the niche disappears, species highly adapted to it may not be able to compete effectively with species adapted to other niches, leading to its extinction.[152]



It is a sobering fact that greater than ~98% of all species that have or now live on Earth are estimated to be extinct. As an example, which of the honeycreepers illustrated above would you expect be most likely to become extinct in response to environmental changes and why?[153] In a complementary way, the migration of a species into a new environment may produce a range of effects on the existing (native) species present.[154] An organisms' influenceon its environment can be complex. A profound example is provided by the appearance, early in the history of life on Earth, of photosynthetic organisms that released molecular oxygen (O_2) into the atmosphere as a waste product. Because of its chemical reactivity, the accumulation of molecular oxygen led to loss of some ecological niches and the creation of new ones. The recent anthropogenic increase in atmospheric CO_2 is another example. While dramatic, similar events occur on more modest levels all the time. It turns out that extinction is a fact of life – at the same time, life has continued and diversified in an uninterrupted manner for over ~4,000,000,000 years.

Gradual or sudden environmental changes, ranging from the activity of the sun, to the drift of continents, and the impacts of meteors and comets can lead to the disappearance of existing ecological niches and the appearance of new ones. For example, the collision of the continents with one another led to the formation of mountain ranges and regions of intense volcanic activity, both of which can influence climate and the connectedness of populations. There have been periods when Earth appears to have been completely or almost completely frozen over and periods when the planet was ice free.[155] Such geological processes are active today, with the Atlantic ocean growing (slowly) wider and the Pacific ocean shrinking, the splitting of Africa along the Great Rift Valley, and the ongoing collision of India with the rest of Asia. As continents move and sea levels change, organisms that evolved on one continent may be able to migrate into another. All of these processes combine to lead to extinctions, which open ecological niches for new organisms, and so it goes.

> *As the Red Queen said to Alice ... "Here, you see, it takes all the running you can do to keep in the same place"*
>
> *-Lewis Carroll, Though the Looking Glass*

By this point you may have come to appreciate the fact that evolution never actually stops. Aside from various environmental factors, each species is part of the environment of other species. Changes in one species can impact others as the selective landscape changes. A particularly obvious example is the interrelationship between predators, pathogens, and prey. Which organisms survive to reproduce will be determined in part by their ability to avoid predators or avoid or recover from infection. Certain traits may make their prey more or less likely to avoid, elude, repulse, discourage, or escape a predator's attack. As the prey population evolves in response to a specific predator or pathogen, these changes will impact the predator or pathogen, which will also have to adapt. This situation is often call the Red Queen hypothesis (→), and it has been invoked as a major driver for the evolution of sexual reproduction, which we will consider in greater detail as we go on.[156]

References

[150] An interesting exception occurs in some plants (which can self-fertilize), where there are instances of new species forming in one generation due to changes in ploidy: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2442920/

[151] Hawaiian honeycreepers and their tangled evolutionary tree

[152] A great video of organisms that have survived (often with human help) the extinction their partners: The Ghosts of Evolution: Nonsensical fruit, missing partners, and other ecological anachronisms

[153] The Perils of Picky Eating: Dietary Breadth Is Related to Extinction Risk in Insectivorous Bats

[154] Humans spread through South America like an invasive species

[155] One “snowball Earth” period appears to have been involved in the emergence of macroscopic multicellular life.

[156] Running with the Red Queen: the role of biotic conflicts in evolution. Another interesting example has been interplay between snake venoms and the resistance of prey animals to that venom. See How Snake Venom Sparked An Evolutionary Arms Race

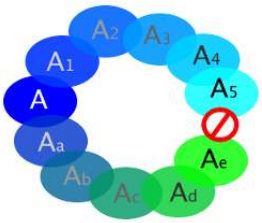
---

This page titled 3.19: Mechanisms of speciation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.20: Mechanisms of reproductive isolation

Think about a population that is specializing to fill a particular ecological niche. What is the effect of cross breeding with a population that is, perhaps, on a path to adapting to another niche? Most likely the offspring will be poorly adapted to both niches. This leads to a new selective pressure, selection against cross-breeding. Even small changes in a particular trait or behavior can lead to significant changes in mating preferences and outcomes. Consider Darwin's finches or Hawai'ian honeycreepers. A key feature that distinguishes these various types of birds is the size and shapes of their beaks. These adaptations represent both the development of a behavior – a behavioral preference to seek food from particular sources and the physical traits needed to harvest that food source effectively. Organisms have to display the behavior, even if it is in a primitive form, to make selection of the physical trait beneficial. This is a type of positive feedback loop, where behavioral and physical traits are closely linked. You can ask yourself, thinking about the ancestor of giraffes, could a long neck have evolved if members of the ancestral population did not choose to eat the higher leaves of trees?

Back to finches and honeycreepers. Mate selection in birds is often mediated by song, generally males sing and females respond (or not). As beak size and shape changes, the song produced also changes.[157] This change is, at least originally, an unselected trait that accompanies the change in beak shape. It can become a selected trait if females recognize and respond to songs more like their own. This would lead to preferential mating between organisms with the same trait (beak shape). Over time, this preference could evolve into a stronger mating preference, until it becomes a reproductive barrier between organisms adapted to different ecological niches.[158] Similarly, imagine that the flowers that a particular subpopulation feeds on open and close at different times of the day. This could influence when an organism is active and sexually receptive. You can probably generate your own scenarios in which a behavioral or physical trait has an influence on reproductive preferences and success. If a population is isolated from others, such effects may develop but are initially irrelevant; they become important only when two closely related but phenotypically distinct populations come into contact. Now matings between individuals in two different populations, sometimes termed hybridization, can lead to offspring poorly adapted to either niche. The result is a selective pressure to minimize hybridization. Again, the reproductive isolation of two populations can arise spontaneously, such as when two populations mate at different times of the day or the year or respond to different behavioral queues, such as mating songs. Traits that enhance reproductive success by reducing the chance of detrimental hybridization will be selected. The end result? what is known as reproductive isolation.[159] As reproductive isolation grows, what was once one species becomes two. A number of different mechanisms ranging from the behavioral to the structural and the molecular are involved in generating reproductive isolation. Behaviors may not be "attractive," genitalia may not fit together,[160] gametes may not recognize and fuse with one another, or embryos might not be viable - there are many possibilities.



Ring species: Ring species demonstrate a version of allopatric speciation. Imagine populations of the species A. Over its geographic range there exist a number of subpopulations. Subpopulations (A_1 to A_5) and (A_a to A_e) have limited regions of overlap with one another and where they overlap they interbreed successfully (→). But populations A_5 and A_e may no longer interbreed successfully – are these populations separate species? In this case, there is no unambiguous answer (and we have to get used to the idea of ambiguity). In part this ambiguity is a basic biological trait. Populations are continuous over time, but individuals within a population vary, and this variation leads to evolutionary change. In the real world, "intact" ring species are unlikely; over time the links between the various subpopulations will be broken and one or more species may arise. Consider the black bear *Ursus americanus*. Originally distributed across all of North America, its distribution is now much more fragmented. Isolated populations are free to adapt to their own particular environments and migration between populations is limited. Clearly the environment in Florida is different from that in Mexico, Alaska, or Newfoundland. Different environments will favor different adaptations. If, over time, these populations come back into contact with one another, they might or might not be able to interbreed successfully - reproductive isolation may have occurred and one species may become many.

References

[157] A good background article on Darin's finches and speciation is here: Sisyphean evolution

[158] Beaks, Adaptation, and Vocal Evolution in Darwin's Finches & Vocal mechanics in Darwin's finches: correlation of beak gape and song frequency

[159] Beak size matters for finches' song: http://news.nationalgeographic.com/news/2004/08/0827_040827_darwins_finch.html

[160] Causes and Consequences of Genital Evolution: http://icb.oxfordjournals.org/content/early/2016/09/13/icb.icw101.abstract

This page titled 3.20: Mechanisms of reproductive isolation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.21: Sympatric speciation

While the logic and mechanisms of allopatric speciation are relatively easy to grasp (we hope), there is a second type of speciation, known as sympatric speciation. Its mechanism(s) were originally more controversial. Sympatric speciation occurs when a single population of organisms splits into two reproductively isolated communities within the same physical region. How could this occur? What stops (or inhibits) the emerging sub-populations from inbreeding? How might these subpopulations become reproductively isolated? A number of plausible mechanisms have been identified. One involves host selection.[161] In host selection, animals (such as insects) that feed off a specific host may find themselves reproducing in distinct zones associated with their hosts. For example, organisms that prefer blueberries may mate in a different place, time of day, or time of year than those that prefer raspberries. There are blueberry- and raspberry-specific niches, and organisms that specialize to one or the other may have a reproductive advantage when they restrict themselves to that food source. Through a process of disruptive selection (see above), organisms that live primarily on one particular plant (or part of a plant) can be subject to different selective pressures. Reproductive isolation will enable the populations to "stay evolutionarily focussed" and so adapt more rapidly. Mutations that reinforce an initial, perhaps weak, mating preference can lead to reproductive isolation. One population becomes two distinct, reproductively isolated populations, one species has become two.

Questions to answer:

32. What factors would lead to positive selection for traits that associated with reproductive isolation?
33. How are sympatric and allopatric speciation the same and how do they differ?
34. What is the role of behavior in the evolution of a trait (e.g. the giraffe's neck).

Questions to ponder:

- What would justify assigning two related asexual organisms to different species?

- How might you decide whether an organism, identified through fossil evidence, was part of a extant species?

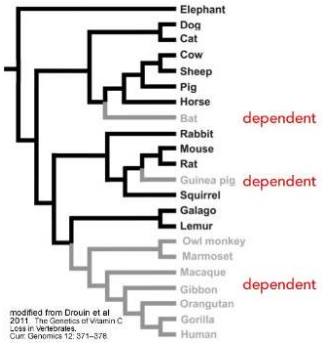
References

[161] Sympatric speciation by sexual selection & Sympatric speciation in phytophagous insects: moving beyond controversy?

This page titled 3.21: Sympatric speciation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

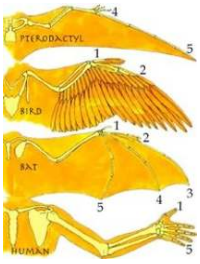
3.22: Signs of evolution- homology and convergence

When we compare two different types of organisms we often find traits that are similar. On the basis of evolutionary theory, these traits can arise through either of two processes: the trait could have been present in the ancestral population that gave rise to the two species or the two species could have developed their versions of the trait independently. In this latter case, the trait was not present in the most recent common ancestor shared by the organisms. Where a trait was present in the ancestral species it is said to be a homologous trait. If the trait appeared independently in the two lineages, it is known as an analogous trait that arose through convergent evolution.

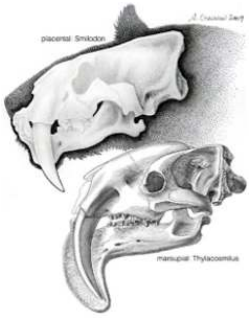


Consider the trait of vitamin C dependence found in primates discussed previously. Based on a number of lines of evidence, we conclude that the ancestor of all Haplorrhini primates was vitamin C dependent; vitamin C dependence in Haplorrhini primates is a homologous trait. On the other hand Guinea pigs (*Cavia porcellus*), which are in the order Rodentia, are also vitamin C dependent, but other rodents are not (→).[162] It is estimated that the common ancestor of primates and rodents lived more than ~80 million years ago, well before the common ancestor of the Halporrhini. Given that most rodentia are vitamin C independent, we can assume that the common ancestor of the rodent/primate lineages was itself vitamin C independent. We therefore conclude that vitamin C dependence in Guinea pigs and Halporrhini (and bats) are analogous traits, they arose as the result of independent events. If we looked at the molecular details, we would not be surprised to discover different mechanisms (different genomic changes) leading to vitamin C dependence in the two groups.

As we consider traits in detail, we have to look carefully, structurally and molecularly, that is, directly at the genotype, to determine at least tentatively whether they are homologous or analogous - the result of evolutionary convergence or ancestry. Consider the flying vertebrates. The physics of flight, and many other behaviors that organisms perform, are constant. Organisms of similar size face the same aerodynamic and thermodynamic constraints. In general there are only a limited number of workable solutions to deal with these constraints. Under these conditions different populations in a position to exploit the benefits of flight will, through the process of variation and selection, end up with structurally similar solutions. This process is known as convergent evolution. Convergent evolution occurs when only certain solutions to a particular problem are evolutionarily accessible.



Consider the wing of a pterodactyl, an extinct flying reptile, a bird, and a bat, a flying mammal (←). These organisms are all tetrapod (four legged) vertebrates. Their common ancestor had a structurally similar forelimb, so their forelimbs are clearly homologous. This evolutionary adaptation, using the forelimbs for flight, began from a structurally similar starting point. But most tetrapod vertebrates do not fly, and forelimbs have become adapted to many different functions. An analysis of tetrapod vertebrate wings indicates that each took a distinctly different approach to generating wings. In the pterodactyl, the wing membrane is supported by the 5th finger of the forelimb, in the bird by the 2ndfinger, and in the bat, by the 3rd, 4th and 5th fingers. The wings of pterodactyls, birds, and bats are analogous structures, while their forelimbs are homologous.



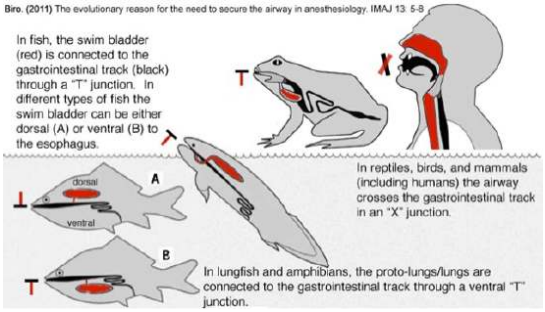
As another example of evolutionary convergence consider teeth. The use of a dagger is an effective solution to the problem of killing another organism. Variations of this solution have been discovered or invented independently many times. Morphologically similar dagger-like teeth have evolved independently from ancestors without such teeth, in a number of distinct lineages. Consider, the placental mammal *Smilodon* and the marsupial mammal *Thyacosmilus* (→); both have similarly-shaped highly elongated canine teeth. Marsupial and placental mammals diverged from a common ancestor ~160 million years ago. This common ancestor, like most mammals, appears to have lacked dagger-like teeth. While teeth are a homologous feature of *Smilodon* and *Thyacosmilus*, elongated dagger-like teeth are analogous structures, the result of convergent evolution.

Recognizing phylogenic relationships: A major challenge when trying to determine a plausible relationship between organisms has been to distinguish homologous from analagous (convergent) traits. Homologous traits, known as synapomorphies, are the basis of placing organisms together within a common group. In contrast, convergent traits are independent solutions to a similar problem, and so are irrelevant when it comes to defining evolutionary relationships. It is, however, also true that evolution can lead

to the loss of traits; this can confuse or complicate the positioning of an organism in a classification scheme. If a trait, whether an enzyme, an eye, or the lack of an eye does not contribute to an organism's reproductive success it will not be selected for. On the other hand, if it is expensive to build, but has no useful function, its loss may be selected for. As organisms adapt to a specific environment and lifestyle, traits once useful can become irrelevant or distracting, and may be lost. A classic example is the reduction of hind limbs during the evolution of whales [↓]. Another is the loss of eyes often seen as populations adapt to environments in which light is absent. The most dramatic cases of loss involve organisms that become obligate parasites of other organisms. In many cases, these parasitic organisms are completely dependent on their hosts for many essential functions. This dependence allows them to become "simplified" even though they are, in fact, highly evolved. They can lose genes as they become dependent upon the host. The loss of traits can itself be an adaptation if it provides an advantage to organisms living in a particular environment. This fact can make it difficult to determine whether an organism is primitive (that is, retains ancestral features) or highly evolved.



Evolution is an ongoing experiment in which random mutations are selected for based on their effects on reproductive success. Various non-adaptive processes can impact evolutionary trajectories. The end result is that adaptations are based on past selective pressures and are rarely perfect and may have become outdated, if the environment the organisms live in changes. One wants to keep this in mind when one considers the differences associated with living in small groups in a pre-technological culture on the African savannah and living in New York City, Nairobi, or Singapore. Evolution is not a designed process that reflects a predetermined goal but is shaped by an organism's history and involves responses to past and current constraints and opportunities - it is a type of tinkering in which selective and non-selective processes interact with pre-existing behaviors and structures which together with the cost and benefits associated with various traits and their effects on reproductive success.[163] What evolution can produce depends on the alleles present in the population, or those generated by mutation, and the current form of the organism. Not all desirable phenotypes may be accessible from a particular genotype; even if they are, the cost of attaining a particular adaptation, no matter how beneficial, may not be repaid by the reproductive advantage it provides within a population.



As an example, our ability to choke on food could be considered a serious design flaw, but it is the result of the evolutionary path that produced us, a path that led to the crossing of our upper airway (leading to the lungs) and our pharynx (leading to our gastrointestinal system). That is why food, lodged in the airway, can cause choking or death. It is possible that the costs of a particular "imperfect" evolutionary design are offset by other advantages (→). For example, the small but significant possibility of death by choking may, in an evolutionary sense, be worth the ability to make the more complex sounds (speech) involved in social communication.[164] Or there may be no accessible evolutionary path available to "fix it".

As a general rule, evolutionary processes generate structures and behaviors that are as good as they need to be or can be in order for an organism to effectively exploit a specific set of environmental resources and behaviors, and to out reproduce its neighbors. If being better than good enough does not enhance reproductive success, it will not be selected for, and variations in that direction will be lost, particularly if they come at the expense of other important processes or abilities.

It is worth noting that we are always dealing with an organism throughout its life cycle. Different traits can have different benefits at different developmental stages. Being cute can have important survival benefits for babies and toddlers but may be less useful in a faculty meeting or corporate board room. A trait that improves survival during early embryonic development or enhances reproductive success as a young adult can be selected for even if it produces negative effects on older, post-reproductive individuals. Since the probability of death by accident or disease, increases with age, selection for traits that benefit the old will inevitably be weaker than selection for traits that effect the young. That said the presence of the old, for example, grandparents (bubbies and babas), may positively influence the reproductive success of the young. Teaching and babysitting come to mind. Of course survival and fertility curves may change in response to changing environmental factors, altering selective pressures. In fact, lifespan itself is a selected trait, since it is the population not the individual that evolves.[165] In this light, while most large mammals have long lifespans, a number of large and complex invertebrates, such as squid, octopus, and cuttlefish have surprisingly short lifespans.[166]

We see the evidence for various evolutionary compromises all around us.[167] They explain the limitations of our senses, as well as our tendency to get backaches, need hip-replacements,[168] and our susceptibility to diseases and aging.[169] For example, the design of our eyes leaves a blind spot in the retina. Complex eyes have arisen a number of times during the history of life, apparently independently, and not all have such a blind spot - a blind spot is not a necessary feature of a complex eye. We have adapted to this retinal blind spot through the use of saccadic eye movements, an easier fix than rebuilding the eye from scratch, something likely to be impossible evolutionarily. An intelligently designed human eye, that is, an eye designed from scratch would presumably not have such an obvious design flaw, but given the evolutionary path that led to the vertebrate eye, it may simply have been impossible to “back up” and fix this flaw. More to the point, since the vertebrate eye works well, there is no apparent reward in terms of reproductive success associated with removing the blind spot. This is a general rule: current organisms work, at least in the environment that shaped their evolution. Over time, organisms that diverge from the current optimal, however imperfect, solution will be at a selective disadvantage. The current vertebrate eye is maintained by stabilizing selection. The eyes of different vertebrates differ in their acuity, basically how fine a pattern of objects they can resolve at what distance, and sensitivity, what levels and wavelengths of light they can perceive. Species have eyes, and their connections to the brain, adapted for their specific ecological niche. For example, an eagle sees details at a distance four to five times as far as the typical human. Why? Because such visual acuity is useful in terms of the eagle’s life-style, whereas excessive visual detail might result in non-useful distractions in humans.[170]

References

[162] see Drouin et al., 2011. "The genetics of vitamin C loss in vertebrates."

[163] Evolutionary tinkering: Jacob 1977

[164] How the Hyoid Bone Changed History: http://www.livescience.com/7468-hyoid-bone-changed-history.html

[165] Methusaleh's Zoo: clues for extending human health span & Why Men Matter: Mating Patterns & Evolution of Lifespan

[166] As described in Peter Godfrey-Smith’s Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness

[167] Wikipedia: Evidence of common descent

[168] Hip pain may be 'hangover from evolution’: http://www.bbc.com/news/health-38251031

[169] How Bipedalism Arose

[170] What If Humans Had Eagle Vision?

This page titled 3.22: Signs of evolution- homology and convergence is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.23: Homologies provide evidence for a common ancestor

The more details two structures share, the more likely they are to be homologous. In the 21st century molecular methods, particularly (relatively) inexpensive genome (DNA) sequencing, have made it possible to treat gene sequences and genomic organization as traits that can be compared quantitatively. Detailed analyses of many different types of organisms reveals the presence of a common molecular signature that strongly suggests that all living organisms share large numbers of homologies, which implies that they are closely related - they share a common ancestor. These universal homologies range from the basic structure of cells to the molecular machinery involved in energy capture and transduction, information storage and utilization. All organisms

- use double-stranded DNA as their genetic material;

- use the same molecular systems to access the information stored in DNA;

- express that information initially in the form of RNA molecules;

- use a common genetic code, with a few variations, and messenger RNAs (mRNAs) to specify

- the sequence of polypeptides (proteins);

- use ribosomes and transfer RNAs (tRNAs) to translate the information stored in mRNAs into polypeptides; and

- share common enzymatic (metabolic) pathways and structures (lipid-based boundary membranes).

Questions to answer:

35. How would you decide whether a trait (or a gene) in two different species is homologous or analogous?
36. Describe a scenario in which the loss of a trait or a gene is beneficial?
37. Describe a scenario in which the simplification of a complex organism would be selected for?

This page titled 3.23: Homologies provide evidence for a common ancestor is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

3.24: Anti-evolution arguments

The theory of evolution has been controversial since its introduction largely because it deals with issues of human origins and behaviors, our place in the Universe, life and its meaning. Its implications can be disconcerting, but many observations support the fact that all organisms on Earth are the product of evolutionary processes and these processes are consistent with what we know about how matter and energy behave. As we characterize the genomes of diverse organisms, we see evidence for these interrelationships, observations that non-scientific (theological) models would never have predicted and do not explain. That evolutionary mechanisms have generated the diversity of life and that all known organisms found on Earth share a common ancestor is as well-established as the atomic structure of matter, the movement of Earth around the Sun, and the solar system around the Milky Way galaxy. The implications of evolutionary processes remain controversial, but not the fact of evolution. We would argue that religions and other belief systems that deny the evolutionary relationships between organisms, and the role of evolutionary mechanisms in shaping organisms, including humans, run the risk of making themselves look ridiculous, at least in terms of data-based (scientific) discussions.[171] On the other hand science (and evolution theory) have little to say on how we should behave, what it means to be moral, basically a good person, or why being an excessively selfish, unfeeling, narcissist person is bad.

> *Scientific knowledge is a body of knowledge of varying degrees of certainty-some most unsure, some nearly sure, but none absolutely certain … Now we scientists are used to this, and we take it for granted that it is perfectly consistent to be unsure, that it is possible to live and not know. - Richard Feynman.*
>
> *...it is always advisable to perceive clearly our ignorance.*
> *– Charles Darwin*

This page titled 3.24: Anti-evolution arguments is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

4: Social Evolution and Sexual Selection

4.0: Introduction
4.1: Selecting social (cooperative) traits
4.2: Quorum Sensing
4.3: Active (altruistic) cell death
4.4: Inclusive fitness, kin and group selection, and social evolution
4.5: Group selection
4.6: Defense against social cheaters
4.7: The appearance of multicellular organisms
4.8: Origins and implications of sexual reproduction
4.9: Sexual dimorphism
4.10: Sexual Selection
4.11: Curbing runaway selection

This page titled 4: Social Evolution and Sexual Selection is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.0: Introduction

In which we consider how unicellular organisms can and do cooperate with one another and how cooperation led to the appearance of multicellular organisms composed of distinct cell types. Similar evolutionary mechanisms have produced a range of cooperative (social) behaviors as well as opportunities for cheating and adaptations to defend against cheaters. One important social behavior is sexual reproduction; we consider its effects on organisms and their evolution.



The naturalist Ernst Mayr (1904-2005) stressed the differences in thinking in biology compared to physics and chemistry. The history of an electron, an atom, or a molecule is irrelevant to its physical and chemical properties. Each carbon isotope atom, for example, is identical to all others - one could be replaced by another and you could never, in practice or in theory, be able to tell the difference. In contrast, each organism, how it is built, how it behaves, how it interacts with other organisms, and the possible futures of its descendants emerges from evolutionary and adaptive processes involving both selective, non-selective, and non-adaptive processes stretching back ~4 billion years. This history encompasses an unimaginable number of individually unpredictable events (mutations, noisy gene expression, accidents and environmental insults, isolated and merging populations). Because of its molecular and cellular complexity and distinct history, each organism is unique and distinguishable from all others. [172]

In biology, we talk about organisms but this is often too simplistic. When does an organism begin? What are its boundaries? The answers can seem obvious, but then again, perhaps not. When a single-celled organism reproduces it goes through some form of cell division, and when division is complete, one of the two organisms present is considered a new organism and the other the old one, but it may not be clear which is which. In fact, both are old, both reflect a continuous history stretching back to the origin of life. When an organism reproduces sexually, the new organism arises from the fusion of two pre-existing cells and produces a cell that forms the next generation. If we trace the steps back in time from any modern organism, we find no clear line between the different species. When, exactly, did humans (*Homo sapiens*) emerge from pre-humans, or modern birds from their dinosaurian progenitors? The answer is arbitrary, since cellular and organismic continuity is never interrupted - life does not start, stop, and start again, it continues until it stops irreversibly in death. Because of superfecundity, selection, and speciation, it can generate branches, distinct lineages.

We typically define the boundaries of an organism in physical terms, but organisms interact with one another, often in remarkably complex ways. For example, some unicellular organisms live so closely together that it is impossible for them to live apart.[173] A dramatic example of this type of situation are what are known as eusocial organisms. While many of us are familiar with the social structure of ants and bees, fewer (we suspect) are aware of naked (*Heterocephalus glaber*) and Damaraland (*Cryptomys damarensis*) mole rats. In these organisms reproduction occurs at the group level; only select females, termed queens, produce offspring. Most members of the group are smaller, effectively sterile female workers. A few males are present, they inseminate the queen.[174] So what, exactly, is the organism? individuals or the social group that makes it up? From an evolutionary perspective, selection is occurring at a social level as well as at a cellular and organismic level.

Consider yourself and other multicellular organisms (animals and plants). Most of the cells in your body, known as somatic cells, do not directly contribute to the next generation; they cooperate to ensure that a subset of cells, known as germ line cells (sperm and eggs), have a chance to form new organisms. The somatic cells sacrifice their reproductive potential so that germ line cells can produce a new organism. They are the sterile workers to the germ line's queen. The term "sacrifice" in the context of a multicellular organism may seem weird, and anthropomorphic, since both germ line and somatic cells are part of a single organism, and it is the organism, rather than the cells that compose it, that is the biologically meaningful object. Similarly, in a eusocial organism, it is the social group that "matters".

We find examples of social behavior at the level of unicellular organisms as well, and even in viruses.[175] Think about a unicellular organism that divides but in which the offspring stick together. As this process continues, we get what we might term a colony. Is such a clump of cells one or many organisms? If all of the cells within the group can produce new cells, and so new colonies, we consider it a colony of organisms. So where does a colony of organisms turn into a colonial organism? The distinction can be fuzzy,

but we can adopt a set of guidelines.[176] One criterion would be that a colony becomes an organism when it displays traits that are more than just sticking together or failure to separate, that is, when it acts more like a coordinated group. This involves the differentiation of cells; certain cells specialize to carry out specific roles. Producing the next generation of organisms is one such specialized functional role. Other cells may become specialized for feeding or defense, they support the process of reproduction, in part by enabling the resulting organism to occupy a particular ecological niche. The differentiation of cells from one type to another within a multicellular aggregate has moved a colony of organisms to a multicellular organism. What is tricky about this process is that originally reproductively competent cells have given up their ability to reproduce, and are now acting, in essence, to defend or support the cells that do reproduce. This is a social event and is similar to the behavior of naked mole rats. Given that natural selection acts on reproductive success, one might expect that the evolution of this type of cellular and organismic behavior might be selected against or impossible to produce, yet multicellularity and social interactions have arisen independently many times during the history of life on earth.[177] Is this a violation of evolutionary theory or do we have to get a little more sophisticated in our thinking?

Questions to answer:

38. What features (behaviors) are important when defining an organism? Does your definition include both uni- and multi-cellular organisms?
39. How would you characterize humans in terms of sociality?

References

[172] While these events obey physical and chemical laws, in practice, the number of variables involved makes them unpredictable. At the same time, because they are based on natural processes, when we consider large numbers of such events, they become predictable. So while the mutation rate is predictable, which mutations occur in which organism is not.

[173] Cultured Asgard Archaea Shed Light on Eukaryogenesis

[174] An Introduction to Eusociality: http://www.nature.com/scitable/knowledge/library/an-introduction-to-eusociality-15788128

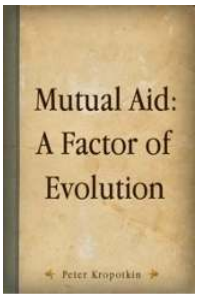
[175] The secret social lives of viruses

[176] A twelve-step program for evolving multicellularity and a division of labor

[177] The Origins of Multicellularity

This page titled 4.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

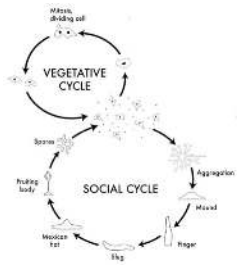
4.1: Selecting social (cooperative) traits





So how does evolution produce multicellularity? To answer this question, we need to approach evolutionary processes more broadly. The first new idea we need to integrate into our theoretical framework is known as inclusive fitness, which is sometimes referred to as kin selection. For the moment, let us think about traits that favor the formation of a multicellular organism - later we will consider traits that have a favorable effect on other, related organisms. Finally, we consider social situations in which behaviors have become fixed to various extents and are extended to strangers. Humans can, but do not always, display such social (moral) behaviors. The importance of mutual aid in evolutionary thinking, that is the roles of cooperation, empathy, and altruism in social populations, was emphasize by the early evolutionary biologist and anarchist (Prince) Peter Kropotkin (1842–1921)(→).

All traits can be considered from a cost-benefit perspective. There are costs (“c”) in terms of energy needed to produce a trait and risks associated with expressing the trait, and benefits (“b”) in terms of the trait’s effects on reproductive success. To be evolutionarily preferred, that is, "selected for", the benefit b must be greater than the cost c, that is b > c. Previously we tacitly assumed that costs and benefits applied to one and the same organism, but when we consider cooperative (social) behaviors and traits, this is not the case. We can extend our thinking as follows: assume that an organism displays a trait. That trait has a cost to produce and yet may have little or no direct benefit to the organism that produces it; it may even harm it. Now let us assume that this same trait benefits neighboring organisms, a situation similar to the fireman who risks their life to save a child from a burning building. How could a biological system (the fireman), the product of evolutionary processes, display this type of self-sacrificing behavior? The answer (at least in part) is social systems.



An example of this type of behavior is displayed by the social amoebae *Dictyostelium discoideum*.[178] Dictyostelium has a complex life style that includes a stage in which unicellular amoeba-like organisms crawl around in the soil eating bacteria, growing and dividing. In this "vegetative" cycle the cells divide asexually (as if vegetables don’t have sex). If or rather when the environment turns hostile, theseamoeba sense the change and secrete small molecules that influence their own and their neighbor’s behaviors. The is amoeba migrating toward one another (←), forming aggregates of thousands of cells.

Now something amazing happens: these aggregates begin to migrate around as multicellular “slugs”. They respond to environmental signals. They move toward light, a behavior known as positive phototaxis, and then settle down and undergo a rather coordinated process of cellular differentiation.[179] All through the aggregation and slug migration stages, the original amoeboid cells remain distinct. Upon differentiation ~20% of the cells in the slug specialize to form non-dividing "stalk" cells that go on to die through a process known as programmed cell death or apoptosis. Before they die the stalk cells act together, through changes in their composition and shape, to lift the non-stalk cells out of the soil. The non-stalk cells then go on to form spores, specialized cells that can survive harsh conditions. When the spores are released they float in the air and can be transported by the wind and other mechanisms into new environments. The stalk cells "sacrifice" themselves so that non-stalk cells can form spores.

When a spore lands in a new, and hopefully hospitable environment, it converts back into a motile, unicellular amoeba and begins to feed and reproduce vegetatively. The available evidence indicates that within the slug the “decision” on whether a cell will form a stalk or a spore cell is not pre-determined, it arises from stochastic molecular level processes. The decision is not based on genetic (genotypic) differences - two genetically identical cells may both form spores, both stalk cells, or one might become a stalk and one a spore cell within the differentiating slug.[180]

References

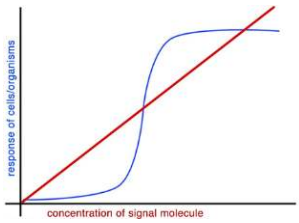
[178] Molecular phylogeny and evolution of morphology in the social amoebas & A Simple Mechanism for Complex Social Behavior. A nice video here: http://youtu.be/bkVhLJLG7ug

[179] Behavior of cellular slime molds in the soil: http://www.mycologia.org/content/97/1/178.full

[180] A type of behavior that occurs in a number of organisms: see From cell differentiation to cell collectives

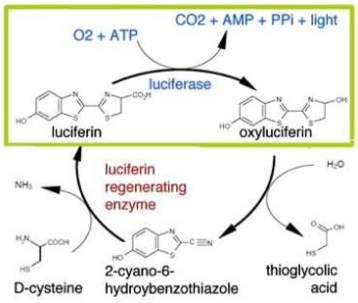
This page titled 4.1: Selecting social (cooperative) traits is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.2: Quorum Sensing



Community behaviors initiated at the unicellular level often involves a process known as quorum sensing. Through this process organisms estimate the density (number of individuals per volume) in their immediate environment. Each individual secretes specific molecules; they also respond to that molecule through specific receptors. The organisms' response is dependent on the signaling molecule's extracellular concentration. More importantly, the response (blue line →) is non-linear; it displays a "threshold" effect (red line → linear response - X axis signal concentration, Y-axis response). Below the system's "threshold" concentration there is little if any cellular response; above the threshold concentration the cells respond fully. When cells or organisms are present at a low density, the signal molecule's concentration is below the threshold concentration. As the density of organisms increases, the molecule's concentration increases and can exceed the threshold concentration; that is when interesting things begin to happen. There can be changes in cellular behaviors, generally associated with changes in gene activity.[181] We can think of non-linear (threshold) responses as a strategy to avoid over-reacting to minor environmental fluctuations or when there are not enough organisms around to produce a useful cooperative response. Only when the signal gets high enough does the system respond. The threshold concentration reflects the signaling molecule' binding affinity to its target receptor, and other factors influencing molecular stability and diffusion.

A classic example of threshold behaviors is provided by the light emitting marine bacteria *Vibrio fischeri*. These bacteria colonize a dedicated organ of the Hawai'ian bobtail squid (*Euprymna scolopes*) shortly after the squid "hatch".[182] The colonization process has a number of steps. The bacteria enter and adhere to an opening leading the light emitting organ. The bacteria begin to divide; as their numbers increase they pass a threshold and begin to secrete molecules that form of gooey matrix, known as a biofilm. The bacteria swim through the biofilm toward signals released by the light organ cells (a process known as chemotaxis). The result is that bacteria enter and colonize the light organs.



Within the light organs the bacteria emit light through a reaction system involving the molecules luciferin and O_2 (→). The light emitting reaction is catalyzed (that is, sped up) by the enzyme luciferase, a protein catalyst. The luciferase protein is encoded by a bacterial gene. Its original role has been proposed to be in the “detoxification of deleterious oxygen derivatives".[183]

A small number of bacteria would emit little light, not enough to be useful, while using energy (all costs, no benefit). The components of the light emitting system are regulated so that they are synthesized only when the numbers of bacteria are high enough to make the emission of light useful, which decreases the cost to benefit ratio. As with *Dictyostelium*, the bacteria use a quorum sensing system. Each bacterium releases a signaling molecule; as the numbers of bacteria increase, the signaling molecule's concentration increases and passes through the response threshold. Receptors on the bacteria are activated leading to the synthesis of the components of the light reaction, luciferase and the proteins involved in the synthesis of luciferin. In the presence of O_2 and ATP (in the bacteria) light is emitted. Nutrients to feed the bacteria are supplied by the squid.

Mechanistically similar systems are involved in a range of processes including the generation of toxins (virulence factors), secreted digestive enzymes, and antibiotics directed against other types of organisms. These are produced when the density of bacteria rises above a threshold concentration. This ensures that the synthesis of biologically costly molecules, such as luciferase and luciferin, occurs only when they are likely to be useful and effective – that is, they are produced at a level high enough to carry out their intended roles. Such high concentrations can only be attained through cooperative behaviors involving many individuals.

Questions to answer:

40. Why (generally) does a quorum signal need to be secreted (released) from the organism?
41. Is a population of bacteria that display quorum sensing behavior a single organism, justify your answer.

Question to ponder:

- How might it impact the social behavior of slime molds if the percentage of spore cells were 1% rather than 80%?
- Why are non-linear responses to stimuli important in biological systems? How might they be achieved

References

[181] Quorum sensing in bacteria: http://www.ncbi.nlm.nih.gov/pubmed/11544353

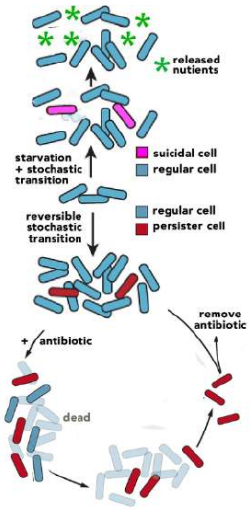
[182] Zink et al (2021). A Small Molecule Coordinates Symbiotic Behaviors in a Host Organ

[183] Experimental evidence for the physiological role of bacterial luciferase: http://www.ncbi.nlm.nih.gov/pubmed/14669913

This page titled 4.2: Quorum Sensing is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.3: Active (altruistic) cell death

A type of behavior you might think impossible for evolutionary processes to produce would be the active and intentional death of a cell or an organism. Yet, such behaviors are surprisingly common in a wide range of systems.[184] The death and release of leaves from deciduous trees in the autumn is an example of a built-in or "programmed" cell death process, known as apoptosis, from the Greek meaning to fall off. Apoptosis plays important roles in the formation of various structures within multicellular organisms, such as the fingers of hands. Programmed cell death also plays a critical role in the development of the immune and nervous systems, important topics beyond our scope here.[185] Programmed cell death is distinct from accidental death, such as occurs when a splinter impales a cell or you burn your skin. Such accidental death leads to what is known as necrosis. In necrosis, cellular contents are spilled out in an uncontrolled manner from the dying cell. The release of cellular debris provokes various organismic defense systems to migrate into the damaged area and (primarily) fight off invading bacteria and remove damaged cells. The swelling and inflammation associated with injury is an indirect result of necrotic cell death. In contrast, apoptotic cell death occurs using well-defined pathways that requires energy to carry out. Cell contents are retained during the process; no inflammatory, immune system response is provoked. Surrounding cells actively remove the remains of the apoptotic cells. Programmed cell death/apoptosis is a tightly controlled process that plays specific and important roles within the context of the organism.



Consider programmed cell death in the context of communities of unicellular organisms. In such systems, programmed cell death can be triggered by environmental stresses together with quorum sensing. In this situation, a subset of cells can stochastically “decide” to undergo cell death by activating a cell death pathway. When a cell dies, its contents are released and can be used by the living cells that remain (→). These living cells gain a benefit, and we would predict that the increase in nutrients will impact their chances of survival and successful reproduction. This strategy works because as the environment becomes hostile, not all cells die at the same time. It also makes no evolutionary sense for an isolated cell/organism to die through programmed cell death, since the release of its nutrients would fail to benefit its (related) neighbors. Instead of dying, better to change into what is known as a “persister”. In such a state the bacterium stops growing and minimizes its use of (and need for) energy (→). In the persister state, a bacterium can survive until the stressor (e.g. an antibiotic, a molecule that leads to the death of susceptible bacteria) disappears from the environment. Such behaviors, death or the adoption of a persister phenotype, occur in groups of genetically identical cells and involve stochastic choices.

So how do cells kill themselves? Many use a similar strategy. They contain what is known as an addiction module, which consists of two genes - the first encodes a toxic molecule. The toxic molecule can kill the cell and is synthesized (expressed) continuously. Many distinct toxic molecules have been identified, so they appear to form analogous rather than homologous systems – they appear to have evolved independently. Now you may wonder how such a gene could exist, how can a cell survive the presence of a lethal toxin. One answer is that the cell contains a second gene that encodes an anti-toxin; the anti-toxin acts on the toxin and inhibits its activity. Within the cell, the toxin-anti-toxin complex forms and the cell survives. So far, so good. But the system has one more wrinkle. The toxin and anti-toxin molecules differ in an important way. The toxin molecule is slowly degraded by molecular systems within the cell; once synthesized it has a long "half-life". In contrast, the anti-toxin molecule is rapidly degraded; it has a short half-life. Under normal conditions the steady state concentration of the anti-toxin, a function of its synthesis and degradation rates, is sufficient to inhibit all of the toxin present. The cell has become addicted to the anti-toxin, which must be made continuously in order to inhibit the toxin and avoid cell death.

What happens if the cell is stressed, either by changes in its environment or infection by a virus? Generally cellular activity, including gene expression and the synthesis of cellular components, such as the anti-toxin, slows or stops. What happens next? The level of the toxin molecule, which has a long half-life, decreases slowly, whereas the level of the short lived anti-toxin drops rapidly. When the level of the anti-toxin falls below that needed to inhibit the toxin, the now active toxin initiates the process of cell death, leading to the release of the dying cell’s components.

In addition to the dying cell "sharing" its resources with its (presumably related) neighbors, programmed cell death can be used as a population-wide defense mechanism against viral infection. One of the key characteristics of viruses is that they must replicate within a living cell. Once a virus enters a cell, it typically disassembles itself and sets out to reprogram the cell’s biosynthetic

machinery in order to generate new copies of the virus. During the period between viral disassembly and the assembly of newly synthesized viruses, infectious virus disappears - it is said to be latent. If the cell kills itself before new viruses are synthesized, it also "kills" (or rather inactivates and eliminates) the infecting virus. By killing the virus (and itself) the infected cell acts to protect its neighbors from viral infection - this can be seen as a form of the altruistic, self-sacrificing behaviors we have been considering. [186]

References

[184] See On the paradigm of altruistic suicide in the unicellular world: http://www.ncbi.nlm.nih.gov/pubmed/20722725

[185] Apoptosis in the nervous system & Apoptosis in the immune system

[186] The evolution of eusociality

This page titled 4.3: Active (altruistic) cell death is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.4: Inclusive fitness, kin and group selection, and social evolution

The question that troubled Darwin and others was, how can evolutionary processes produce this type of social, self-sacrificing behavior? Consider the behavior of bees. Worker bees, who are sterile females, “sacrificed themselves to protect their hives” even though they themselves cannot reproduce, they are sterile.[187] Another example, taken from the work of R.A. Fisher (1890-1962), involved the evolution of noxious taste as a defense against predators. We can assume that the organisms eaten by predators do not directly benefit from this trait, after all, they have been eaten. So how can the trait of “distastefulness” arise in the first place? If evolution via natural selection is about an individual’s differential reproductive success, how are such traits possible? W.D. Hamilton (1936-2000) provided a formal answer, expressed in the equation rb > c. As in our consideration of costs and benefits, “b” stands for the trait’s benefit to the organism and others, “c” stands for the cost of the trait to the individual, while “r” indicates the extent to which two organisms within the population are related and genetically similarity.

Let us think more about what this means. How might active cell death in bacterial cells be beneficial evolutionarily? In this case, reproduction is asexual. The organism’s (cell’s) offspring, and its likely neighbors, are clonally-related to one another, much as the cells of a multicellular organism are related. Aside from occasional mutations, the cells in a clone and within an organism are genetically identical, with DNA molecules that are identical in sequence.[188] We can characterize the degree of relationship through their r value, the coefficient of relationship. In two genetically identical organisms r is close to 1. Two totally unrelated organisms, with minimum possible genotypic similarity would have an r very close to 0.[189] Now let us return to our cost-benefit analysis of a trait’s effect on reproductive success. Each trait has a cost c to the organism that produces it, as well as a potential benefit b in terms of reproductive success. Selection leads to a trait becoming prevalent (frequent or fixed) within a population if b >> c. But this equation ignores the effects of a trait on other related and neighboring organisms. Let us call the benefits to the individual that result from their cooperative/altruistic behavior b_i and the benefits to others/neighbors b_o. To generate our social equation, known as Hamilton’s rule, we consider what is known as the inclusive fitness, which changes b > c into $b_i + r \times b_o > c$. This allows for the selection of social (potentially detrimental) traits provided that their impact on the reproductive success of closely related individuals is increased sufficiently. Given the clonal nature of many microbes, inclusive fitness can be powerful in these organisms. It can also be significant in small populations of sexually reproducing, genetically related organisms.

Inclusive fitness requires a close relationship to the recipient of the beneficial act. How does one individual “know” that it is making a sacrifice for a relative and not an unrelated strangers? As social groups get larger, identifying relatives becomes more and more difficult. One approach is to genetically link the social, altruistic behavior to a visible trait, called a “green beard” trait. The likelihood that an organism will behave socially is linked to the visible trait. The presence of the green beard trait indicates that an organism is "prepared" to "sacrifice" itself for you in the same way you are prepared to sacrifice for it. In some cases, a trait may evolve to such a degree that it becomes part of an interconnected set of behaviors, a type of biosocial moral system.[190]

Once humans developed a brain sufficiently complex to do what it was originally selected for, this complexity may have produced various unintended byproducts. Empathy, self-consciousness, or a tendency to anxiety may not be directly selected for but could be side effects of processes that were. As a completely unsupported but plausible example, the development of a good memory as an aid to hunting might leave us susceptible to nightmares. If empathy is an “unintended” by-product of selective processes, once present they could alter future selection pressures. They might not be easy to evolve away from, particularly if they are mechanistically linked to a trait that is highly valued, that is, selected for. The effects of various genetic mutations on personality and behavior support the idea that such traits may be influenced by one’s genotype.

References

[187] Dugatkin, L.A. 2007. Inclusive Fitness Theory from Darwin to Hamilton

[188] There is an exception to this role involving a subset of the cells of the immune system, but it is not important here.

[189] You are not genetically identical to your parents or siblings (unless you are have an identical twin). You share ~50% of your genetic material with either of your parents and ~25% with your siblings.

[190] Organisms that fail to live by these rules might be sociopaths or suffering from pernicious narcissism.

This page titled 4.4: Inclusive fitness, kin and group selection, and social evolution is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

https://bio.libretexts.org/@go/page/181667

4.5: Group selection

A alternative to inclusive fitness (kin selection) is the concept of group selection. In this scenario, small groups of organisms of the same species are effectively acting as single (perhaps colonial) organism. It is the reproductive success of the group, rather than the individuals within the group, compared to other groups that is the basis of selection. In certain situations, groups that display cooperative and altruistic traits may have a selective advantage over groups that do not. It has been claimed that group and kin selection are mathematically equivalent, even though one occurs between population groups and the other within a population group.[191] The costs of a trait must be offset by the benefits, but now the key factor is membership in a group that tend to be closely related to one another. The life cycle of the bacterium *Myxococcus xanthus* provides an example of this type of behavior. When environmental conditions are harsh, the cells aggregate into dense, 100 μm diameter "fruiting bodies", each containing ~100,000 stress resistant spores. When the environment improves, and nutrients become available, the spores are released en mass and return to active life. They move and feed in a cooperative manner through the release of digestive enzymes that, because they are acting in a quorum mode, can reach high functional levels.[192] A well-coordinated group is expected to have a significant reproductive advantage over a more anarchic collection of individuals.



While their functional roles are different, analogous types of behaviors are seen in flocks of birds, schools (or shoals) of fish, swarms of bees, blooms of algae, and groups of slime mold cells (→).[193] Each of these examples represents a cooperative strategy by which organisms gain a reproductive advantage over those that do not display the behavior. While the original behavior is likely the result of kin selection, in the wild it is possible that different groups (communities) are in competition with one another, and the group(s) that produces the most offspring, that is, the most reproductively successful group will come to dominate.

References

[191] Mathematics of kin- and group-selection: formally equivalent? http://www.ncbi.nlm.nih.gov/pubmed/19929970

[192] Evolution of sensory complexity recorded in a myxobacterial genome: http://www.ncbi.nlm.nih.gov/pubmed/17015832

[193] How Does Social Behavior Evolve?

This page titled 4.5: Group selection is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.6: Defense against social cheaters

Within a social group, such as cooperating microbes[194] or human hunters, we can expect that, through mutation and other effects, cheaters will arise. What is a cheater? Imagine a bacterium within a swarm, a cell in an organism, or an animal in a social group that fails to obey or ignores the rules - it benefits from social cooperation without contributing to it.[195] For example when an individual accepts help from others, but fails to help others. In the case of slime mold aggregates, imagine a cell that always differentiates into a reproductively competent spore and that this trait has a genetic basis. What happens over time? A likely outcome is that when such a cell begins its own clone this group of cels will fail to form a functional or useful stalk. If stalk-dependent spore dispersion is important for long term survival and reproduction, there will be selection for populations with low levels of cheaters.

Multicellular organisms are social systems, composed of cells that have given up their ability to reproduce new organisms in order to enhance the reproductive success of the organism as a whole. In this context cancer can be viewed as an anti-social disease. Cancers typically arise from mutations that lead to a loss of social control. Normally the growth and division of cells is strictly controlled. In cancer, cells begin to divide in an uncontrolled manner; the result is disruption of normal tissue organization. If they become malignant, they migrate away from their site/tissue of origin and colonize other areas of the body, a process known as metastasis. Their uncontrolled growth often lead to damage of multiple organs and the death of the organism.

How can organisms defend themselves against social cheaters? When we think about maintaining a social behavior, there are two general mechanisms: intrinsic and extrinsic policing. Assume that a trait associated with the social behavior is also linked to, or required for, cell survival. In this case, a mutation that leads to the loss of the social trait may lead to cell death (apoptosis). In the context of cancer, normal cells can be considered "addicted to normality". When normality is disrupted they undergo apoptosis. An abnormal cell growing in an uncontrolled manner will undergo apoptosis before it can produce significant damage.[196] For a "pre-cancer" to progress, other mutations must inactivate this normal (wild-type) apoptotic response. The apoptotic process is an intrinsic policeman enforcing social control. It is a little like the guilt experienced by (some) people when they break social rules. True psychopath's loss of social guilt is analogous to the inhibition of apoptosis in response to abnormal behaviors.[197]

There are also extrinsic social control systems, analogous to the presence of external policeman. Mutations associated with the loss of social integration – that is, the transformation of a cell to a cancerous state – can lead to changes in the character or behavior of the cell. Cells of the immune system can recognize these changes as "non-self" and induce the mutated cell's death.[198] Of course, given that tumors occur and kill people, we can assume that there are mutations or phenotypic changes, associated with altered gene expression, that enable tumor cells to circumvent immune system surveillance. One part of the cancerous phenotype is often the loss of normal mutation repair systems. In effect, the mutant cell increases the number of unrepaired mutations, and consequently the genetic variation in the cancer cell population. While many of these variants are lethal, the overall effect is to increase the rate of cancer cell evolution. The result is an arms race. If the cancer is killed by intrinsic and extrinsic social control systems, no disease occurs. If, however, the cancer evolves so as to avoid death induced by these systems, the cancer can progress and spread. As we look at a range of social systems, from cooperating bacteria to complex societies, we see examples of intrinsic and extrinsic control.

References

[194] An interesting read: The stag hunt and the evolution of social structure.

[195] As an example, consider a person who accepts the protection of police and firefighters, but avoids paying their taxes.

[196] Apoptosis in cancer: http://carcin.oxfordjournals.org/content/21/3/485.full

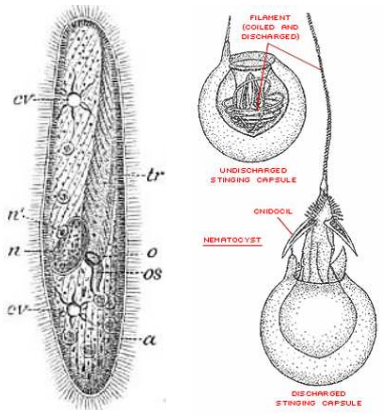
[197] In an age of rampant narcissism and social cheating – the importance of teaching social evolutionary mechanisms.

[198] Immune recognition of self in immunity against cancer & Anti-cancer drugs that reactivate the immune survalience

This page titled 4.6: Defense against social cheaters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.7: The appearance of multicellular organisms

What drove the appearance of multicellular organisms? A number of theoretical and empirically supported and mechanistically plausible models have been proposed. Some suggest that predation was an important driver, either enabling the organisms to become better predators or to avoid predation. In an experimental study, when the small unicellular algae *Chlorella vulgaris* (5 to 6 μm in diameter) was grown together with a unicellular predator *Ochromonas vallescia*, which engulfs its prey. Over time *Chlorella* formed multicellular colonies that *Ochromonas* could not ingest.[199] At this point what we have is more like a colony of organisms than a colonial organism or a true multicellular organism.



The change from a multi-individual colony to a multicellular organism involves cellular specializations, so that different types of cells within the organism come to carry out different functions. The most dramatic specialization involves differentiations between somatic and germ line cells. At the other extreme, instead of producing distinct types of specialized cells to carry out distinct functions, a number of unicellular eukaryotes, such as the protist *Paramecium* (→), have developed complex cells that display specialized behaviors such as directed motility, predation, osmotic regulation, and digestion. Cellular specialization can be carried out even further in multicellular organisms. The stinging cells of jellyfish provide a classic example; highly specialized cells deliver poison to any organism that touches them through a harpoon-like mechanism (←). The structural specialization of these cells can make processes such as cell division impossible and typically a stinging cell dies after it discharges. Presumably, it is simpler to generate a new stinging cell than it is to reset a discharged cell. The production of these new cells involves both cell division and differentiation, which we will consider in more detail later on. While we are used to thinking about individual organisms, the same logic can apply to groups of organisms. The presence of cooperation can extend beyond a single species, leading to ecological interactions in which organisms work together to various degrees to achieve something that would be difficult or impossible to achieve on their own, while maintaining their ability to reproduce.

Based on the study of a range of organisms, much progress has been made in clarifying the origins of multicellular organisms. Such studies indicate that multicellularity has arisen independently in a number of eukaryotic lineages. This indicates that becoming multicellular is a successful way to establish an effective relationship with the environment.

Questions to answer

42. What type(s) of mutation would enable an organism to escape a cell death module?
43. Make a model for the process that could lead to the evolution of social interactions.
44. What factors limit the complexity of a unicellular organism?
45. Is the schooling or herd behavior seen in various types of animals a homologous or an analogous trait?

Questions to ponder:

- What strategies might be used to defend against the effects of cheaters in a population?
- Why is r (the relationship between organisms) never 0.
- What are some of the advantages of multicellularity? What are the drawbacks? Why aren't all organisms unicellular or multicellular?

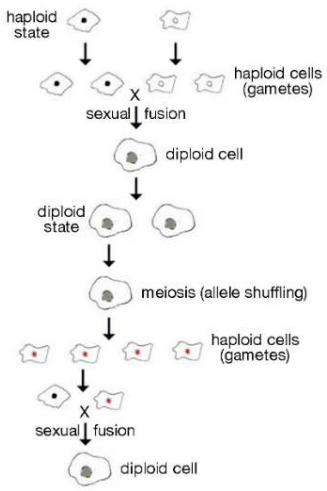
References

[199] Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity

This page titled 4.7: The appearance of multicellular organisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.8: Origins and implications of sexual reproduction

One type of social interaction is sexual reproduction, which involves cooperative interactions between distinct organisms. While two distinct sexes (male and female) is common, it is not universal. Many unicellular eukaryotes are characterized by a number of distinct “mating types”. Typically, sexual reproduction involves the fusion of two specialized cells, known as gametes, of different mating types or sexes. Through multiple mechanisms to be considered later, the outcome of sexual reproduction leads to increased genetic diversity among offspring.



What are the common hallmarks of sexual reproduction? Let us return to the slime mold *Dictyostelium* as an exemplar. We have considered its asexual life cycle, but *Dictyostelium* also has a sexual life cycle. Under specific conditions, two amoeboid cells of different mating types will fuse together to form a single cell. The original cells are haploid (→), meaning that they each have a single copy of their genome. After fusion, the resulting cell has two copies of its genome and is referred to as diploid. This diploid cell then goes through a series of events, known collectively as meiosis, a process that results in the shuffling of genetic material and the production of four haploid cells. The critical point is that the genotypes of the haploid cells that emerge from meiosis are different from the haploid cells that originally fused together. Some organisms can spend a significant amount of time in the haploid state, while others spend most of their lives in the diploid state. You, for example, had a reasonably short haploid stage (as both an egg AND a sperm cell); your diploid stage began when these two cells fused.

The oscillation between haploid and diploid states has some interesting implications. The first is that in the diploid state, there are (generally) two copies of each gene. The different versions of a gene present in a diploid cell can be the same or different. If they are the same, the cell/organism is said to be homozygous for that gene; if they are different, it is heterozygous for that gene. Alleles can have a range of effects on phenotype, from cellular lethality to more subtle effects arising from differences in the activity, localization, stability, specific form, or amount of the gene product. These effects may be influenced by the products of other genes, leading to genetic background effects. In diploid cells the effects of one allele can be masked by the presence the other normal (wild type) allele. Such "maskable" alleles are commonly referred to as recessive. We will return to these topics later on. Where genes are actively expressed and functionally important in the haploid state, which is not always the case, the presence of a lethal allele can lead to death or dysfunction. The presence of an extended haploid phase of an organisms’ life cycle can lead to the elimination of such alleles from the population.

This page titled 4.8: Origins and implications of sexual reproduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.9: Sexual dimorphism

What, biologically, defines whether an organism is female or male, and why does it matter? The question is meaningless in unicellular organisms with multiple mating types. For example, the protozoan *Tetrahymena* has seven different mating types, all of which appear morphologically identical. An individual *Tetrahymena* cell (organism) can mate with another individual of a different mating type but not with an individual of the same mating type as itself. Mating involves cell fusion and so the identity of the parents is lost; the four cells that are produced by the fused cell, through the process of meiosis) are of one or the other of the original mating types.

In multicellular organisms, the parents do not fuse with one another. Rather they produce specialized cells that do. Instead of multiple mating types there are two, male and female. How are male and female defined? The biological answer is simple but its implications can be profound. Sex is defined by the relative size and morphology of the fusing haploid gametes that the organism produces. The larger, generally immobile gamete is termed an egg; an organism that produces eggs is termed female. The smaller, generally motile gamete is termed a sperm; organisms that produce sperm are termed male. At this point, we should note the limits of these definitions. There are organisms that can produce both types of gametes; either at the same time or sequentially. These are known as hermaphrodites, after the Greek gods Hermes (male) and Aphrodite (female). A hermaphroditic organism can self-fertilize. In such cases, males (which produce only sperm) may appear only under certain circumstances. There are organisms that can change their sex, a behavior known as sequential hermaphroditism. In the clownfish *Amphirion ocellaris* (think Nemo), individuals that originally developed as sperm-producing males can, based on environmental cues, change to produce eggs and so become female.[200]

The size and morphological differences between male and female gametes influences the reproductive stakes for the two sexes. Because of the larger size of eggs, females invest more energy in their production (per egg) than a male invests in the production of the much smaller sperm cells. It is therefore relatively more important, from the perspective of reproductive success, that each egg produce a viable and fertile offspring. As the costs to females of generating eggs and rearing their offspring increases, the more important the egg's reproductive success becomes. Because sperm are relatively "cheap" to produce, and because males may make little investment in rearing their offspring, the selection pressure on males can be significantly less than on female. This can result in a conflict of interest between females and males. The "size" of the conflict increases as the disparity in the relative investment per gamete or offspring increases.

Sex-associated reproductive conflicts are an example of evolutionary economics based on an implicit cost-benefit analyses. First there is what is known as the two-fold cost of sex. Each asexual organism can, in theory at least, produce offspring but two sexually



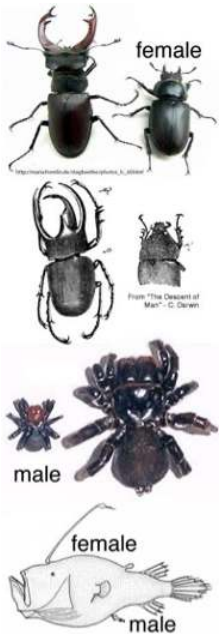
reproducing individuals must cooperate to produce offspring. Other, more specific factors influence an individual's reproductive costs. The cost to a large female laying a small number of small eggs that develop independently is less than that of a small female laying a large number of large eggs. Similarly, the cost to an organism that feeds and defends its young after they are born (that is, leave the body of the female) is larger than the cost to an organism that lays eggs and leaves them to fend for themselves. The costs to a female that raises its young on its own are different from those of a male that simply supplies sperm and leaves. As you can imagine, there are many different reproductive strategies, more than we can consider here and they all have distinct bio-economic implications, benefits, and constraints. For example, a contributing factor in social evolution is that when raising offspring is biologically expensive, cooperative child rearing can improve reproductive success significantly and increase the parents "return on investment" (ROI).

Consider the situation in placental mammals where fertilization occurs within the female and relatively few offspring are born from any one female. The female must commit resources to supporting the development and nurturing of the new organisms during the period from fertilization to birth and often beyond. Female mammals typically protect their young and feed them with milk, generated by specialized milk-secreting (mammary) glands. Depending on the species, the young are born at various stages of development, from the active and frisky, such as goats (→), to the helpless (humans). During the period when the female feeds and protects its offspring, the female is often more stressed and vulnerable than at other times. Cooperation with other females or with a specific male can influence the rate of survival of both mother and offspring, and impact the male's reproductive success. At the same time, protecting mother and offspring can increase the male's vulnerability. An important question is how does a male know

that the offspring he is acting to protect and nurture are his? Time protecting and gathering food for unrelated offspring is time and energy diverted from the male's search for a new mate and so reduce the male's overall reproductive success.

Looking at the natural world, we find a range of sexual behaviors, from males who sexually monopolize multiple females (polygyny) to polyandry, where females have multiple male "partners." In some situations, no pair bond forms between male and female, whereas in others male and female pairs are stable and (largely) exclusive. In some cases these pairs are long lasting, in others there is what has been called serial monogamy, pairs form, break up, and new pairs form. Sometimes females mate with multiple males, a behavior that is thought to confuse males (they cannot know which offspring are theirs) and so reduces infanticide by males.[201] .

Often, while caring for their young, females are sexually inactive. When males monopolize females, the arrival of a new male who displaces the previous male can lead to behaviors such as infanticide. By killing the young, fathered by another male, the female becomes reproductively active sooner, and so able to produce offspring related to the new male. There are situations, for example in some spiders, in which the male may risk, or even allow itself to be eaten during sexual intercourse as a type of "nuptial gift" that serves to block other males from mating with the female (who is, after all, busy eating and mating) and increases the number of the offspring that result from the mating event. This is an effective reproductive strategy for the male if its odds of mating with a female are low. Better (evolutionarily) to mate (reproduce) and die than never to have mated (reproduced) at all. An interesting variation on this behavior is described by Albo et al.[202] Male *Pisaura mirablis* spiders offer females "nuptial gifts", perhaps to avoid being eaten during intercourse. Of course where there is a strategy, there are counter strategies. Instead of an insect wrapped in silk, a male can offer a worthless gift, an inedible object (a silk-wrapped stone). Females cannot initially tell that the gift is worthless but quickly terminate mating when they do. Over time, as deceptive male strategies become more common, females may come to develop counter strategies. For example, a number of female organisms store sperm from a mating; they can eject that sperm and replace it with sperm derived from subsequent mating events.[203] Female wild fowl (*Gallus gallus*) can bias the success of a mating event in favor of dominant males. Following mating with a more dominant male, they can eject the sperm of subdominant males. The result is the production of more robust offspring.[204] Such behaviors are known as cryptic female choice; cryptic because the female's choice of sperm is not obvious. These are not (apparently) conscious decisions on the part of the female but physiological responses to various cues. And so it goes, each reproductive strategy leads, over time, to counter measures (another "Red Queen" situation). For example, in species in which a male guards a set of females (its harem), groups of males can work together to distract the guarding male, allowing members of their group to mate with the females. Many mating and reproductive strategies exist.[205] Molecular studies that can distinguish an offspring's parents suggest that "cheating" by both males and females can occur even among highly monogamous species. The extent of cheating will, of course, depend on the stakes. The more negative the effects on reproductive success, the more evolutionary processes will select against it.



In humans, a female can have at most one pregnancy a year, while a totally irresponsible male could, in theory at least, make a rather large number of females pregnant during a similar time period. Moreover, the biological costs of generating offspring is substantially greater for the female, compared to the male.[206] There is a low but real danger of the death of the mother during pregnancy, whereas males are not so vulnerable, at least in this context. So, if the female is going to have offspring, it would be in her evolutionary interest that those offspring are as robust as possible, meaning that they are likely to survive and reproduce. How can the female influence that outcome? One approach is to control fertility, that is, the probability that a "reproductive encounter" results in a pregnancy. This is accomplished physiologically, so that the odds of pregnancy increase when the female has enough resources to successfully carry the fetus to term. One might argue that the development of various forms of contraception are just another facet of this type of behavior, but one in which females (and males) consciously control reproductive outcomes.

References

[200] Gender-bending fish: http://evolution.berkeley.edu/evolibrary/article/fishtree_07

[201] Promiscuous females protect their offspring

[202] Worthless donations: male deception and female counter play in a nuptial gift-giving spider

[203] Evolution: Sperm Ejection Near and Far & Sperm Competition and the Evolution of Animal Mating Systems

[204] Female feral fowl eject sperm of subdominant males & Cryptic female choice favors sperm from major histocompatibility complex-dissimilar males

[205] The Evolution of Alternative Reproductive Strategies: Fitness Differential, Heritability, and Genetic Correlations

[206] ‘ Parental investment

This page titled 4.9: Sexual dimorphism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.10: Sexual Selection

It is not uncommon to see morphological and behavioral differences between the sexes. Sometimes the sexual dimorphism and behavioral differences are profound; they can even obscure the fact (at least for human observers) that the two sexes are members of the same species (→). Specific traits associated with one sex may even appear maladaptive; they might be expected to reduce rather than enhance an organism's reproductive potential.[207] The male peacock's tail, the gigantic antlers of male moose, or the bright body colors displayed by some male birds are obvious examples. Darwin recognized the seriousness of this problem for evolutionary theory and addressed it in The Descent of Man and Selection in Relation to Sex (1871). Where the investment of the two sexes in successful reproduction is not the same, the two sexes may have different and potentially antagonistic reproductive strategies. Organisms of different sexes may be "looking" for different traits in their mates. In general, the larger parental investment in the production and rearing of offspring, the less random is mating and the more prominent are the effects of sexual selection, that is, the choice of who to mate with.[208] It is difficult not to place these behaviors in the context of conscious choices, that is looking, wanting, etc., but they appear to be the result of selected behaviors and do not imply self-conscious decision making or moral judgements. Presumably, they arise from preferences based on costs and benefits. In humans, how consciousness, self-conscious-ness, social organization, ideological and theo-political choices influence sexual behavior (and selection) is even more complex and way beyond our scope here.

> *One of the most robust and reliable findings in the scientific literature on interpersonal attraction is the overwhelming role played by physical attractiveness in defining the ideal romantic partner. Both men and women express marked preference for an attractive partner in a non- committed short-term (casual, one night stand) relationship.*
>
> *For committed long-term relationships, females appear to be willing to relax their demand for a partner's attractiveness, especially for males with high social status or good financial prospects.*
>
> *Males also look for various personality qualities (kindness, understanding, good parental skills) in their search for long-term mating partners, but unlike females, they assign disproportionately greater importance to attractiveness compared to other personal qualities.*
>
> *The paramount importance of attractiveness in males' mate choices has been recently demonstrated by using the distinction between necessities (i.e., essential needs, such as food and shelter) and luxuries (i.e., objects that are sought after essential needs have been satisfied, such as a yacht or expensive car) made by economists.*
>
> *Using this method, Li et al., reported that males treat female attractiveness as a necessity in romantic relationships; given a limited "mating budget," males allocate the largest proportion of their budget to physical attractiveness rather than to other attributes such as an exciting personality, liveliness, and sense of humor.*

Consider the situation in which females have help in raising their offspring. Selection would be expected to favor females who mate preferentially with the most robust, but not necessarily the most supportive males based on the reasonable assumption that the most robust appearing male will lead to the most robust offspring. This prediction has been confirmed experimentally in a number of systems; the robustness of offspring correlates with the robustness of the male, a win for evolutionary logic.[209] In this context, the reproductive success of a male will be enhanced if it "advertises" its genetic robustness through visible and unambiguous features. A useful sign of robustness needs to be difficult to fake.[210] The size and symmetry of a beetle's or an elk's antlers accurately communicate their state of health.[211] Consider a male peacock's tail; a large, colorful and symmetrical tail is a sign of health an the ability to survive.

Another approach involves territoriality. Individuals, typically males, establish and defend territories. Since there are a limited number of such territories and females only mate with males that have established and can defend a territory, only the most robust males are reproductively successful. An alternative scenario involves males monopolizing females sexually. Because access to females is central to their reproductive success, males may interact with one another to establish a dominance hierarchy, typically in the form of one or more “alpha” males. Again, the most robust males are likely to emerge as alpha males, which in turn serves the reproductive interests of the females. This type of dominance behavior is difficult to fake. But, cooperation between non-alpha males can be used to thwart the alpha male’s monopolization of females.

Now consider how strategies change if the odds of successful reproduction are improved when a male can be counted on to help the female raise their joint offspring. In this situation, there is a reproductive advantage for females who can identify those males who will display reproductive loyalty.[212] Under conditions of cooperative rearing of offspring, females may compete with other females for access to (perhaps rare) loyal males. Moreover, it can be in the male’s interest to cooperate with fertile females; often females (but not human females) advertise their state of fertility, through external signals.

There are of course, alternative strategies. For example, groups of females, including sisters, mothers, daughters, aunts, and grandmothers may cooperate, thereby reducing the importance of male assistance. At the same time, there can be conflicts. What happens if the most robust male is not the most "loyal" male? A female could maximize their reproductive success by mating with a robust male and bonding with a loyal male, who helps rear another male’s offspring. While not in the "loyal"male’s immediate reproductive interest, it could enhance their future reproductive "success". Selection could favor males that cooperate with one another to ward off robust but transient males. Since loyal males already bond and cooperate with females, it may be simple for them to cooperate with other males. Another possibility is that a loyal male may be attractive to multiple females, who in turn compete for his attention and loyalty. The outcome of such interactions is likely to be influenced by the number of females a male can protect and the impact the male has on a female's reproductive success.

It is critical to correctly read and/or respond to various traits, an ability likely to be selected. Males that can accurately "read" other males can determine whether they are likely to win a fight, while an inaccurate assessment could result in crippling injuries. A trickier question is how to determine whether a potential mate will be loyal? As with advertisements of overall robustness, we might expect that traits that are difficult or expensive to generate will play a key role. So how does one unambiguously signal one’s propensity to loyalty and a willingness to cooperate? One could use the size and value of nuptial gifts. The more valuable, that is, the more expensive and difficult the gift is to attain, the more loyal the recipient can expect the gift giver to be. On the other hand, once valuable gift-giving is established, one can expect the evolution of traits in which the cost of the gift given is reduced and by which the receiver tests the value of the gift, a behavior we might term rational skepticism, as opposed to naive gullibility.

This points out a general pattern. When it comes to sexual and social interactions, organisms have evolved to “know” the rules. If costs of cooperation are "expensive" there will be selective pressure to cheat. Cheating can be suppressed by making the signs of social or sexual cooperativity difficult or impossible to fake, or by generating counter-strategies that identify fakes. These biological realities produce many behaviors, some of which are disconcerting. These include sexual cannibalism, male infanticide, and various forms of infidelity, mentioned above. What we have not considered as yet is the conflict between parents and offspring. Where the female makes a major and potentially debilitating investment in its offspring, there can be situations where continuing a pregnancy can threaten the survival of the mother. In such cases, spontaneous abortion (ending the pregnancy) could save the female, who can go on and mate again. In a number of organisms, spontaneous abortion occurs in response to signs of reproductive distress in the fetus. Of course, spontaneous abortion is not in the interest of the offspring and we can expect that mechanisms will exist to maintain pregnancy, even if it risks the life of the mother, in part because the fetus and the mother, while related are not identical and can have conflicts of interest.[213] There are many variations of reproductive behavior to be found in the world and a full discussion is beyond our scope here.

Questions to answer

46. How it is possible that a parent’s, particularly a mother's interests can conflict with the interests of its offspring?
47. Why do the different sexes often display different traits?
48. What might the absence of sexual dimorphism indicate about their reproductive behaviors?

References

[207] “Flaunting It' - Sexual Selection and the Art of Courtship: http://youtu.be/g3B8hS80k6A

[208] R. Trivers, Parent investment and Sexual selection : http://joelvelasco.net/teaching/3330/trivers72-parentalinvestment.pdf

[209] Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character

[210] In Male Rhinoceros Beetle, Horn Size Signals Healthy Mate

[211] Attractiveness of grasshopper songs correlates with their robustness against noise

[212] From an evolutionary standpoint what is the meaning of romantic love?

[213] Maternal-Fetal Conflict: and Wildman et al (2011). Spontaneous abortion and preterm labor and delivery in nonhuman primates: evidence from a captive colony of chimpanzees (Pan troglodytes). *PloS one*, *6*(9), e24509.

This page titled 4.10: Sexual Selection is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

4.11: Curbing runaway selection

Sexual selection can lead to what has been termed, but is not really, runaway selection. For example, the more prominent the peacock male's tail the more likely he will find a mate even though larger and larger tails can have significant negative effects of survival. All of which is to say that there are both positive and negative selection pressures for tail size, influencing the probability that a particular male mates successfully. Selection never really runs away, but settles down when the positive benefit(s), in terms of sexual success, and the negative cost(s) of a trait come to be (roughly) equal to each other. In the wild, there will be reproductively successful male peacocks with large tails, even if those tails may make them more susceptible to predators.

For another example, consider the evolution of extremely large antlers associated with male dominance and mate accessibility, such as occurred in *Megaloceros giganteous* (↓). Large antlers influence the animal's ability to move through heavily wooded areas or to run fast. In a stable environment, the costs and benefits of antlers are expected to balance out. Selection will produce an optimal solution. If the environment changes however, this balance may be disrupted; pre-existing behaviors and phenotypes may influence the impact of selective pressures. Organism's may not be able to adapt fast enough to avoid extinction. There is a balance between costs and benefits, particularly within a changing environment.



Social and ecological interactions apply to all organisms, from bacteria to humans. They serve as a counter-balance to the common caricature of evolution as a ruthless and never ceasing competition between organisms. This hyper-competitive view, often known as the struggle for existence or Social Darwinism, may be appealing to ruthless (anti-union / anti-social) capitalists but is not supported by scientifically-established evolutionary scenarios. It has been promulgated by a number of pundits who have used it to justify various political and inherently non-scientific positions, arguing against social programs that help the poor, often characterized as the unfit ,at the "expense" of the wealthy (who might be viewed as parasites). Assuming that certain organisms are inherently less fit, and that they could be identified, this view of the world gave rise to eugenics, the view that genetically inferior people should be removed from the population or sterilized, before their "bad" traits overwhelmed a particular culture. Eugenics was an influential ideology in the United States during the early part of the 20th century and inspired the genocidal programs of the Nazis in Germany. It re-emerges periodically for various reasons. What is particularly odd about this pseudo-evolutionary perspective is that it is actually anti-evolutionary, since if the unfit really were unfit, they could not possibly take over a population. In addition, it completely ignores the deeply social (cooperative) aspect of the human species.

Questions to answer

49. What does it mean to cheat, in terms of sexual selection - is a "cheating" organism consciously deceptive?
50. Are there specific types of "cheating" behaviors that females use with males? or males with females?
51. What are the costs involved when a male tries to monopolize multiple females? What are the advantages?
52. What limits runaway selection, or better, why is runaway selection impossible?

Questions to ponder

- Should human ethical or ideological beliefs and decisions be more important than evolutionary cost-benefit calculations?

Short chapter summary

- Cooperation evolves when benefits (kin, reciprocity, group structure) outweigh costs.
- Cheater control is essential for stable cooperation—from microbes to animals.
- Sex mixes alleles, enabling faster adaptation; sexual selection can be powerful (and risky).
- Runaway traits face ecological and physiological brakes.

This page titled 4.11: Curbing runaway selection is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

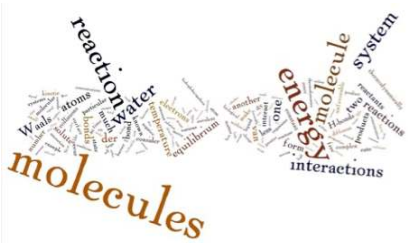
CHAPTER OVERVIEW

5: Molecular Interactions, Thermodynamics and Reaction Coupling

5.0: Introduction
5.1: Approaching Thermodynamics
5.2: Coupling thermodynamically favorable and unfavorable reactions
5.3: Reaction rates
5.4: Coupling Reactions
5.5: Inter- and intra-molecular interactions
5.6: Covalent bonds
5.7: Bond Stability and Thermal Motion (a non-biological moment)
5.8: Bond polarity, inter- and intramolecular interactions
5.9: The implications of bond polarity
5.10: Interacting with Water

This page titled 5: Molecular Interactions, Thermodynamics and Reaction Coupling is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.0: Introduction

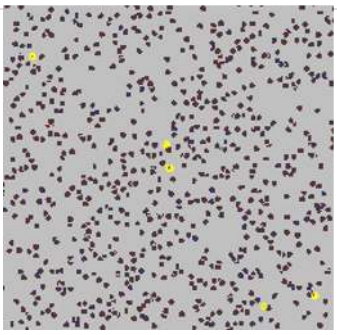


In which we change gears, from evolutionary biology to the physical and chemical properties of organisms. These properties shape and constrain biological behaviors. We consider how molecules interact and react with one another and how these interactions and reactions determine the properties of substances and systems, particularly the continuous, bounded, non-equilibrium system that is life.

All biological systems and processes, from those involved in cellular integrity, growth, replication, movement, and differentiation to thoughts and feelings, obey the rules of chemistry and physics. What makes biological systems unique is that, unlike simpler systems that move toward thermodynamic equilibrium, living organisms maintain a non-equilibrium state. Chemical reaction systems can be assembled *de novo*. So far, this is not true for organisms and the cells they are composed of; every biological system has an uninterrupted history going back billions of years. So, we start by clarifying what it means when we say that a system is at equilibrium versus when it is in a obligate non-equilibrium (living) state. A biological system at equilibrium is dead, and not ever, ever coming back to life.

This page titled 5.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.1: Approaching Thermodynamics



Brownian motion (Wikipedia)

We begin by distinguishing between the macroscopic world that we perceive directly and the molecular world that we can come to understand through scientific observations and experimentation. Perhaps the best example of the connection between the two was Albert Einstein's (1879–1955) explanation of Brownian motion, named after the botanist Robert Brown (1773-1858). In 1827 Brown described the jiggly movement of pollen grains in water viewed using a microscope. In 1905 Einstein recognized that the "random" movements of the pollen grain (→) was evidence for the existence of atoms (and molecules) and the transfer of energy associated with collisions between water molecules and pollen grains.[214] This type of movement continues even though the system is at thermodynamic equilibrium, a state in which nothing macroscopic appears to be going on. This insight revealed the connection between temperature, energy, and molecular movements, the focus of the laws of thermodynamics.[215]



To apply the laws of thermodynamics to life, we first need to define the system(s) we are considering. Our system of interest could range from the Universe as a whole to a single living cell or multicellular organism. Next, can energy and matter enter or leave the system? In a "closed" system, neither energy or matter can enter or leave. Matter and/or energy can enter an "open" system, typically in a controlled manner. The laws of thermodynamics enable us to predict the behavior closed systems; a closed system will, over time, come to equilibrium, a state in which there are no more macroscopic changes. In an open system, the system's behavior will depend upon the degree to which energy and matter are exchanged with the external world, the world outside the system's boundary. Organisms and the cells that composed them are obligate open systems. They require the exchange of matter and energy with the external world to maintain their non-equilibrium living state.

Cryptobiosis: We have characterized organisms as dynamic, open, non-equilibrium systems. An apparent exception to the dynamic aspect of life are organisms that display a rather special phenotypic adaptation, known as cryptobiosis. Some organisms, such as the tardigrad or water bear (→), can be freeze-dried and persist in a state of suspended animation for decades.[216] What is critical, however, is that when the organism is in this cryptobiotic state it is not at equilibrium, in much the same way that a battery is not at equilibrium, but can be reanimated when returned to normal conditions, specifically by the addition of water.[217] Cryptobiosis is a genetically-based adaptation that takes energy to produce and energy is needed to emerge from stasis. While the behavior of tardigrads is extreme, many organisms display a range of adaptive behaviors that enable them to survive hostile environmental conditions.

References

[214] Albert Einstein: The Size and Existence of Atoms http://youtu.be/nrUBPO6zZ40

[215] Laws of thermodynamics (Wikipedia) https://en.wikipedia.org/wiki/Laws_of_thermodynamics

[216] Everything you need (and want) to know about tardigrades

[217] On dormancy strategies in tardigrades & Towards decrypting cryptobiosis

This page titled 5.1: Approaching Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.2: Coupling thermodynamically favorable and unfavorable reactions

Biological systems are complex. Both their overall structural elements and many of their molecular components, including DNA and proteins, are the products of thermodynamically unfavorable reactions and are intrinsically unstable. What does that imply? In a closed, isolated system we expect that it would be impossible for such molecules to form. So the question arises, what drives the synthesis of these molecules in organisms? how do these biosynthetic reactions take place? The answer involves the coupling of thermodynamically favorable reactions to thermodynamically unfavorable reactions.

We can distinguish a favorable from an unfavorable reaction by the change in free energy, known as Gibbs free energy, associated with the reaction. A negative ΔG indicates a favorable reaction that releases energy, while a positive ΔG indicates an unfavorable reaction that requires the addition of energy. When ΔG for the system = 0 no observable, that is macroscopic changes will occur. The system is at equilibrium. You may recognize the Gibbs free energy equation: $\Delta G = \Delta H - T\Delta S$, where T is the system temperature.[218] Under defined conditions (e.g. constant pressure) ΔH is the change in enthalpy, the amount of energy transferred between the system and the surroundings during a reaction, and ΔS as the change in a system factor known as entropy. Entropy is related to the ways that energy and matter can be arranged; the more possible equivalent ways, the greater the entropy. A simple way to think of entropy is the more organized a system, the lower the entropy (See CLUE: Chemistry for a more detailed discussion). Entropy is often used colloquially to describe "random" or disordered systems. A gas has greater entropy than the solid form of the same substance because there are more ways for the gas particles can be arranged, compared to a solid where the particles are fixed in place.

For any change in an isolated system, the system's entropy always increases - which is the Second Law of Thermodynamics. At this point you might be saying wait a minute, aren't there systems in which entropy decreases? For example, it is certainly possible to change a gas (higher entropy) into a liquid or a solid (lower entropy), but the critical part is that this system is not isolated. While the system may decrease in entropy, the entropy of the universe as a whole increases. When gas condenses to a liquid, energy is removed and that energy is transferred to the surroundings, a process that increases the entropy of the surroundings by making molecules move and vibrate faster. While the entropy of a particular region of the universe (the system, the cell, the organism, etc.) may decrease, the total entropy of the universe always increases.

Questions to answer:

53. How is it that a dried out tardigrad can be alive?.
54. What are the common components of a non-equilibrium system; how might you identify such a system.

References

[218] The value of ΔG depends upon the concentrations of solute and solvent, but we will ignore that complexity for the moment. https://en.wikipedia.org/wiki/Bioenergetics

This page titled 5.2: Coupling thermodynamically favorable and unfavorable reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.3: Reaction rates

A reaction is characterized by its equilibrium constant, K_{eq}, a function of the reaction itself, the concentrations of the reactants, and system temperature and pressure. In biological systems we generally ignore pressure (and only occasionally consider temperature), although both can be important for organisms that live on the sea floor, mountain tops, or hydrothermal vents. The equilibrium constant (K_{eq}) for the reaction A + B ⇆ C + D is defined (→) as the concentrations of the products (C + D) divided by the product of the concentrations of the reactants at equilibrium, where nothing macroscopic is happening. At equilibrium the concentrations do not change, that is why K is a constant. For a thermodynamically favorable reaction, one that favors the products, K will be greater, often much greater, than one. The larger K_{eq}, the more product and the less reactant there will be when the system reaches equilibrium. If the equilibrium constant is less than 1 then, at equilibrium the concentration of reactants will be greater than the concentration of products.

$$K = \frac{[C][D]}{[A][B]}$$

While the concentration of reactants and products of a reaction at equilibrium remains constant it is not the case that the system is static. If we were to peer into (or imagine) the system at the molecular level we would find that reactants are continuing to form products and products are rearranging to form reactants; the rate of these two reactions will be equal, although slow.[219] Most reactions involve collisions between molecules. The frequency of productive collisions between reactants or products increases as their concentrations increase. At equilibrium for a highly favorable reaction; the high concentration of products (produced by the reaction) x low probability of effective collisions will equal the low concentration of reactants (remaining) x higher probability of effective collisions.

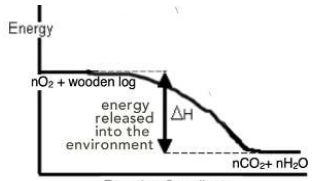
Knowing that a reaction is thermodynamically favorable does not tell us much (or really anything) about whether the reaction occurs to a significant extent under a particular set of conditions. For example, consider a wooden log, which is composed primarily of the carbohydrate polymer cellulose $(CH_2O)_n$. In the presence of molecular oxygen (O_2) the reaction:

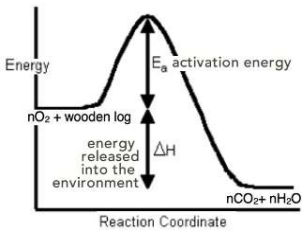
$$nO_2 + \log\left((CH_2O)_n\right) \leftrightarrows nCO_2 + nH_2O + \text{ heat}$$

is extremely favorable thermodynamically, that is. It has a large negative ΔG and a large equilibrium constant; once the reaction starts it goes completely to CO_2 and H_2O. Yet logs are stable - they do not spontaneously burst into flames. The question is why not?

The answer lies in the fact that both the equilibrium constant and ΔG (or for the more chemically rigorous, ΔG°) tell us only whether a reaction is thermodynamically favorable; they tell us nothing about whether, or how fast that reaction will proceed under a specific set of conditions. For that we have to turn to the study of reaction rates. This requires us to consider the various factors that affect the reaction. In general a reaction will go faster if there are more reactant molecules. Reactant molecules must collide to initiate a reaction. For example, in the case of the log and oxygen, oxygen molecules (O_2) must come in contact with the log. In air (at sea level) O_2 molecules amount to ~20% of the total molecules present in the atmosphere. If we increase the atmospheric O_2 concentration, the log will burn faster and brighter, because there are more collisions that initiate the reaction.[220] Under normal conditions, however, the log will not start burning spontaneously - added energy is needed. Why? Because the transition between reactants and products requires the breaking of bonds. Breaking a bond requires the addition of energy. Generally, the energy needed is more than the energy available through molecular collisions. The energy required for bonds to break and the reaction to proceed is known as the activation energy. The presence of activation energy explains why chemical systems do not always move spontaneously to equilibrium, why thermodynamically unstable molecules can be "stable". Why nucleic acids and proteins do not quickly breakdown to produce more stable (but rather more boring) molecules such as CO_2, H_2O, and NH_3 from which they are composed.

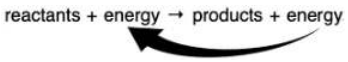
To explore the idea of activation energy further, let us consider the very simplified model of a log burning in air to produce CO_2 and H_2O, a reaction that is, in fact, complex. We could represent this process on a graph of energy (or more accurately Gibbs Free Energy (G)) vs reaction progress like this (↓). As the reaction proceeds, a great deal of heat is released into the surroundings. This released energy corresponds to the ΔH between reactants and products. The graph also indicates that the products are more stable (lower energy) than the reactants. But, the reaction energy graph does not give us any indication that energy must be added to start the reaction, or why. If we add in this energy the graph would look like this (↓). The activation energy (E_a) is the energy needed to





break the bonds within wood molecules and in O_2. These steps, in which pre-existing bonds are broken but new bonds have yet to form is also known as the transition state. In general the amount of activation energy needed to reach the transition state determines the rate of the reaction. If most collisions with surrounding molecules supply this (or more) energy, the reaction will occur rapidly and spontaneously. If, on the other hand and in the case of a log at room temperature, few if any collisions supply enough energy to break the bonds necessary to reach the transition state, the reaction rate will be slow or the reaction will not occur at all (more below). For the wood burning reactions, the energy needed to start the reaction may involve a downed electrical line, a lightning strike, or a burning match.

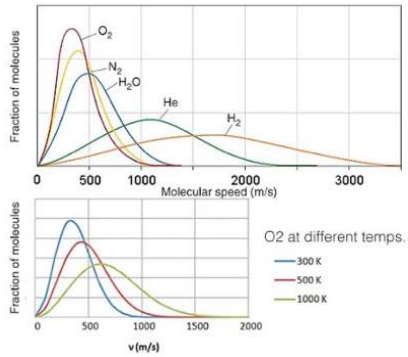
Once the "forward" reaction starts energy will be released as new bonds are formed. The resulting increase in the temperature



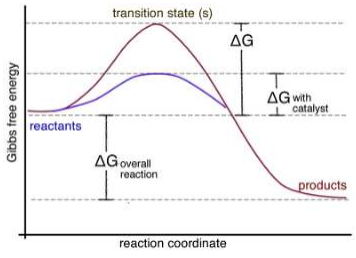
(average kinetic energy of molecules) leads to increased kinetic energy of colliding reactants; more collisions reach the transition state and produce products. The reaction rate increases and becomes "self-sustaining" - a positive feedback loop (→). As reactants are used up productive collisions between reactants will decrease, while productive collisions between products (the "back reaction") with sufficient activation energy (more energy in fact than the forward reaction) will increase. Unless further fuel (reactants) is added, the reaction system will reach equilibrium. A equilibrium the rates of the forward and back reactions will be equal.

Activation energy and catalysis in biological systems.

As noted above, the reason why (most) thermodynamically favorable reactions do not occur spontaneously when reactants come into contact is that bonds must be broken for the reaction to occur. Breaking bonds, particularly covalent bonds, requires a large amount of energy. In biological systems there are two major sources for this bond-breaking energy: light and collisions with other molecules. A molecule can absorb a photon (a particle of light) or energy can be transferred through collisions with other molecules. At room temperature liquid water molecules are moving on average at ~640 meters/second. That is not to say that all molecules are moving at that same speed. If we were to look at the population of molecules, we would find a distribution of speeds known as a Boltzmann (or Maxwell-Boltzmann) distribution (←). As they collide with one another, the molecules exchange energy; one molecule can emerge from a collision with more energy than it entered with. Since reactions occur at temperatures well above absolute zero, there is plenty of energy available in the form of the kinetic energy (movement) of molecules.



Biological systems are constrained in a number of ways. As we will see, the three-dimensional structure of many macromolecules, specifically proteins and nucleic acids, is critical to their normal function, and their 3D structure is often unstable - even small changes, by the standards of a typical chemistry lab, in temperature can lead to what is known as denaturation and the loss of function. The take home message is that biological systems use alternative strategies to control the rates of the reactions they depend upon.



Their solution are molecules, typically proteins or small RNAs known as enzymes or ribozymes respectively, that act as catalysts. But what exactly does a catalyst do? Basically, it lowers the "activation" energy of a reaction by interacting with the reactants (→). The result is that, at a particular temperature and environmental conditions, the reaction rate will be increased in the presence of an active catalyst. An important feature of biological catalysts is that their activity can be regulated; their effectiveness as a catalyst for specific reaction can be turned on or off. As we will see, the regulate-ability of biological catalysts is central to maintaining the dynamic, non-equilibrium state of the cell.

Questions to answer:

55. How might changes in an organism's internal temperature influence reaction rates?
56. A reaction is at equilibrium; we increase the amount of reactant or product. What happens (over time) to the amounts of reactants and products? Drawing some graphs may be helpful.

57. What does reducing the activation energy of a reaction do to a system at equilibrium? What does it do to a system far from equilibrium? Drawing some graphs may be helpful.
58. How and why does the feedback system of a burning log change over time?

Question to ponder:

- Propose a model for how (at the molecular level) a catalyst might lower a reaction's activation energy?

References

[219] This, of course, assumes that we have a closed system, that is, that neither the products or the reactants can leave the system, and that the volume of the system also remains constant. If the reactants can “leave the scene” of the reaction, then of course the back reaction, Products ⇆Reactants, will be much less likely to occur.

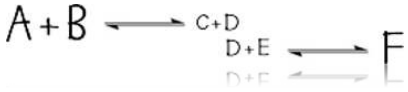
[220] This is one reason why smoking is not a good idea for people who have to use supplementary oxygen to breathe

This page titled 5.3: Reaction rates is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.4: Coupling Reactions

There are large numbers of different types of reactions that occur continually within cells. As a crude rule of thumb, a reaction that produces smaller molecules from larger ones will be thermodynamically favored, while reactions that produce larger molecules from smaller ones will be unfavorable. Similarly a reaction that leads to a molecule moving from a region of higher concentration to a region of lower concentration will be thermodynamically favorable. So how exactly can we build big molecules, such as DNA and proteins, and generate the concentration gradients upon which life depends?

Thermodynamically favored reactions are characterized by negative ΔG^o, with an equilibrium constant greater, typically much greater, than 1.They are associated with the breakdown of food molecules and the capture of energy released, known generically as catabolism. Unfavorable reactions, known as anabolism, are driven by the energy made available by catabolism. An organism's metabolism is the sum total of these various reactions. The question is, if a reaction is unfavorable - how, exactly, does it occur? The answer lies in coupling unfavorable synthesis reactions to favorable breakdown reactions. By coupling we mean that the two reactions share a common intermediate. In this example (↓) there are two reactions that share the component "D". Let us assume



that the upper reaction is unfavorable while the lower reaction is favorable. Assume that both reactions are occurring at measurable rates and that E is already present in the system. What happens? At the start of our analysis, the concentrations of A and B are high, and C, D, E and F are low. We can then use Le Chatelier's principle to predictions. This principle states that if a change is made to a system at equilibrium, then the system will shift to counteract that change, basically because the number of productive collision events associated with one direction of the reaction will increase compared to those associated with the other direction.[221]

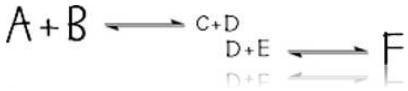
Consider how how Le Chatelier's principle works in this reaction system.

$$A + B \leftrightarrows C + D$$

At equilibrium the rates of the forward and reverse reactions are equal. What happens if other cellular reactions consume C and reduce it concentration? The rate of the reverse reaction will decrease; there is less C to collide with D to initiate the reaction: the reaction is no longer at equilibrium. More A and B will react to give C and D. Similarly if the concentration of B were to increase (to the active state of the cell), the rate of the forward reaction will increase; the reaction will produce more products. The increase in C + D will continue until their concentration reaches a point where:

$$\textbf{the rate of } A + B \leftarrow C + D \textbf{ reaction is equal to } A + B \rightarrow C + D \textbf{ reaction.}$$

This type of behavior arises from the fact that at equilibrium reaction systems are dynamic at the molecular level – things are still occurring but at equal rates, so there is no net change. When you add or take something away from the system, it becomes unbalanced. The system response moving to restore the equilibrium condition – but it can't really, since C is constantly being consumed by other biological reactions.



So back to our system of coupled reactions in a biological (non-equilibrium) system (→). The unfavorable A+B reaction will generate a small amount of C+D. Assuming that E is present (the produce of other reactions) the favorable D+E reaction means that as D is formed it will react with E to produce F, removing D from the system. As the concentration of D decreases, the C+D "back reaction" becomes less probable while the A+B reaction continues producing more C and D. The end result is that, even though energetically unfavorable, C and D will be produced and D will be used up to make F, something that the chemically-naive might think impossible. By linking reactions in this way biological systems use energy and matter from the outside world to produce and regulate the concentrations of molecules needed for maintenance, growth, and reproduction of the living system.

Questions to answer:

59. How is LeChatelier's principle involved in reaction coupling? How does removing "waste" components influence reaction coupling?
60. How would you go about deciding whether a biosynthetic system involves coupled reactions?
61. Why are catalysts important to reaction coupling in biological systems. Why are they essential for life (at least as we know it)?

References

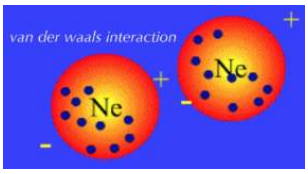
[221] http://en.wikipedia.org/wiki/Le_Chatelier's_principle

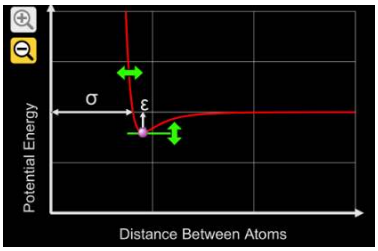
This page titled 5.4: Coupling Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.5: Inter- and intra-molecular interactions

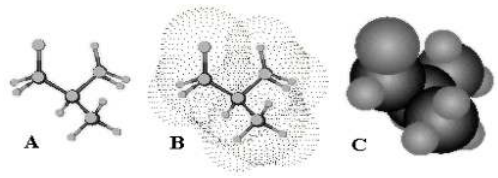
We have briefly considered energy and how it can be transferred within reaction systems. Now consider what we mean by matter, which implies an understanding of the atomic organization of molecules. As you may know by now, all matter is composed of atoms. The internal structure of atoms is the subject of quantum physics and we will not go into it in any depth. Suffice it to say that each atom consists of a tiny positively charged nucleus, composed of protons and neutrons, and a cloud of negatively charged electrons. Typically molecules, which are collections of atoms connected by covalent bonds (see below), interact with each other to various extents. Molecules interact with one another through two types of interactions. The first of these are London Dispersion Forces or LDFs named after their discoverer Fritz Wolfgang London (1900–1954).

Attractive LDFs arise from the fact that the relatively light (in terms of mass) negatively-charged electrons of an atom or molecule are in continual movement, compared to the relatively massive positively-charged nuclei (←). As two atoms/molecules approach one another the distorted electron cloud of one will induce a distortion (an induced dipole) in the electron cloud of the other. Note that because charges on the protons and electrons are equal in magnitude the atoms/molecules as a whole remain electrically neutral.



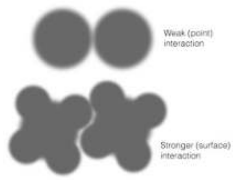


PhET LDFs vary as ~$1/R^6$ where R is the distance between the molecules (→). As a result LDFs act over very short distances, typically less than a nanometer (1 nm = 10^{-9} m). The ˚magnitude of this attractive force reaches its maximum when the two molecules are separated by what is known as the sum of their van der Waals radii (the van der Waals radius of a carbon atom is ~0.17 nm). If they move closer than this distance, the attractive LDF is quickly overwhelmed by the rapidly increasing, and strongly repulsive forces that arise from the electrostatic interactions between positively charged nuclei and negatively charged electrons. When atoms form a covalent bond (see below), their van der Waals surfaces merge to produce a new molecular van der Waals surface. This surface is the closest distance that two molecules can approach one another before repulsion kicks in and drives them away from one another. Every molecule generates LDFs when it approaches another molecule, so LDF-mediated interactions are universal.



There are a number of ways to draw molecules, but the space-filling or van der Waals surface view is the most useful for our purposes. While realistic it can be confusing, since it obscures how the atoms in the molecule are linked together. Consider this set of representations of the simple molecule 2-methylpropane (→). As molecules become larger, as is the case with many biologically important molecules, it can be difficult to appreciate their underlying organization based on a van der Waals surface representation.[222]

The strength of the LDF-mediated interactions between molecules is influenced by their shapes. The greater the surface complementarity between two molecules, the stronger their interaction. Compare the interaction between two spherical monoatomic "Noble" atoms, such as helium, neon, argon, or xenon (top), and two molecules with more complex shapes (bottom)(→). The two monoatomic particles interact via LDFs at a single point, so the strength of the interaction is minimal. On the other hand, the more complex molecules interact over extended surfaces, so the LDFs between them are greater, resulting in a stronger van der Waals interaction. In the face of thermal (molecular) collisions, you can predict which pair of interacting molecules (or atoms) is more likely to remain intact.



References

[222] Explicit Concepts of Molecular Topology: http://www.chem.msu.ru/eng/misc/babaev/match/top/top02.htm

This page titled 5.5: Inter- and intra-molecular interactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.6: Covalent bonds

In the case of van der Waals interactions, the atoms and molecules involved retain a hold on their electrons, they remain distinct and discrete. There are cases, however, where atoms come to "share" electrons; sharing involves pairs of electrons, one or more from each atom. When electron pairs are shared, the atoms stop being distinct; their shared electrons are no longer restricted to one or the other. Since one electron cannot, even in theory, be distinguished from any other electron, they become a part of the molecule's electron system and form a new stable entity. It is not surprising (perhaps) that the properties of a molecule are distinct from, although influenced by, the properties of the atoms from which they are composed. Some atoms, common to biological systems, such as hydrogen (H), can form only a single covalent bond. Others (oxygen (O) and sulfur (S)) can make two, three (nitrogen (N)), four (carbon (C)), or five (phosphorus (P)) covalent bonds.

The sharing of electrons produces what is known as a covalent bond. Covalent bonds are 20 to 50 times stronger than the interactions based on LDFs. What exactly does that mean? Basically, it takes much more energy to break a covalent bond than to break an LDF-mediated interaction. Different covalent bonds between different atoms in different molecular contexts differ in terms of bond stability. A molecule is stable if its bond energies are high enough to remain intact when energy is delivered to the molecule through collisions with neighboring molecules or the absorption of light.

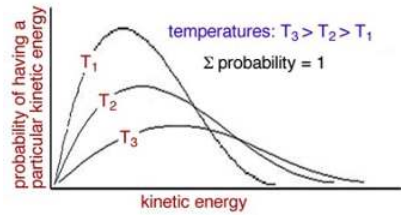
In addition to smaller molecules, biological systems contain distinct types of extremely large molecules, composed of many thousands to billions of atoms; these are known as macromolecules. Such macromolecules are not rigid; they can often fold back on themselves leading to **intramolecular** interactions (interactions within a given molecule). There are also interactions between molecules - which are referred to as **intermolecular** interactions. The strength and specificity of these interactions can vary dramatically and even small changes in a molecule's structure, such as result from mutations and allelic variations, can lead to dramatic effects on molecular shape and interactions with other molecules and function(s). Similarly, increasing temperatures can break weak interactions, leading to changes in molecular shapes and functions.

Molecules and molecular interactions are dynamic. Collisions with surrounding molecules can lead to parts of a molecule rotating with respect to one another around a single bond. The presence of a double bond restricts rotation; it involves what amounts to breaking and then reforming one of the bonds. If you have mastered some chemistry you know that it can be an over-simplification to consider bonds as distinct entities isolated from one another and their surroundings. In some structures the electrons in bonds are best considered as delocalized (that is not "stuck" between two adjacent atoms). These are often shown as "resonance structures" that behave as mixtures of single and double bonds. Again this restricts free rotation around the bond axis and acts to constrain molecular geometries. As we will come to see, the peptide bond that occurs between a carbon (C) and a nitrogen (N) atom in a polypeptide chain, displays such resonance behavior. Similarly, the ring structures found in the various "bases" present in nucleic acids result in flat structures that pack one on top of another. These various geometric complexities combine to make predicting a molecule's three dimensional structure increasingly challenging as its size increases, although computational tools, such as Alphafold, have led to dramatic improvements in structure prediction.

This page titled 5.6: Covalent bonds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.7: Bond Stability and Thermal Motion (a non-biological moment)

Molecules do not exist out of context. In the real, or at least the biological world, they do not sit alone in a vacuum, they are surrounded by other, mostly water, molecules. When we think about a system, we inevitably think about its temperature. Temperature is a concept that makes sense only at the system level. Individual molecules do not have a temperature, they have kinetic energy. The temperature of a system is a measure of the average kinetic energy of the molecules within it. The average kinetic energy is: E_k = 1/2 (average mass) x (average velocity)^2. As you may already know there is a lowest possible temperature, known as absolute zero (0ºK, −273.15ºC, or −459.67°F). At this biologically irrelevant temperature, molecular movements are minimal but not, apparently, absent all together.[223] It does not matter whether the system is composed of only a single type of molecule or many different types of molecules, at a particular temperature the average kinetic energy of all of the different molecules has one value. This is not to say that all molecules have the same kinetic energy, they certainly do not; each forms part of a distribution that is characterized by its average energy and shape, this is the Maxwell-Boltzmann distribution (introduced previously →). The higher the temperature, the more molecules will have a higher kinetic energy.



In a gas we can largely overlook the attractive inter-molecular interactions because the average kinetic energies of the molecules is sufficient to disrupt them – that is, after all why they exist as a gas. As we cool the system's the average kinetic energy of the molecules decreases. When the average kinetic energy gets low enough, the molecules will form a liquid. In a liquid, the movement of molecules is not sufficient to disrupt all of the interactions between them. This is a bit of a simplification, however. Better to think of it more realistically. Consider a closed box partially filled with a substance in a liquid state. What is going on? Assuming there are no changes in temperature over time, the system will be at equilibrium. What we will find, if we think about it, is that there is a partial phase change going on, that is:

$$\text{Molecule (gas)} \leftrightarrows \text{Molecule (liquid)}$$

At a particular temperature, the liquid phase is favored, although there will be some molecules in the system's gaseous phase. The point is that at equilibrium, the number of molecules moving from liquid to gas will be equal to the number of molecules moving from the gas to the liquid phase. If we increase or decrease the temperature of the system (that is add or remove energy), we will alter this equilibrium state; the relative numbers of molecules in the gaseous versus the liquid states will change. The equilibrium is dynamic, in that different molecules may be in the gaseous or the liquid states, even though the distribution of molecules between these two states will be steady, assuming a large enough system, very large numbers of molecules - a form of stochastic behavior.

In a liquid molecules are transiently associate with one another. That is why liquids can be poured, and why they assume the shape of the (solid) containers into which they are poured. This is in contrast to the container, whose shape is independent of what it contains. In a solid the molecules are much more tightly associated and so do not translocate with respect to one another, although they may jiggle in various ways. Solids do not flow. The cell, or more specifically, the cytoplasm, acts primarily as a liquid. Most biological processes take place in the liquid phase: this has a number of implications. First molecules, even very large macromolecules, move with respect to one another. Driven by thermal motion, molecules will move stochastically, in what is known as Brownian motion or a "random" walk.

Thermal motion influences whether, how, and for how long molecules associate with one another. Let us think about this process in the context of an ensemble of molecules, call them A and B that interact to form a complex, A:B. Assume that this complex is held together by LDF-interactions. In an aqueous solution, the A:B complex is constantly colliding with water molecules. These water molecules have various energies (from low to high), as described by the Boltzmann distribution. There is a probability that at any moment one or more collisions will deliver energy greater than the interaction energy that holds A and B together leading to the disassociation of the A:B complex into separate A and B molecules. Assume we start with a population of 100% A:B complexes, the time it takes for 50% of these molecules to dissociate into A and B is considered the "half-life" of the complex. We use the term half-life repeatedly to characterize the stability of a complex or a macromolecule. Now here is the tricky part, much like the situation with radioactive decay. Assuming a large enough population of A:B, we can confidently conclude that 50% of the A:B complexes will have disassembled into A and B at the half-life time, but we can not predict which A:B complexes will have disassembled and which will remain intact. Why? Because we cannot predict exactly which collisions will provide sufficient

energy to disassociate a particular A:B complex.[224] Dissociation is a stochastic process, and like all stochastic processes (such as genetic drift) is best understood in terms of probabilities.

Stochastic processes are particularly important within biological systems because, generally, cells are small. They contain small numbers of specific molecules. A typical cell contains one or two copies of a particular gene. The expression of those genes depends upon the binding of specific proteins to specific regions of a DNA molecule. If there are relatively small numbers of those proteins present in a cell, whether or not a copy of the protein is bound to a specific DNA sequence will be stochastic.[225] If there are enough cells, then the group average may be predictable, but the behavior of any one cell may not be.[226] In an individual cell, sometimes the protein will be bound and the gene will be expressed and sometimes not, all because of thermal motion and the small numbers of interacting molecules involved. This stochastic property of cells can play important roles in the control of cell and organismic behaviors.[227] It can transform a genetically identical population of cells and organisms into subpopulations that display distinct behaviors with functional implications.

Questions to answer:

62. Explain how temperature influences intermolecular interactions? How might changes in temperature influence macromolecular shape?
63. Under what conditions will the effects of temperature on covalent bonds be significant in biological systems?
64. Why does population size matter when generating a graph that describes radioactive decay or the dissociation of a complex, like the A:B complex discussed above?

Questions to ponder:

Why is the Boltzmann distribution more asymmetric around the highest point as temperature decreases.

References

[223] zero point energy (from wikipedia)

[224] In theory at least, we might be able to make this prediction if we mapped the movement of every water molecule. This is different from radioactive decay, where it is not even theoretically possible to predict the behavior of an individual radioactive atom.

[225] This is illustrated here and we will return to this type of behavior later on.

[226] Biology education in the light of single cell/molecule studies

[227] Single Cells, Multiple Fates, and Biological Non-determinism: https://www.ncbi.nlm.nih.gov/pubmed/27259209

This page titled 5.7: Bond Stability and Thermal Motion (a non-biological moment) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.8: Bond polarity, inter- and intramolecular interactions

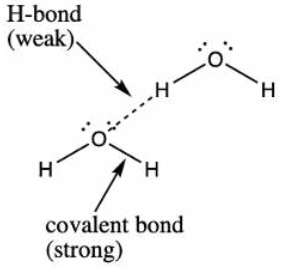
Atoms of each element have a characteristic "electronegativity", a measure of how tightly they hold onto their electrons. If the electronegativities of the two atoms in a covalent bond are equal or close, then the electrons are shared more or less equally between the two atoms. The resulting bond is said to be non-polar, meaning without direction. There are no stable regions of net negative or net positive charge on the resulting molecular surface due to the presence of the bond. When the electronegativities of the two bonded atoms are unequal the electrons of the bond will be shared un-equally. On average more electrons will spend more of their time around the more electronegative atom and fewer around the less electronegative atom. The result is what is known as a polar bond, with partially negatively and partially positively-charged regions – a polar bond has a direction and leads to an electrical field known as a dipole.

O and N atoms are more electronegative than C and H (see table, next page). When O or N form a bond with C or H, the O and N become partly negative and the C and H become partly positive - the bond is polarized. In contrast, there is no significant charge polarization in bonds between C and H; such bonds are termed non-polar. The presence of polar bonds can lead to electrostatic interactions between molecules that is stronger than "simple" van der Waals interactions. While stronger than non-polar LDF-mediated interactions they are weaker than covalent bonds. Like covalent bonds, polar interactions have a directionality to them – the three atoms involved are arranged more or less along a straight line. There is no such geometric constraint on simple LDF-mediated interactions. Since the intermolecular forces arising from polarized bonds often involve an H atom interacting with an O or an N atom, these are known generically, and perhaps unfortunately, as hydrogen or H-bonds (→). Why unfortunate? H atoms take part in covalent bonds, but H-bonds are not covalent, they are weaker. It takes much less energy to break an H-bond between molecules or within (generally macro-) molecules than it does to break a covalent bond involving a H atom.

This page titled 5.8: Bond polarity, inter- and intramolecular interactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.9: The implications of bond polarity

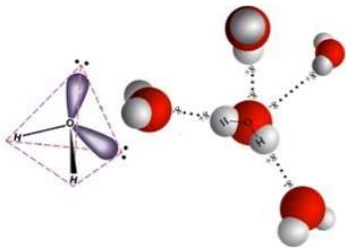
Melting and boiling points are important physical properties of small molecules. They apply to pure samples composed of large numbers of molecules. They are irrelevant to larger molecules, which generally change their structure (denature) at temperature extremes. Let us start at a temperature at which the sample is liquid. The molecules are moving with respect to one another, there are transient interactions between the molecules, who are constantly switching neighbors. As we increase the temperature of the system, the energetics of collisions are such that all interactions between neighboring molecules are eventually broken, and the molecules move away from one another. When they to collide with one another, they (generally) do not adhere; the bond that might form between them is not strong enough to resist the kinetic energy delivered by collisions with other molecules. The molecules are in a gaseous state and the transition from liquid to gas is the boiling point. Similarly, starting with a liquid, when we reduce the system temperature, the interactions between molecules become longer lasting until a temperature is reached at which the energy transferred through collisions is no longer sufficient to disrupt the interactions between molecules.[228] As more and more molecules interact, the positions of neighboring molecules become more highly constrained - the liquid transitions into a solid. Because neighboring molecules in a liquid can move with respect to one another, liquids flow and assume the shape of their containers. Solids, in contrast, maintain their shape – neighboring molecules stay put. The temperature at which a liquid changes to a solid is known as its melting point. These temperatures mark what are known as phase transitions: solid to liquid and liquid to gas.



At the macroscopic level, we see the dramatic effects of bond polarity on melting and boiling points by comparing molecules of similar size with and without polar bonds and the ability to form H-bonds (↓). For example, CH_4(methane) or Ne (neon) do not contain polar bonds and so do not form intra-molecular H-bonds. In contrast NH_3 (ammonia), H_2O (water), and FH (hydrogen fluoride) have three, two and one polar bonds, respectively, and so can take part in one or more intra-molecular H-bonds. All five compounds have the same number of electrons, ten. When we look at their melting and boiling point temperatures, we see how the presence of polar bonds influences these properties. In particular, water stands out as dramatically different from the rest, with significantly higher (> 70ºC) melting and boiling points than its neighbors.

Compounds	CH_4	NH_3	OH_2	FH	Ne
molecular weight	16.04	17.02	18.02	20.01	20.18
bond electronegativity	0.45	0.94	1.34	1.88	N/A
# of electrons	10	10	10	10	10
# of bonds	4	3	2	1	0
melting point	-182°C	-77.7°C	**0°C**	-83°C	-248.6°C
boiling point	-161.5°C	-33.4°C	**100°C**	19.5°C	-246.1°C

Why is water so different?



In addition to the presence of polar covalent bonds, we have to consider a molecule's shape. Each water molecule has two partially positive H atoms and two partially negative regions on its O atom. These sites of potential H-bonds are arranged in a nearly tetrahedral geometry (→). Because of this arrangement, each water molecule can interact through H-bonds, with four neighboring water molecules. To remove a molecule from its neighbors, four H-bond-type electrostatic interactions must be broken, which is relatively easy, energetically, since they are each rather weak. In the liquid state, molecules jostle one another and change their H-bonded interaction partners constantly. Even if one or two interactions are broken the water molecule is likely to remain linked to multiple neighbors via the remaining H-bonds.

This molecular hand-holding leads to water's comparatively high melting and boiling points as well as its high surface tension. We can measure the strength of surface tension in various ways. The most obvious is the weight that the surface can support. Water's surface tension has to be dealt with by organisms that interact with a liquid-gas interface. Some, like the water strider, use it to cruise along the surface of ponds (←). For the water strider to "walk" on the water surface the molecules of its feet cannot form H-bonds, they are said to be hydrophobic. That is clearly a bad name - they are not afraid of water, rather they are apathetic to it. Hydrophobic molecules interact with other molecules, including water molecules only through LDF-mediated interactions.



Molecules that can make H-bonds or other polar interactions with water are termed hydrophilic. As molecules increase in size they can have both hydrophobic and hydrophilic regions. Molecules with distinct hydrophobic and hydrophilic regions are termed amphipathic; we consider them in the next chapter.

References

[228] The nature of the geometric constrains on inter-molecular interactions will determine whether the solid is crystalline or amorphous. see: https://en.wikipedia.org/wiki/Crystal

This page titled 5.9: The implications of bond polarity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

5.10: Interacting with Water

We get an idea of the hydrophilic, hydrophobic, and amphipathic nature of molecules when we try to dissolve them in water. Molecules like sugars (carbohydrates), alcohols, and most amino acids are primarily hydrophilic, they dissolve readily in water. Molecules like fats are highly hydrophobic; they do not dissolve significantly in water. Why the difference? To answer this question we have to be clear about what we mean when we say that a molecule is soluble in water. We consider this from two perspectives. The first is what the solution looks like at the molecular level, the second is how the solution behaves over time. Consider what pure water looks like. Because of its ability to make multiple H-bond-type electrostatic interactions in a tetrahedral arrangement, water molecules form a dynamic three-dimensional intermolecular interaction network. In liquid water the H-bond mediated interactions between the molecules break and form rapidly.

To insert a molecule A, known as a solute, into this network you have to break some of the H-bonds between the water (solvent) molecules. If the A molecules can make H-bonds with water molecules, that is, if they are hydrophilic, then there is little net effect on the free energy of the system. One H-bond is replaced by another. Such a molecule is soluble in water. So what determines how soluble the solute is? As a first order estimate, each solute molecule will have at least one layer of water molecules around it, otherwise it will be interacting with other solute molecules. If the number of the interacting solute molecules is large enough, the solute will no longer be in solution, but associated with itself. Small aggregates of solute molecules can, when small enough, remain suspended in the solution, a situation known as a colloid. The cytoplasm of a cell behaves like a colloid in many ways. While a solution consists of individual solute molecules surrounded by solvent molecules, the cytoplasm has molecules in solution and colloidal states of varying stability.[229]

Now consider the behavior of hydrophobic solute molecules. Such molecules cannot make H-bonds with water molecules, so when inserted into water the total number of H-bond-type electrostatic interactions per unit volume decreases - the energy of the system increases (remember, bond forming lowers potential energy). Much of this “enthalpy” change, indicated as ΔH, is compensated for by LDF-interactions between the molecules. Generally, the net enthalpic effect is small. Something else must be going on to explain the insolubility of such molecules.

Turning to entropy

In a liquid, water molecules will be found in a state that maximizes the number of H-bonds present. Because these interactions have a tetrahedral geometry, their presence constrains the possible orientations of molecules with respect to one another. This constraint is captured when water freezes; it is the basis for ice crystal formation, why the density of water increases before freezing and then decreases with freezing, and why ice floats in liquid water.[230] Introduction of a small hydrophilic (H-bond forming) solute molecule will have little effect on the number of equivalent ways water and solute molecules interact. But introducing hydrophobic molecules that cannot make H-bonds will lead to significant constraints on the energetically favorable orientations of water molecules, leading to increased order (decreased entropy) in the system. How such hydrophobic and mixed hydrophilic and hydrophobic molecules, known generically as lipids behave is the subject of the next chapter.

Questions to answer:

65. How does the ability to make H-bond influence a molecule’s boiling point.
66. Predict and explain, the factors that influence the solubility of a molecule in water
67. Why does the separation of oil and water represent a more disordered state?
68. How would you explain to a "normal" person how it is possible for a water strider to walk on water; or why ice floats – what concepts would you need to introduce them to?

Question to ponder:

What would happen to a water strider if its "feet" suddenly became hydrophilic?

Short Chapter Summary

- Thermodynamics sets the possible; coupling makes the useful happen.
- Rates depend on activation energies; enzymes reshape pathways, not laws.
- Polar/non-polar interactions, H-bonds, ionic/van der Waals forces drive structure & function.

- Water's polarity + entropy dominate biological self-assembly.

References

[229] McSwiggen et al., 2021. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences

[230] Why does ice float in water? http://youtu.be/UukRgqzk-KE

This page titled 5.10: Interacting with Water is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

6: Membrane Boundaries and Capturing Energy

6.0: Introduction
6.1: Defining the Cell’s Boundary
6.2: The origin of biological membranes
6.3: Transport across membranes
6.4: Channels and carriers
6.5: Generating gradients- using coupled reactions and pumps
6.6: Simple Phototrophs
6.7: Chemo-osmosis (an overview)
6.8: Oxygenic Photosynthesis
6.9: Chemotrophs
6.10: Using the energy stored in membrane gradients
6.11: Osmosis and living with and without a cell wall
6.12: An evolutionary scenario for the origin of eukaryotic cells
6.13: Making a complete eukaryote

This page titled 6: Membrane Boundaries and Capturing Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.0: Introduction

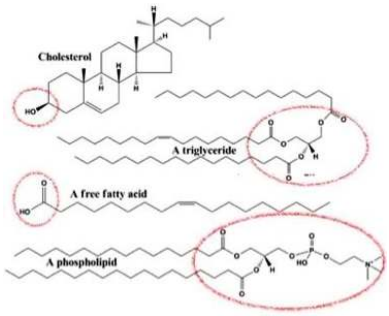


In which we consider how the aqueous nature of biological systems drives the formation of lipid-based barrier membranes and how such membranes are used to capture and store energy from the environment and chemical reactions. We consider how coupled reactions are used to drive macromolecular syntheses and growth, and how endosymbiotic events, involving the capture of aerobic and photosynthetic bacteria, played a critical role in the evolution of eukaryotic cells

This page titled 6.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

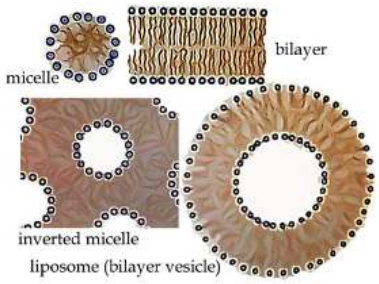
6.1: Defining the Cell's Boundary

An apparently necessary step in the origin of life was the generation of a discrete boundary layer that separates the living non-equilibrium reaction system from the rest of the universe. This boundary layer, the precursor of the plasma membrane, maintains the integrity of the intracellular environment and mediates the movement of molecules and energy into and out of the cell. So how is this membrane built and maintained, and how does it work? Simply put, when a cell divides, the plasma membrane of the "new" cell forms by budding off from the pre-existing membrane.



As a cell grows, new molecules are added into the existing membrane increasing its surface area. Biological membranes are composed primarily of two general classes of molecules, proteins (discussed in the next chapter) and lipids. Lipids are not a structurally coherent group, that is they do not have one common structure. Structurally distinct molecules, such as cholesterol and phospholipids, are both lipids (→). Lipids have two distinct domains: a hydrophilic domain (circled in red) characterized by polar regions and one or more hydrophobic domains that are usually made up of C and H atoms. While there is a relatively small set of common lipid types, there are many different lipids found in biological systems and the characterization of their structures and functions has led to a new area of analysis known as lipidomics.[231]

Lipids are amphipathic. In aqueous solution, entropic effects act to drive the hydrophobic parts of the lipid out of an aqueous solution. In contrast to totally non-polar molecules, like oils, the hydrophobic part of the lipid is attached to a hydrophilic domain. Lipid molecules deal with this structural dichotomy by associating with other lipid molecules in multimolecular structures in which the interactions between the hydrophilic parts of the lipid molecule and water molecules are maximized and the interactions between the lipid's hydrophobic parts and water are minimized. Many such multi-molecular structures can fulfill these constraints (→). The structures that form depend upon the details of the system, including the shapes of the lipid molecules involved and the relative amounts of water and lipid present. In every case, the self-assembly of these structures involves an increase in the total overall entropy of the system, a perhaps counterintuitive result. For example, in a micelle the hydrophilic region is in contact with the water, while the hydrophobic regions are inside, away from direct contact with water. This leads to a more complete removal of the lipid's hydrophobic domain from contact with water than can be arrived at by a purely hydrophobic oil molecule, so unlike oil,



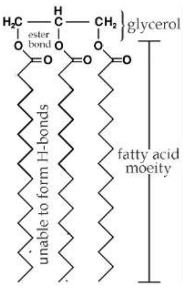
lipids can form stable structures in solution. The diameter and shape of the micelle is determined by the size of its hydrophobic domain. As this domain gets longer, the center of the micelle becomes more crowded. A type of organization that avoids "lipid-tail crowding" is known as a bilayer vesicle. Here there are two layers of lipid molecules, pointing in opposite directions. The inner layer surrounds a water-filled region, the lumen of the vesicle, while the outer layer interacts with the external aqueous environment. In contrast to the situation within a micelle, the geometry of a vesicle means that there is significantly less crowding as a function of lipid tail length. Crowding is further reduced as a vesicle increases in size to become a cellular membrane. Micelles and bilayer vesicles can form colloid-like systems with water; they exist as distinct structures that can remain suspended in a stable state. We can think of the third type of structure, the planar membrane, as an expansion of the vesicle to a larger and more irregular size. Both surfaces of the membrane face aqueous environments, the cytoplasm and the extracellular environment. For the cell to grow, new lipids need to be inserted into both inner and outer layers. Since lipids are synthesized in the cytoplasm, they are generally initially inserted into the cytoplasmic-facing membrane layer and then "flipped" to the extracellular-facing layer, an event that involves interactions with proteins, known as flippases. When we consider proteins, consider the energetics of this reaction and plausible flipping mechanisms.

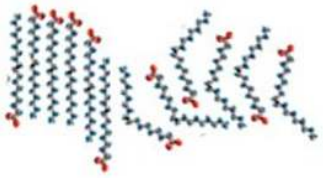
A number of mechanisms are used to insert molecules into membranes, but they all involve a pre-existing membrane – another aspect of the continuity of life. Totally new cellular membranes generally do not form. For example, a vesicle is a spherical lipid bilayer that can fuse into or emerge from a planar bilayer membrane. These processes are typically driven by protein-based molecular machines coupled to thermodynamically favorable reactions. When the membrane involved is the plasma (boundary) membrane, these processes are known as endocytosis and exocytosis (into and out of the cell), respectively. These terms refer

explicitly to the fate of the material within the vesicle. Exocytosis releases material in the vesicle interior into the outside world, whereas endocytosis captures material from outside of the cell and brings it into the cell. Within a cell, vesicles can fuse with and emerge from one another.

Cells contain many hundreds of different types of lipids, generated by a variety of biosynthetic pathways catalyzed by gene-encoded proteins. We consider two generic classes, the glycerol-based lipids (←) and cholesterol. Considerations of their structures illustrate general principles related to membrane behavior. In bacteria and eukaryotes, glycerol-based lipids are typically formed from the highly hydrophilic molecule glycerol combined with two or three fatty acid molecules (a three fatty acid chain molecule →). Fatty acids contain a long hydrophobic hydrocarbon chain with a polar (carboxylic acid) head group. The molecular nature of these fatty acids influences the behavior of the membrane formed. Often these fatty acids have what are known as saturated hydrocarbon tails. A saturated hydrocarbon contains only single bonds between the carbon atoms of its tail domain. While these chains can bend and flex, they tend to adopt a more or less straight, closely packed configuration, that maximizes the lateral (side to side) LDF-mediated interactions between them. Because of the extended surface contact between the chains, lipids with saturated hydrocarbon chains are typically solid around room temperature. Solid means that the molecules rarely exchange positions with one another. On the other hand (←), there are cases where the hydrocarbon tails are "unsaturated"; they contain CC double bonds (–C=C–). Because unsaturated hydrocarbon chains have permanent kinks due to the rigid nature and geometry of C=C bonds they cannot pack as regularly as saturated hydrocarbon chains. The less regular packing means that there is less interaction area between the molecules, which lowers the strength of the LDF-mediated interactions between them. Lower LDF-mediated interaction energy in turn, lowers the temperature at which these bilayers change from a solid (no or little movement of the lipids relative to one another) to a more liquid state with relatively free movements within the plane of the membrane. Recall that the strength of interactions between molecules determines how much energy is needed to overcome the interactions between them.





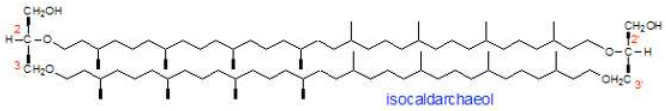


Because LDF-mediated intermolecular interactions are relatively weak, changes in temperature influence the physical state of the membrane. The liquid-like state is often referred to as the fluid state. The membrane's state is important because it can influence the movement, behaviors, and activities of proteins embedded in the membrane. If the membrane is in a solid state, proteins within the membrane will be relatively immobile. If it is in the liquid state, proteins can move by collision-driven movements within the plane of the membrane (diffusion). Since lipids and proteins closely associate with one another in the membrane, the physical state of the membrane can influence the activity of embedded proteins, a topic to which we will return.

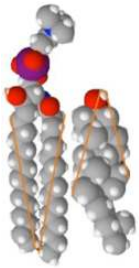
Cells can manipulate the solid-to-liquid transition temperature of their membrane(s) by altering the membrane's lipid composition. Increasing the ratio of saturated to unsaturated chains increases the melting temperature. Controlling chain saturation involves altering the activities of the enzymes involved in various saturation/desaturation reactions. That these enzymes can be regulated implies a feedback mechanism by which either temperature, membrane fluidity, or protein activity act to regulate metabolic processes and gene expression. This type of feed back mechanism is part of the homeostatic and adaptive systems of cells (and organisms), a topic we will return to repeatedly.



There are a number of differences between the lipids used by bacterial and eukaryotic organisms and archaea.[232] Most dramatically difference is found in archaea which use branched isoprene (CH_2=C(CH_3)CH=CH_2) polymers, instead of straight chained hydrocarbons, that are linked to the glycerol group through an ether, rather than an ester linkage (→). The bumpy and irregular shape of the isoprene groups, compared to the relatively smooth saturated hydrocarbon chains, means that archaeal membranes tend to melt (go from solid to liquid) at lower temperatures.[233] At the same time the ether linkage is more stable (requires more energy to break) than the ester linkage. One speculation to explain this difference is that the archaea were originally (or became) adapted to living at higher temperatures, where the greater stability of the ether linkage would provide a critical advantage.



Some archaea and bacteria, known generically as thermophiles and hyper-thermophiles, live (happily, apparently) at temperatures up to 110 ºC.[234] At the highest temperatures, thermal motion might be expected to disrupt membrane integrity, allowing small charged molecules (ions) and larger hydrophilic molecules to pass through the plasma membrane.[235] Given the importance of membrane integrity, you may not be surprised to find “double-headed” lipids in such thermophilic organisms (→). These molecules have two distinct hydrophilic glycerol "heads", one located at each end of the molecule. A single molecule can span the membrane. The presumption is that such lipids act to stabilize the membrane against the disruptive effects of high temperatures.



The solid-fluid nature of biological membranes, as a function of temperature, is complicated by the presence of cholesterol and structurally similar lipids. For example, in eukaryotes the plasma membrane can contain as much as 50% cholesterol, in terms of the number of molecules present. Cholesterol has a short bulky hydrophobic domain (hydrocarbon chain lipid left, cholesterol to the right →) that does not pack well with other lipids. The presence of cholesterol dramatically influences the solid-liquid behavior of the membrane. The various roles of lipids is complex and goes beyond our scope here.

References

[231] On the future of "omics": lipidomics & Lipidomics: new tools and applications

[232] A re-evaluation of the archaeal membrane lipid biosynthetic pathway

[233] The origin and evolution of Archaea: a state of the art

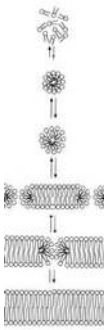
[234] You might consider how this is possible and under want physical conditions you might find these “thermophilic” archaea.

[235] Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea

This page titled 6.1: Defining the Cell’s Boundary is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.2: The origin of biological membranes

The hydrophobic “tails” of modern lipids range from 16 to 20 carbons in length. The earliest membranes, however, were likely to have been composed of molecules with shorter hydrophobic chains. Based on the properties of lipids, we can map out a plausible scenario for the appearance of membranes. Lipids with very short hydrophobic chains, from 2



to 4 carbons in length, dissolve in water (can you explain why?) As the lengths of the hydrophobic chains increases, the molecules begin to self-assemble into micelles. By the time the hydrophobic chains reach ~10 carbons in length, it becomes more difficult to fit the hydrocarbon chains into the interior of a micelle without making larger and larger spaces between the hydrophilic heads. Water molecules can begin to move through these spaces and interact with the hydrocarbon tails. At this point, the hydrocarbon-chain lipid molecules begin to associate with one another into semi-stable bilayers (→). One interesting feature of bilayers is that the length of the hydrocarbon chain is no longer structurally limiting, in contrast to the situation in micelles. One problem, though, are the edges of the bilayer, where the hydrocarbon region of the lipid would come in contact with water, a thermodynamically unfavorable situation. This problem is avoided by linking edges of the bilayer to one another, forming a closed balloon-like structure. Such bilayers can capture regions of solvent, that is water and the solutes dissolved within it.

Bilayer stability increases further as hydrophobic chain length increases. At the same time, membrane permeability decreases. It is a reasonable assumption that the earliest biological systems used shorter chain lipids to build their "proto-membranes" and that these membranes were relatively leaky.[236] The appearance of more complex lipids, capable of forming more impermeable membranes, likely depended upon the appearance of mechanisms, presumably protein-based, that enabled hydrophilic molecules to pass through the membranes. The interdependence of these changes is known as co-evolution. Co-evolutionary processes were apparently common enough to make the establishment of living systems possible at least under the conditions that existed on the early Earth.

Questions to answer:

69. Draw diagrams to show how increasing the length of a lipid's hydrocarbon chains affects the structures that it can form and use your diagrams to predict how the effects at the hydrophobic edges of a lipid bilayer are minimized? Why are there no triple bonds in fatty acid chains?
70. Some lipids have negatively-charged phosphate groups attached to the glycerol as well as fatty acids - how does the presence of “phospho-lipids" will impact membrane structure and stability.
71. Make a set of general rules on the effects of size and composition on the ability of a molecule to pass through a membrane.

Questions to ponder:

- Why do fatty acid and isoprene lipids form similar bilayer structures?
- Why might early (evolutionarily) membrane be expected to be leakier than modern membranes?

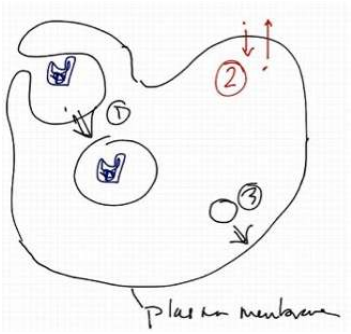
References

[236] Jack Szostak (two videos): The origin of life on Earth & Protocell membranes

This page titled 6.2: The origin of biological membranes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.3: Transport across membranes

As we have said before (and will say again), the living cell is a historically continuous, and ancient non-equilibrium system. To maintain its living state both energy and matter have to move into and out of the cell. This leads us to consider intracellular and extracellular environments and the properties of the boundary membrane that separates them. The differences between the regions inside and outside of the plasma membrane are profound. Outside, even for cells within a multicellular organism, the environment is generally mostly water, with relatively few complex molecules. Inside the membrane-defined space is the cytoplasm, a highly concentrated (300 to 400 μg/μl) solution of proteins, nucleic acids, smaller molecules, and thousands of interconnected chemical reactions.[237]



A lipid bilayer membrane poses a barrier to the movement of molecules. For larger molecules, particles or other organisms it also acts as a physical barrier. Typically when larger molecules, particles (viruses), and other organisms enter a cell, they are first engulfed by the membrane (process 1 known as endocytosis)(→).[238] A superficially similar process, exocytosis, runs in "reverse" (process 3 →) and is involved in moving molecules to the cell surface and releasing them into the extracellular space. Both endocytosis and exocytosis involve membrane vesicles emerging from or fusing into the plasma membrane. These processes leave the topology of the cell unaltered; a molecule within a vesicle is still "outside" of the cell, or at least outside of the cytoplasm. These movements are driven by various protein-based molecular machines (considered in more specialized cell biology courses). There are also processes that directly move molecules through the membrane (process 2 ↑).

How does the membrane "decide" which molecules to allow into or out of the cell? If we think about it, there are three possible mechanisms (can you think of others?) Molecules could move on their own through the membrane, driven by Brownian motion. Alternatively, their diffusion-based movement could be mediated by specific "carriers" or "channel" molecules. Finally they could be moved directionally using a "pump", an energy dependent process involving coupled reactions. In the majority of cases, these carriers, channels, and pumps are protein-based molecular machines. The types of carriers, channels, and pumps that are present determine the types of molecules that move through the cell's membrane, as well as their net flux into or out of the cell. We can think of this molecular movement as a reaction:

$$[\ \text{Molecule}_{\ \text{outside}}\ ] \rightleftharpoons [\ \text{Molecule}_{\ \text{inside membrane}}\ ] \rightleftharpoons [\ \text{Molecule}_{\ \text{inside cell}}\ ]$$

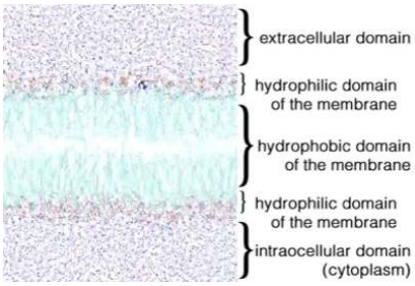
As with standard chemical reactions, movements through membranes involve an activation energy; this includes the energy needed to remove a water soluble molecule from aqueous solution and then for it to pass through the membrane. Ask yourself why does the membrane, particularly the hydrophobic region of the membrane, pose a barrier to the movement of hydrophilic molecules. The answer involves the difference in the free energy of the moving molecule within an aqueous solution, including the hydrophilic surface region of the membrane, where H-bond type electrostatic interactions are common between molecules, and the hydrophobic region of the membrane, where only LDF-mediated interactions are present. The situation is exacerbated for charged molecules, since water molecules are typically organized in a dynamic shell around ions.

The identity of the moving molecules does not change during the transport reaction. If the concentrations of the molecules are the same on both sides of the membrane, then their Gibbs free energies are also equal, the system will be in equilibrium with respect to this reaction. In this case, as in the case of chemical reactions, there will be no net flux of molecules across the membrane, but molecules will continue to move back and forth at equal rates. The rate at which they move will depend on the size of the activation energy associated with moving across the membrane as well as the concentrations of the molecules on each side of the membrane. Molecules can be concentrated within a cell, but that requires coupling molecular movement to another, energy-dependent reaction.

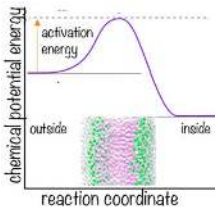
To think about how molecules cross lipid membranes, let us begin with water itself, which is small and uncharged, although polarized. Typically, the concentration of water outside of a cell is greater than the concentration of water inside a cell. This implies that the reaction:

$$H_2O\ \text{outside} \rightleftharpoons H_2O\ \text{inside}$$

will be favorable; there will be a net flux of water molecules into the cell. What is happening in this reaction? As a water molecule moves through water, H-bonds are broken and reform - there is no net energetic change. In contrast, when a water molecule begins

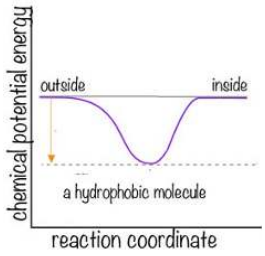


to leave the aqueous phase the H-bonds between it and neighboring water molecules must be broken but no new H-bonds form as the molecule enters the hydrophobic (central) region of the membrane. This asymmetry in H-bonding results in water molecules being "pulled back" into the water phase (←). In this video of a water molecule is moving through a membrane; can you explain why the water molecules moves faster in the hydrophobic region than in water?



The $\mathbf{Water}_{\mathbf{outside}} \rightleftharpoons \mathbf{Water}_{\mathbf{inside}}$ reaction's activation energy (→) involves breaking H-bonding interactions between water molecules and the membrane's hydrophilic lipid head domains. Thermal movement is generally sufficient to break these H-bonds. Once they enter the membrane's hydrophobic region, water molecules pass through more easily, since only weak LDFs are involved.

Small non-polar molecules, such as O_2 and CO_2 can readily pass through biological membranes. There is more than enough energy available through collisions with other molecules (thermal motion) to provide the energy needed to overcome the activation energy involved in leaving the aqueous phase and passing through the molecular domains of the membrane. As with water, there are often differences in the free energies of the molecules on the inside and outside of the cell. Consider animals that depend upon O_2 (obligate aerobes). The $[O_2]$ outside of the cell (produced by plants as a waste product) is carried throughout the organism's interior by its circulatory system. After O_2 enters the cell, it takes part in the reactions of respiration (considered soon), leading to a decrease in the O_2 concentration gradient, $[O_2]_{outside} > [O_2]_{inside}$. The result is a net flux of O_2 into the cell.



Another perspective into membrane behavior involves considering the interactions of different types of molecules within a bilayer membrane. If a molecule is hydrophobic (non-polar) it will be more "soluble" (concentrated) in the membrane's central hydrophobic region than in the surrounding aqueous environment (←). A totally hydrophobic molecule will accumulate within the membrane; its entropic effects on water structure will oppose its entry to the aqueous phase.

Questions to answer:

72. Consider the reaction diagram for flipping a lipid molecule's orientation by 180º perpendicular to the plane of the membrane: what energy barriers are associated with such a movement?
73. Draw a graph to show how the potential energy changes as an ion moves across a membrane. What is involved when an ion leaves the aqueous phase? How would this differ from a hydrophobic molecule?
74. What do you expect to happen to the O_2 gradient if an aerobic cell's ability to use O_2 is inhibited?

References

[237] A model of intracellular organization + Wee & Chisowa 2025. The Great Unknown: How Chemistry Remains the Last Frontier to Understanding Life

[238] These processes, ranging from pinocytosis (cell drinking) to phagocytosis (cell eating) involve different molecular machines.

This page titled 6.3: Transport across membranes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.4: Channels and carriers

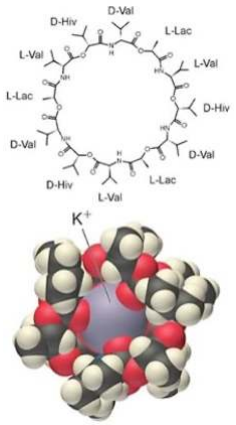
Beginning around the turn of the last century, a number of scientists began working to define the nature of the cellular boundary layer. Small, water soluble molecules were found to enter cells faster than predicted based on the assumption that the membrane acts like a simple hydrophobic barrier. Ernest Overton (1865-1933) and Runar Collander(1894-1973) postulated that membranes were more than simple barriers, specifically that they contained features that enabled them to act as highly selective molecular sieves.[239] Most of these features are proteins (getting closer to discussing proteins, promise) that can act as channels, carriers, and pores. If we think about crossing the membrane as a reaction, the activation energy will be high for highly hydrophilic and larger molecules. We will need a catalyst to reduce the activation energy so that the reaction proceeds at a reasonable rate. There are two generic types of membrane permeability catalysts: carriers and channels.

Carrier proteins are membrane proteins that shuttle back and forth across the membrane. They bind to specific hydrophilic molecules, hold on to the bound molecule as they traverse the membrane's hydrophobic region, and then release their "cargo" when they reach the other side of the membrane. Both the movements of carrier and cargo across the membrane, and the release of transported molecules, are stochastic, driven by thermal motion (molecular collisions), so no other energy source is needed. We can write this class of reactions as:

$$\begin{aligned}\text{Molecule}_{\text{outside}} + \text{carrier}_{\text{empty}} &\rightleftharpoons \text{carrier}-\text{Molecule}_{\text{outside}} \rightleftharpoons \text{carrier}-\text{Molecule}_{\text{inside}} \\ &\rightleftharpoons \text{carrier}_{\text{empty}} + \text{Molecule}_{\text{inside}}\end{aligned}$$

There are many different types of carrier proteins and each type of carrier has a preferred cargo. Related molecules may be bound and transported, but with less specificity and so at a much lower rate. Exactly which molecules a particular cell will allow to enter the cell will be determined in part by which carrier protein genes it expresses. Mutations in a gene encoding a carrier can change (or abolish) the range of molecules that that carrier can transport across a membrane.

Non-protein carriers: An example of a non-protein membrane carrier is a class of antibiotics, known generically as ionophores. They kill cells by disrupting the normal ion balance across the cell's membrane. By carrying specific ions across membranes they disrupt the cell's normal metabolic activity.[240] One of these ionophore antibiotics is valinomycin (→), a molecule made by *Streptomyces* type bacteria.[241] The valinomycin molecule has a hydrophobic periphery and a hydrophilic core. It binds K^+ ions ~10^5 times more effectively than it binds Na^+ ions.



In the absence of specific K^+ channels and pumps, K^+ cannot pass through the membrane, the activation energy is too high. Valinomycin molecules continually shuttle back and forth across the membrane. In the presence of a K^+ gradient, that is a higher concentration of K^+ on one side of the membrane compared to the other, K^+ will tend to bind to the valinomycin molecule on the high K^+ concentration (cytoplasmic) side and be released on the low K^+ concentration (extracellular) side. The result is an increase in the net flux of K^+ across the membrane and the dissipation of the normal K^+ gradient. To be clear, in the absence of a gradient, K^+ ions will move across the membrane (in the presence of valinomycin), but there will be no net movement of K^+, no net flux. As you might imagine, there are analogous carrier systems that move hydrophobic molecules within the aqueous phase.

Channels: Channel proteins sit within a membrane and contain an aqueous channel that spans the membrane's hydrophobic region. Hydrophilic molecules of particular sizes and shapes can pass through this aqueous channel. Their movement involves a significantly lower activation energy than would be associated with their moving through the lipid part of the membrane in the absence of the channel. Channels are highly selective in terms of which molecules will pass. For example, there are channels which will, on average, pass 10,000 K^+ ions for every one Na^+ion.

Channel proteins can exist in two or more distinct structural states. For example, in one state the channel can be open and allow particles to pass through or it can be closed. Often the properties of these channels can be regulated, they can be turned on or off. As an example, the binding of small molecules to a channel protein can lead to channel opening. Channels do not, however, determine in which direction an ion will move - net flux is based on the gradients that pre-exist across the membrane.

Another method of channel control depends on the fact that channel proteins are embedded within a membrane and contain charged groups. As we will see, cells can (and generally do) generate ion gradients, that is a separation of charged species across their membranes. For example if the [K^+] concentration is higher on one side of the membrane, ions will (if movement is possible)

display a net flux down the gradient, from the region of higher to lower K^+ concentration.[242] In some cases, the generation of ion gradients can, in turn, produce an electrical field across the plasma membrane. As these fields change, they can produce (induce) changes in channel protein structure that can switch the channel from open to closed and vice versa. Organisms typically have many genes that encode specific channel proteins involved in a range of processes from muscle contraction to thinking. Again, channels do not determine the direction of molecular motion. The net flux of movement is determined by the presence of molecular gradients, with the thermodynamic driver being entropic factors. The movement of the molecules through the channel is driven by thermal motion.

Questions to answer:

75. What does it mean to move against a concentration gradient? Is this a favorable or unfavorable event?
76. Where does the energy involved in moving molecules come from?
77. What happens to the movement of molecules through channels and transporters if we reverse the concentration gradients across a membrane?
78. Draw a diagram to show how K^+ ions are transported by an ionophore across a membrane. Draw a graph to show how the potential energy changes as the ion moves. Be sure to include the relative concentrations.

Questions to ponder:

How might you prove that movements of molecules across a membrane occurs in the absence of a gradient.

References

[239] Does Overton still rule? http://www.nature.com/ncb/journal/v1/n8/full/ncb1299_E201.html

[240] There is little data in the literature on exactly which cellular processes are disrupted by which ionophore; in mammalian cells (as we will see) these molecules act by disrupting the energy storing ion gradients in mitochondria and chloroplasts, apparently.

[241] Valinomycin: https://en.wikipedia.org/wiki/Valinomycin

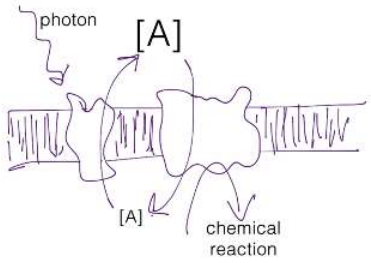
[242] In fact this tendency for species to move from high to low concentration until the two concentrations are equal can be explained by the Second Law of Thermodynamics. Check with your chemistry instructor for more details

This page titled 6.4: Channels and carriers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.5: Generating gradients- using coupled reactions and pumps

Both carriers and channels allow the movement of molecules across a membrane, but there is a net directional flux only when a concentration gradient is present - that is, if the concentration of the molecule is different on the two sides of the membrane. If a membrane contains active channels and carriers (as all biological membranes do) concentration gradients across the membrane will dissipate. The $[\text{molecule X}]_{\text{outside}}$ will become equal to $[\text{molecule X}]_{\text{inside}}$. Removing a concentration gradient across a cell's plasma membrane is a good way to kill the cell. When we look at cells we find lots of concentration gradients, which raises the question, what produces and maintains these gradients. The common sense, thermodynamically correct answer is that there must be molecules (generally proteins) that can transport specific types of molecules across the membrane and against a concentration gradient. We will call these types of molecules pumps and write the reaction they are involved in as:

$$[\text{ Molecule }]_{\text{low concentration}} + \text{pump} \leftrightarrow [\text{ Molecule }]_{\text{high concentration}} + \text{pump}$$



As you might suspect moving this reaction to the right is thermodynamically unfavorable; like a familiar macroscopic pump, it will require the input of energy to work. We will have to "plug in" our molecular pump into a source of energy in order to move a molecule against its concentration gradient. So, what energy sources are available? We have two choices: the system can use electromagnetic energy (light) or it can use chemical energy. In a light-driven pump, there is a system that absorbs photons ($\leftarrow$); the photon's energy is coupled to the pumping system. Where the pump is driven by a chemical reaction a thermodynamically favorable reaction ($\leftarrow$), catalyzed by the pump, is coupled to the movement of one or more molecules against their membrane-associated concentration gradients.

A number of chemical reactions can be used to drive such pumps and pumps can drive various reactions (remember reactions can move in both directions). One of the most common reactions involves the movement of energetic electrons through a membrane-bound, protein-based "electron transport" system; leading to the creation of an H^+ based electrochemical gradient. The thermodynamically favorable movement of H^+ "down" the concentration gradient can be coupled, through a membrane-bound ATP synthase enzyme, to a reaction that leads to the synthesis of adenosine triphosphate (ATP):

$$\text{H}^+_{(\text{extracellular})} \rightleftharpoons \text{H}^+_{(\text{intracellular})}$$
$$\text{ATP synthase (membrane-localized catalyst of intracellular reaction)}$$
$$\text{H}^+ + \text{adenosine diphosphate } (ADP) + \text{ phosphate } \rightleftharpoons \text{ adenosine triphosphate } (ATP) + \text{H}_2\text{O}$$

In the cytoplasm ADP, phosphate and H^+ react resulting in the release ATP and water. The thermodynamically favorable movement of H^+ down its concentration gradient is coupled to the thermodynamically unfavorable ATP synthesis reaction.

The reaction can also run in reverse, so that the thermodynamically favorable ATP hydrolysis reaction drives the export of H^+.

$$ATP + \text{H}_2\text{O} \rightleftharpoons ADP + \text{ phosphate } + \text{H}^+$$
$$\text{ATPase-driven pump (ATP synthase running 'backward')}$$
$$\text{H}^+_{(\text{intracellular})} \rightleftharpoons \text{H}^+_{(\text{extracellular})}$$

The same membrane protein, the ATP synthase/pump, makes it possible to use energy present in a chemical gradient (across a membrane) to drive ATP synthesis within the cell and can enable ATP hydrolysis to generate a concentration gradient.

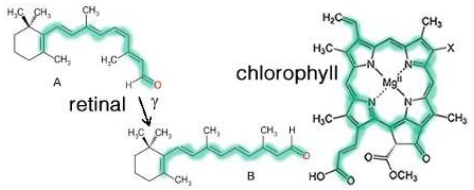
This page titled 6.5: Generating gradients- using coupled reactions and pumps is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

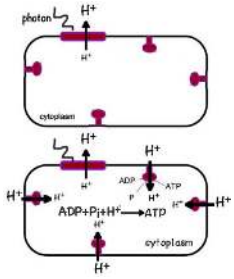
6.6: Simple Phototrophs

Phototrophs are organisms that capture photons (particles of light) and transform their electromagnetic energy into energy stored in unstable molecules, such as ATP and carbohydrates. Phototrophs “eat” light. Light can be considered as both a wave and a particle (that is quantum physics for you) and the wavelength of a photon reflects its "color" as perceived by a brain and the amount of energy it contains. Due to quantum mechanical considerations, a particular molecule can only absorb or emit photons of specific wavelengths (energies). This property makes possible spectroscopic methods, and enables us to identify molecules even when located at interstellar/intergalactic distances based on the photons they absorb or emit. Our atmosphere allows mainly visible light from the sun to reach the earth's surface, but most biological molecules do not absorb visible light. To capture this energy, evolution has led organisms to synthesize molecules, known as pigments, that can absorb visible light. The colors we see for a typical pigment are the colors of light that are not absorbed but reflected. For example chlorophyl appears green because light in the red and blue regions of the spectrum is absorbed while green light is reflected. So, how do organisms' use absorbed electromagnetic energy?

One of the simplest examples of a phototrophic system, a system that can directly capture the energy of light and transform it into the energy stored in a molecule, is provided by the archaea *Halobacterium halobium*.[243] *Halobacteria*is an extreme salt-loving organism, a halophile. It lives in waters that contain up to 5M NaCl. *H. halobium* uses the membrane protein bacteriorhodopsin to capture light. Bacteriorhodopsin consists of two components, a polypeptide, known generically as an opsin, and a non-polypeptide prosthetic group, the pigment retinal, a molecule related tovitamin A.[244] Together the two, opsin + retinal, form the functional bacteriorhodopsin protein.

When a molecule of retinal absorbs a photon, an electron moves from a lower to a higher energy molecular orbital. Extended molecular orbitals (← highlighted here) are associated with molecular regions that are drawn as alternating single and double bonds between carbon atoms, known as conjugated π orbital systems. Conjugated π systems are responsible for the absorption of light by pigments such as chlorophyll and heme, the pigment that makes blood red. Heme (not shown) includes an iron ion while chlorophyll includes a magnesium ion. When a photon of light is absorbed by the retinal group, it undergoes a reaction that leads to a change in molecular shape (↑) and composition, which in turn leads to a change in the structure of the polypeptide to which it is attached. This is called a photo-isomerization reaction. Isomers are similar molecules with different configurations.





Bacteriorhodopsin proteins are embedded within the plasma membrane where they associate with other bacteriorhodopsin molecules to form patches. These patches of bacteriorhodopsin give the organisms their purple color and are known as purple membranes. When a bacteriorhodopsin molecule (→) absorbs light, the change in its retinal group produces a light-induced change in protein structure that results in the movement of an H^+ ion from the inside to the outside of the cell. The protein and its associated pigment molecule then returns to its original low energy (ground) state; its state before it absorbed a photon. The return of bacteriorhodopsin to its ground state is NOT associated with the movement of an H^+ ion across the membrane. Because all of the bacteriorhodopsin molecules have the same orientation, as light is absorbed H^+ ions move in the same direction across the membrane, leading to the formation of an H^+ concentration gradient with $[H^+]_{outside} > [H^+]_{inside.}$ This H^+ gradient is associated with an electrical gradient because the movement of H^+ leads to more positive charge outside the cell. As light is absorbed the concentration of H^+ outside the cell increases and the concentration of H^+ inside the cell decreases. One question is, where are the moving H^+’s coming from? As you (perhaps) learned in chemistry, water can undergo a dissociation reaction (although this reaction is quite unfavorable):

$$H_2O \rightleftharpoons H^+ + OH^-$$

At pH 7.0 water contains 10^{-7} moles of H^+ and it is these H^+ s that move.

As the absorption of light drives the movement of H^+s across the membrane, they leave behind OH^- ions in the cytoplasm. The result is an electrical field with excess positive charges outside and excess negative charges inside the cell. As you may remember from physics, positive and negative charges attract, but the intervening membrane stops them from reuniting. The result is that

when exposed to light there is an accumulation of positive charge on the outer surface of the membrane and negative charge on the inner surface. This charge separation produces an electric field across the membrane. An H^+ ion outside of the cell will experience a force associated with the electric field together with the effect of a H^+ concentration gradient. If there is a way back across the membrane, both the electrical field and the $[H^+]$ gradient will act to drive the movement of H^+ ions back into the cell. The $[H^+]$ gradient acts as a battery, a source of energy that the cell can use.

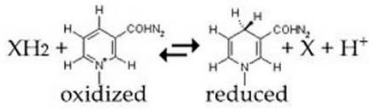


This page titled 6.6: Simple Phototrophs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.7: Chemo-osmosis (an overview)

One of the most surprising discoveries about biological systems is the almost universal use of H^+–based electrochemical gradients to generate ATP. Originally known as the chemiosmotic hypothesis, it was produced by the eccentric British scientist, Peter Mitchell (1920–1992).[245] Before the significance of H^+ transmembrane gradients was widely appreciated, Mitchell proposed that energy captured through the absorption of light (by phototrophs) or the breakdown of molecules (by chemotrophs) relied on the same basic homologous (evolutionarily-related) mechanism, the generation of H^+ gradients across membranes (the plasma membrane in prokaryotes and the internal membranes of mitochondria and chloroplasts, intracellular eukaryotic organelles derived from bacteria.

What makes us think that these processes have a common evolutionary origin? It is the observation that in both captured energy is transferred through the movement of electrons through a structurally similar membrane-embedded “electron transport chain” (ETC). Electron transport chains are composed of a series of membrane and membrane-associated proteins and involve a series of reduction-oxidation (redox) reactions. Where an electron is within a molecule's electron orbital system influences the amount of energy present in the molecule. Adding a negative charge (an electron) to a molecule can increase electron-electron repulsion and raise the molecule’s potential energy. When an electron is added to a molecule, that molecule is said to have been "reduced". It is weird that adding an electron "reduces" a molecule (→). Generally, when an electron is removed, the molecule's energy is decreased and the molecule is said to have been "oxidized".[246] Electrons, like energy, are neither created nor destroyed in biological systems, so the reduction of one molecule is always coupled to the oxidation of another. In a system of redox reactions, the energetic electrons removed from the reduced molecule are used to drive thermodynamically unfavorable reactions. Some of the energy difference is used to move H^+ ions across a membrane, generating a H^+ concentration gradient. Subsequently the thermodynamically favorable movements of H^+s down their concentration gradient (across the membrane) is used to drive ATP synthesis, a thermodynamically unfavorable reaction. ATP synthesis itself involves the rotating ATP synthase. The reaction can be written:



$$H^+_{outside} + ADP + Pi + H^+ \rightleftharpoons ATP + H_2O + H^+_{inside,}$$

where “inside” and “outside” refer to compartments defined by the membrane containing the electron transport chain and the ATP synthase, with the ATP synthesis reaction occurring within the membrane-bound compartment. When run in reverse (as noted above) the ATP synthase acts as an ATP hydrolase, pumping H^+s and other molecules against their concentration gradients. Such pumping ATPases establish most of the important ion gradients across membranes. In such a reaction:

$$ATP + H_2O + \text{ molecule in low concentration region} \rightleftharpoons ADP + \text{ Pi } + \text{ molecule in high concentration region.}$$

The key difference between phototrophs and chemotrophs is, essentially, where do the high energy electrons come from - the absorption of light or the breakdown of unstable molecules.

References

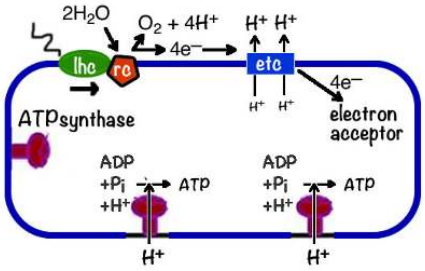
[245] Chemo-osmosis and Peter Mitchell (wikipedia)

[246] you can review redox here or in CLUE

This page titled 6.7: Chemo-osmosis (an overview) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.8: Oxygenic Photosynthesis

Compared to the salt loving archaea *Halobium*, with its purple bacteriorhodopsin-rich membranes, photosynthetic cyanobacteria (which are true or eubacteria), green algae, and higher plants (both eukaryotes) use more complex molecular systems through which to capture and utilize light. The photosynthetic systems of these organisms appear homologous.[247] For simplicity's sake we consider oxygenic or non-cyclic system photosynthetic system of cyanobacteria; look to more advanced classes for more detail about the complete photosystem. The major pigment in this system, chlorophyll, is based on a complex molecule, a porphyrin (see above). These pigments give plants their green color. As in the case of retinal, they absorb visible light due to the presence of a conjugated (resonance) bonding structure, typically drawn as a series of alternating single and double carbon-carbon bonds. Chlorophyll is synthesized by a conserved biosynthetic pathway. Variants of this scheme are used to synthesize heme, which is bound to the protein hemoglobin of animals and cytochromes, which are found within the electron transport chains of both plants and animals (which we will come to shortly), vitamin B_{12}, and other biologically important prosthetic (that is non-polypeptide) groups associated with proteins and required for their normal functions.[248]



Chlorophyll molecules are organized into two distinct membrane-embedded light absorbing protein complexes known as light harvesting (lhc) and reaction center (rc) complexes. Light harvesting complexes are important because photosynthetic organisms often compete with one another for light; increasing the efficiency of the system through which an organism captures light can provide a selective (reproductive) advantage. When a photon is absorbed, an electron is excited to a higher molecular orbital and then passed to a reaction center complex (→). In the oxygenic, molecular oxygen (O_2) generating, reaction system excited electrons are passed from the reaction center to the membrane-associated electron transport chain. As these electrons move through the electron transport chain their energy is used to move H^+s from inside to outside of the cell, or in mitochondria or chloroplasts. This is the same geometry of movement that we saw previously in the purple membrane system. In the non-cyclic process electrons from the electron transport chain are delivered to an electron acceptor. Generally this involves the absorption of a second photon, a mechanistic detail that need not trouble us here. In the cyclic form of photosynthesis, low energy electrons from the electron transport chain are returned to the reaction center directly, where they regenerate the pigment molecules to their original (before they absorbed a photon) state.

The laws of conservation imply that when electrons leave the photosynthetic system in the non-cyclic process they must be replaced. So where do these electrons come from? Here we see what appears to be a major evolutionary breakthrough. During the photosynthetic process, the reaction center couples light absorption to the oxidation (removal of electrons) from water molecules:

$$\text{light} + 2H_2O \rightleftharpoons 4H^+ + 4e^- + O_2.$$

The four electrons, derived from two molecules of water, pass to the reaction center, while the $4H^+$s contribute to the proton gradient across the membrane.[249] O_2 is a waste product of this reaction. Over millions of years, the light-driven release of O_2 changed the Earth's atmosphere from containing essentially 0% molecular oxygen to the current ~21% level at sea level. Because O_2 is highly reactive, this transformation is thought to have been a major driver of a number of subsequent evolutionary changes. There remain organisms that cannot use O_2 and cannot survive in its presence. They are known as obligate anaerobes. This distinguishes them from organisms that normally grow in the absence of O_2 but that can survive in its presence, known as facultative anaerobes.

In the past the level of atmospheric O_2 has changed dramatically; its level is based (primarily) on how much O_2 is released into the atmosphere by oxygenic photosynthesis and how much is removed by various reactions, such as the decomposition of plant materials. When large amounts of plant materials are buried before they can decay, such as occurred from ~360 to ~299 million years ago with the formation of coal beds during the Carboniferous period, the level of atmospheric O_2 increased dramatically, apparently reaching levels as high as ~35%. It is speculated that such high levels of atmospheric molecular oxygen made it possible for organisms without lungs (like insects) to grow to gigantic sizes.[250]

References

[247] Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes

[248] Mosaic Origin of the Heme Biosynthesis Pathway in Photosynthetic Eukaryotes:

[249] Photosystem II and photosynthetic oxidation of water: an overview

[250] When Giants Had Wings and 6 Legs

This page titled 6.8: Oxygenic Photosynthesis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.9: Chemotrophs

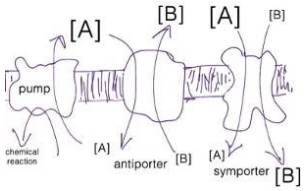
Organisms that are not phototrophic capture energy from other sources, specifically by transforming thermodynamically unstable molecules into more stable species. Such organisms are known generically as chemotrophs. They can be divided into various groups, depending upon the types of food molecules (energy sources) they use: these include organotrophs, which use carbon-containing molecules (you yourself are an organotroph) and lithotrophs or rock eaters, which use various inorganic molecules. In the case of organisms that can "eat" H_2, the electrons that result are delivered, along with accompanying H^+ ions, to CO_2 to form methane (CH_4) following the reaction:

$$CO_2 + 4H_2 \rightleftharpoons CH_4 + 2H_2O.$$

This page titled 6.9: Chemotrophs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.10: Using the energy stored in membrane gradients

The energy captured by organisms is used to drive a number of processes in addition to synthesis reactions. For example, we have already seen that ATP synthases can act as pumps (ATP-driven transporters), coupling the favorable ATP hydrolysis reaction to the movement of molecules against their concentration gradients (→). The resulting gradient is a form of stored (potential) energy, energy that can be used to move other molecules, molecules not moved directly by the ATP-hydrolysis driven transporter.[252] Such processes involve what is known as coupled transport.[253] They rely on membrane proteins that enable a molecule to pass through the membrane, and so allow for a net flux down a concentration gradient. In contrast to simple carriers and channels this thermodynamically favorable net flux from high concentration to low concentration is coupled to the movement of a second net flux against a gradient, that is from low to high concentration. When the two transported molecules move in the same direction, the transporter is known as a symporter; when they move in opposite directions, it is known as an antiporter. Which direction(s) the molecules move will be determined by the nature of the transporter protein and the concentration gradients of the molecules moved.



What is important here is that energy stored in the concentration gradient of one molecule can be used to drive the movement of a second molecule against its concentration gradient. In mammalian systems, it is common to have Na^+, K^+, and Ca^{2+} concentration gradients across the plasma membrane. These gradients are then used to transport other molecules into and out of cells. There are a large number of transporters that use H^+ ion gradients to move different types of molecules across cellular membranes; they appear to be evolutionarily ancient.[254] Of course, the presence of these gradients implies that there are ion-specific pumps that couple an energetically favorable reaction, typically ATP hydrolysis, to an energetically unfavorable reaction, the movement of an ion against its concentration gradient. Without these pumps, and the reactions that drive them, the membrane battery would quickly run down. Many of the immediate effects of death are due to the loss of membrane gradients and much of the energy needs of cells (and organisms) involves running pumps to maintain the non-equlibrium state of the cell.

References

[252] Although we will not consider it hear, membrane gradients are also used to send signals throughout the nervous system.

[253] Structural features of the uniporter/symporter/antiporter superfamily

[254] see Lichtinger et al., 2024. The mechanism of mammalian proton-coupled peptide transporters. eLife13:RP96507

This page titled 6.10: Using the energy stored in membrane gradients is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.11: Osmosis and living with and without a cell wall

Cells are packed full of molecules. These molecules take up space, space that cannot be occupied by water molecules. The concentration of water outside of the cell $[H_2O]_{out}$ will generally be higher than the concentration of water inside the cell $[H_2O]_{in}$. This water concentration gradient leads to the net movement of water molecules into the cell, a process known as osmosis.[255] This movement occurs across the cell's surface membrane, a membrane that is somewhat permeable to water (see above). A surprising finding that earned Peter Agre a share of the 2003 Noble prize in chemistry was that the membrane also contains water channels, known as aquaporins.[256] Follow the video link (→) to a molecular simulation of a water molecule (yellow) moving across a membrane through an aquaporin protein. Without aquaporins the rate of osmotic movement of water is dramatically slower. In addition to water, aquaporin-type proteins can facilitate the movement of other small uncharged molecules across cellular membranes.

The gradient in water concentration across the cell membrane, together with the presence of aquaporins, generates a system that can do work. The water gradient, can lift a solution against the force of gravity, a process involved in how plants stand up straight. How is this possible? If we think of a particular molecule in solution, it moves through collisions with its neighbors. These collisions drive the stochastic movement of particles. But if there is a higher concentration of molecules on one side of a membrane compared to the other, then the random movement of molecules will result in a net flux of molecules from the area of high concentration to that of low concentration, even though each molecule, on its own moves unpredictably. This video is a good illustration of the movement of a water molecule across a membrane. At steady state, the force generated by the net flux of water moving down its concentration gradient is balanced by other forces (e.g. the weight of the solution) acting in the other direction.

The water concentration gradient across the plasma membrane of most free-living organisms leads to an influx of water into the cell. As water enters, the plasma membrane expands. If the influx of water continues unopposed, the membrane will eventually burst like an over-inflated balloon, killing the cell. One strategy to avoid this lethal outcome, adopted by a range of organisms, is to build a semi-rigid “cell wall” external to the plasma membrane (↓). The synthesis of this cell wall involves the controlled assembly of macromolecules secreted by the cell. As osmosis "drives" water through the plasma membrane and into the cell, the plasma membrane is pressed up against the cell wall. The force exerted by the rigid cell wall on the membrane balances the force of water entering the cell. When the two forces are equal, the net influx of water into the cell stops, the systems arrives at a steady state. Conversely, if $[H_2O]_{outside}$ decreases, this pressure is reduced, the membrane moves away from the cell wall and, because they are only semi-rigid, the walls flex. It is this behavior that causes plants to wilt when they do not get enough water. These are passive behaviors, based on the structure of the cell wall; they are built into the wall as it is assembled. Once the cell wall has been built, a cell with a cell wall does not need to expend energy to resist osmotic effects. Plants, fungi, bacteria and archaea all have cell walls. A number of antibiotics work by disrupting the assembly of cell walls. This leaves the cells osmotically sensitive, water enters these cells until they burst and die.

Questions to answer:

88. Draw a graph of the water concentration across a typical cellular membrane for an organism living in fresh water; explain what factors influenced your prediction.
89. How might cell wall-less organisms deal with challenges associated with the absence of a cell wall?
90. Plants and animals are both eukaryotes; how would you decide whether the common ancestor of the eukaryotes had a cell wall. Do you think LUCA had a cell wall? What are some implications if it did not?
91. What are potential evolutionary benefits of losing a cell wall?
92. There is a concentration gradient of A across of membrane, but no net flux – what can we conclude?

Questions to ponder:

Why might an aquaporin channel not allow a Na^+ ion to pass through it?

References

[255] An important note is that in chemistry classes you may be taught that water moves from a region of low to high SOLUTE concentration. These two definitions of osmosis mean the same thing but it is easy to get confused.

[256] Water Homeostasis: Evolutionary Medicine

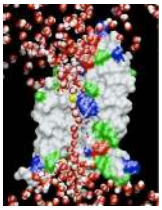
This page titled 6.11: Osmosis and living with and without a cell wall is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.12: An evolutionary scenario for the origin of eukaryotic cells

When we think about how life arose, and what the first organisms might have looked like, we are moving into an area where data is fragmentary or unobtainable and speculation is rampant. These events took place billions of years ago. But there is relevant data present in each organisms' genetic data, its genome, the structure of its cells, and their ecological interactions. It is this type of data that can inform and constrain our speculations.



Animal cells do not have a rigid cell wall. Its absence allows them to be active predators, moving rapidly and engulfing their prey whole or in macroscopic bits through phagocytosis. They use complex "cytoskeletal" and "cytomuscular" systems to drive these thermodynamically unfavorable behaviors (→). Organisms with a rigid cell wall can't perform such functions. Given that bacteria and archaea have cell walls, it is possible that cell walls were present in their common ancestor. This leads us to think more analytically about the nature of the earliest organisms and the path back to the common ancestor. A cell wall is a complex structure that would have had to be developed and assembled through evolutionary processes before it would be useful. If we assume that the original organisms arose in an osmotically friendly, non-challenging environment, then a cell wall could have been generated in steps, and once adequate it could enable the organisms that possessed it to invade new, more osmotically challenging environments. Another plausible scenario might be that the ancestors of the bacteria and archaea originally developed cell walls as a form of protection against predators. So who were these predators? Were they the progenitors of the phagocytic eukaryotes? If so, it might be that organisms in the eukaryotic lineage never had a cell wall, and that neither did the ancestor of bacteria and archaea. In this scenario, the development of eukaryotic cell walls by fungi and plants represents an example of convergent evolution. These structures would then be analogous rather than homologous to the cell walls of prokaryotes (bacteria and archaea).



But now a complexity arises, there are plenty of eukaryotic organisms, including microbes like the amoeba, that live in osmotically challenging environments. How do they deal with the movement of water into their cells? How could they follow their prey into the non-salty world? One approach is to actively pump the water that flows in back out using membrane pumps or perhaps an organelle similar to a "contractile vacuole". Water accumulates within a membrane-bounded intracellular membrane vesicle. As water accumulates the vesicle inflates. To expel this water, the vesicle connects with the plasma membrane and is squeezed by the contraction of a cytomuscular system; as a result the vesicle's contents are squirting out of the cell. The process of vacuole contraction is active; it involves work and requires energy.[257] One might speculate that such as cytomuscular (contractile) system might originally been involved in enabling a cell to move its membranes so to surround and engulf other organisms (phagocytosis). The resulting vacuole became specialized in killing and digesting the engulfed prey. When digestion was complete, this micro-stomach could fuse with the plasma membrane to discharge the waste, a process involving an active contractile system. It turns out that the molecular systems involved in driving active membrane movement are related to the systems involved in dividing the eukaryotic cell into two during cell division; a system distinctly different from that used by prokaryotes.[258] So which came first, distinct cell division mechanisms that led to differences in membrane behavior, with one leading to a predatory active membrane and the other to a passive membrane, perhaps favoring the formation of a cell wall? At this point it is hard (impossible?) to know for sure, we might best be agnostic while still able to make scientifically sensible speculation.

References

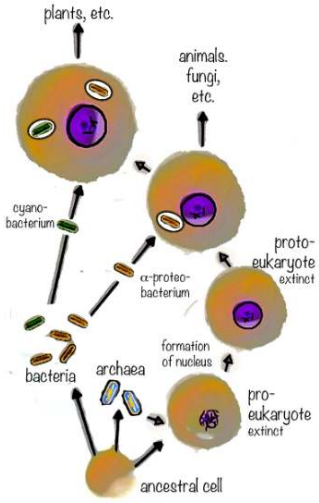
[157] Very cool video of a contractile vacuole in paramecium and explanation

[158] The cell cycle of archaea & Bacterial cell division

This page titled 6.12: An evolutionary scenario for the origin of eukaryotic cells is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

6.13: Making a complete eukaryote

Up to this point we have touched on only a few of the ways that prokaryotes (bacteria and archaea) differ from eukaryotes. A major difference includes the fact that eukaryotes have their genetic material isolated from the cytoplasm by a complex double-layered membrane/pore system known as the nuclear envelope (discussed later on). Exactly how the nucleus came into being in the lineage leading to eukaryotes remains poorly defined, as is often the case in historical processes that occurred billions of years ago.[259] Another difference is the relative locations of chemo-osmotic/photosynthetic systems in the two types of organisms. In prokaryotes, these systems (light absorbing systems, electron transport chains and ATP synthases) are located within the plasma membrane or plasma membrane-derived internal membrane vesicles. In contrast, in eukaryotes (plants, animals, fungi, protozoa, and other types of organisms) these structural components are found in discrete and distinctive intracellular structures. Systems associated with aerobic respiration are found in the inner membranes of mitochondria (singular: a mitochondrion). Photosynthetic eukaryotes (algae and plants) have mitochondria and a second type of membrane-bounded cytoplasmic organelle, known as chloroplasts. Both mitochondria and chloroplasts are characterized by the presence of a double membrane and electron transport chains located within the inner membrane and membranes apparently derived from it.



These are the type of structures one might expect to see if a bacterial cell was engulfed by an ancestral pro-eukaryotic cell, with the host cell's membrane surrounding the engulfed cells plasma membrane. Detailed molecular analyses revealed that the mitochondrial and chloroplast electron transport systems, as well as the ATP synthase proteins, closely resemble (are homologs of) those found in bacteria, rather than in archaea. In fact, detailed analyses of the genes and proteins involved suggest that the electron transport/ATP synthesis systems of eukaryotic mitochondria are homologous to those of α-proteobacteria while the light harvesting/reaction center complexes, electron transport chains and ATP synthesis proteins of algae and plants appear homologous to those of a second type of bacteria, photosynthetic cyanobacteria.[260] In contrast, many of the nuclear systems found in eukaryotes appear more similar to those systems present in archaea. How do we make sense of these observations?

This page titled 6.13: Making a complete eukaryote is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

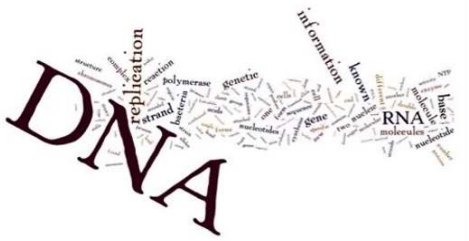
CHAPTER OVERVIEW

7: The Molecular Nature of Heredity

7.0: Introduction
7.1: Discovering how nucleic acids store genetic information
7.2: Locating hereditary material within the cell
7.3: Identifying DNA as the genetic material
7.4: Unraveling Nucleic Acid Structure
7.5: Discovering the structure of DNA
7.6: DNA, sequences, and information
7.7: Discovering RNA- structure and some functions
7.8: DNA Replication
7.9: Replication machines
7.10: Further replication complexities
7.11: Mutations, Deletions, Duplications and Repair
7.12: Genes and alleles
7.13: The origin of new (de novo) genes
7.14: DNA repeat diseases and genetic anticipation

This page titled 7: The Molecular Nature of Heredity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.0: Introduction



In which we discover how the physical basis of genetic inheritance, DNA, was identified and consider the factors that influence how it is that DNA encodes information, how that information is replicated, "read out" into RNAs and can be "translated" into the polypeptides that form proteins, how mutations occur and may be repaired, and how such extravagantly long molecules are organized within small cells.

An amazing fact associated with Darwin and Wallace's original evolutionary model was their lack of a coherent understanding of genetic mechanisms. While it was clear, based on the experiences of plant and animal breeders, that organisms varied with respect to one another and that part of that variation was inherited from the organism's parents, the mechanism(s) by which genetic information was stored and transmitted were unclear and, at the time, essentially unknowable. This situation promoted much speculation, including hypotheses based on supernatural or metaphysical mechanisms.[266] For example, some proposed that evolutionary variation was generated by an "inner drive" acting at organismic or even at the species level - an idea known as orthogenesis. Orthogenesis had the comforting implication that evolutionary processes reflect some form of purposeful design, that things were going somewhere; that there was a purpose to existence. On the negative side, such orthogenic models served to justify toxic racism. Different types of organisms (and different populations of people) represented different levels of "perfection".[267] An alternative model, proposed by Jean-Baptiste Lamarck (1744–1829), suggested that inheritance reflected the desires and experiences of the parent.[268] This model presumed an "internally directed" and purposeful form of evolution; the idea that evolutionary change reflects the desires, needs, and experiences of individuals. In contrast Darwin's model, based on random variations in the genetic material, seemed more arbitrary and unsettling, as it implied a lack of any over-arching purpose to life in general, and human existence in particular.

The scientific study of inheritance, which led to the modern disciplines of genetics and molecular biology, has its origins in the work of Gregor Mendel (discuss above and again in chapter 15). Darwin published multiple revised editions of "On the Origin of Species", so why did he not incorporate a Mendelian view of heredity into his theory? A simple explanation would be that Darwin was unaware of Mendel's work – in fact, the implications of Mendel's work were largely ignored by the scientific community until the early years of the 20th century.

Why was the significance of Mendel's work not immediately recognized? It turns out that Mendel's conclusions were initially not obviously broadly applicable. Mendel studied carefully bred pea plants, *Pisum sativum.* These plants displayed what are known as dichotomous traits (one or the other): smooth versus wrinkled seeds, yellow versus green seeds, grey versus white seed coat, tall versus short plants. In natural population, these traits are not dichotomous but occur along a continuum, with various intermediate forms that (as Weldon noted) can be influenced by genetic, developmental and environmental factors.[269] The majority of traits do not behave in a simple, dichotomous Mendelian manner; most genes influence a number of different traits and the variation in a particular trait is generally influenced by variations in many genes. Finally, in an attempt to established the general validity of his conclusions Mendel went on to examined the behavior of other plants, including hawkweed, *Hieracium*. Unfortunately, hawkweed uses a specialized, asexual reproductive strategy, known as apomixis that does not follow Mendel's rules.[270] This did reassure Mendel or others that his genetic laws were universal or useful. Subsequent work, published in 1900, led to the recognition of the validity of many of Mendel's basic conclusions.[271]

A key, but often over-emphasized conclusion from Mendel's work was that there are stable hereditary "factors" that became known as genes. As noted, each gene can exist in different forms, known as alleles. In some cases specific alleles are associated with specific forms of a trait. For example, in mammals, the ability to digest lactose depends upon the ability to make the enzyme lactase. Lactase is encoded by the *LCT* gene and made when *LCT* is "expressed".[272] In most mammals, expression of *LCT* stops with age. In some human sub-populations, *LCT*expression, and so the ability to digest lactose, persists in adults – a trait known as adult lactose tolerance. Adult lactose tolerance has arisen independently in a number of distinct human populations historically associated with the practice of herding animals. One version of adult lactose tolerance is based on the allele of the *MCM6* gene you carry. The *MCM6* allele that promotes adult lactose tolerance acts to maintain the expression of *LCT* into adulthood. As we proceed, we will consider many of the molecular level details involved in such outcomes. We introduced the terms genes, alleles,

genomes, genotypes and phenotypes in our previous discussion of evolutionary mechanisms; we consider them again and again in greater detail as we proceed.

When a cell divides, its genes must be replicated so that each daughter cell receives a full set of genes, a complete genome. The exact set of alleles a cell inherits determines its genotype. Later it was recognized that sets of genes are linked together in a physical way, but this linkage is dynamic; processes exist that can shuffle the alleles of linked genes. In sexually reproducing organisms, such as the pea plants Mendel studied, and in most multicellular organisms including humans, two copies of each gene are present in each somatic (body) cell. Such cells are said to be diploid. During sexual reproduction, specialized cells, known as gametes, are produced; these cells contain only a single copy of each gene and are referred to as haploid, although monoploid would be a better term. Two such haploid gametes fuse to form a new diploid organism. While gametes can be morphologically identical, in animals and plants gametes are generally quite different in size and shape. The gametes of animals are known as sperm and egg, while in plants they are known as pollen and ovule. An individual sexually reproducing organism generally produces a single type of gamete. The organisms that produce the morphologically larger gametes are known as female and those producing the smaller gametes are known as male. As we discussed earlier (Chapter 4), this difference in size can have evolutionary (selective) implications.

In any particular organism there are thousands of genes and within a population there are often a number of different alleles of each gene.[273] Through sexual reproduction a new organism, generated upon gamete fusion, carries a unique combination of the alleles present in their parents. This increases the genetic variation within the population and enables a population, as opposed to specific individuals, to deal with a range of environmental factors, including pathogens, predators, prey, and competitors. It leaves unresolved, however, exactly how genetic information is replicated, how new alleles arise, and how information is encoded, regulated, and utilized at the molecular, cellular, and organismic levels.

Question to answer

105. Develop a plausible explanation for why adult lactose tolerance is not a universal trait of mammals?

References

[266] The eclipse of Darwin: wikipedia

[267] Evidence for perfection in people, as a species, seems consciously absent.

[268] It is worth reading Evolution in Four Dimensions (reviewed here) which reflects on the factors that influence selection.

[269] Actually more complex that we can address here: see An expanded view of complex traits: from polygenic to omnigenic.

[270] Apomixis in hawkweed: Mendel's experimental nemesis: link & van Dijk & Ellis (2016). The full breadth of Mendel's genetics. *Genetics*, *204*, 1327-1336.

[271] Rediscovery of Mendel's work: link

[272] The Co-evolution of Genes and Culture: link

[273] You can get an idea of the alleles present in the human population by using the gnomAD browser: link

This page titled 7.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

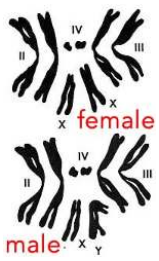
7.1: Discovering how nucleic acids store genetic information

To follow the historical pathway that led to our understanding of how heredity works, we start back at the cell, the basic living unit. As it was established that all organisms are composed of one or more types of cells, and that all cells are derived from pre-existing cells, it became more and more likely that inheritance was a cellular phenomenon. As part of their studies, cytologists (students of cells) began to catalog the common components of cells. Because of resolution limits associated with available microscopes, these studies were initially limited to larger eukaryotic cells. One such visible component of eukaryotic cells is the nucleus. Using early (bright-field) microscopes, cells appear clear and transparent, after all they are ~70% water. To discern structural details cytologists had to stabilize the cell; stabilizing the cell means killing it. Biological samples were "fixed" to ensure that structures were preserved as close to the living state as possible. Originally, this process involved the use of chemicals, such as formaldehyde or organic solvents, that cross-link or precipitated various molecules together. Fixation stops molecular movement; it is not unlike boiling an egg. In more modern studies, using higher resolution optical methods[274] and electron microscopes, such crude fixation methods have been replaced by alternatives that include very rapid freezing and cryoelectron microscopy.

At the level of light microscopy, it can be hard to resolve cellular subcomponents. One approach is to treat fixed cells with various dyes, dyes that bind preferentially to specific types of molecules. These molecules were often found located in particular cellular regions. The most dramatic of these cellular sub-regions is the nucleus, which due to its bulk chemical composition stains differently from the surrounding cytoplasm. One common stain consists of a mixture of hematoxylin and eosin; it leaves the cytoplasm pink and the nucleus dark blue.[275] The nucleus was first described by Robert Brown (1773-1858), the person after which Brownian motion was named. The presence of a nucleus was characteristic of eukaryotic (true nucleus) organisms.[276] Prokaryotic cells (before a nucleus) are typically much smaller which made it technically difficult to determine whether they had a nucleus or not – they do not.

The careful examination of fixed and living eukaryotic cells revealed that the nucleus undergoes a dramatic reorganization during the process of cell division; it loses its roughly spherical shape, which is replaced by discrete stained strands known as chromosomes (colored bodies). In 1887 Edouard van Beneden (1846-1910) reported that the number of chromosomes in a somatic (diploid) cell was constant for each species and that different species had different numbers of chromosomes (←). Within a particular species the individual chromosomes could be recognized based on their distinctive sizes and shapes. For example, in the somatic cells of the fruit fly *Drosophila melanogaster* there are two copies of each of four chromosomes (→). In 1902, Walter Sutton (1877-1916) published his observation that chromosomes obey Mendel's rules of inheritance, that is that during the formation of the gametes that fuse during sexual reproduction, each cell received one and only one copy of each chromosome. This strongly suggested that Mendel's genetic factors were associated with chromosomes.[277] It was recognized that there were many more Mendelian factors than chromosomes, which implied that many factors must be present on each chromosome. These observations provided a physical explanation for the observation that many genetic traits do not behave independently but acted as if they were somehow linked together. The behavior of the nucleus, and the chromosomes that appeared to exist within it, mimicked the type of behavior that a genetic material would be expected to display.

species	chromosome #
Ophioglossum reticulatum (a fern)	1260 (630 pairs)
Canis familiaris (dog)	78 (39 pairs)
Cavia cobaya (guinea pig)	60 (30 pairs)
Solanum tuberosom (potato)	48 (24 pairs)
Homo sapiens (humans)	46 (23 pairs)
Macaca mulatta (monkey)	42 (21 pairs)
Mus musculus (mouse)	40 (20 pairs)
Felis domesticus (house cat)	38 (19 pairs)
Saccharamyces cervisae (yeast)	32 (16 pairs)
Drosophila melanogaster (fruit fly)	8 (4 pairs)
Myrmecia pilosula (ant)	2 (1 pair)



Cellular anatomy studies were followed by studies on the composition of the nucleus. As with many scientific studies, progress is often made when one has the right "model system" to work with. It turns out that some of the best systems for the isolation and analysis of the components of the nucleus were sperm and pus, isolated from discarded bandages from infected wounds (yuck). It was assumed, quite reasonably, that components enriched in this material would likely be enriched in nuclear, genetic information containing components. Using sperm and pus as starting materials Friedrich Miescher (1844-1895) isolated a phosphorus-rich compound called nuclein.[278] Nuclein was later resolved into an acidic component, deoxyribonucleic acid (DNA), and a basic component, primarily proteins known as histones. Because they have different properties (acidic DNA, basic histones), chemical "stains" that bind or react with specific types of molecules and absorb visible light, could be used to visualize the location of these molecules within cells using a light microscope. The nucleus stained for both highly acidic and basic components - which

suggested that both nucleic acids and histones were localized to the nucleus, although what they were doing there was initially unclear.

Questions to answer

97. How was the nucleus first visualized? What was needed to see it?
98. Make (and justify) a prediction: do you expect there to be a correlation between chromosome number of organismic complexity.

Questions to ponder

- In comparing organisms, what does complexity mean?

References

[274] Optical microscopy beyond the diffraction limit: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2645564/

[275] The long history of hematoxylin: http://www.ncbi.nlm.nih.gov/pubmed/16195172

[276] There are some eukaryotic cells, like human red blood cells, that do not have a nucleus, they are unable to divide.

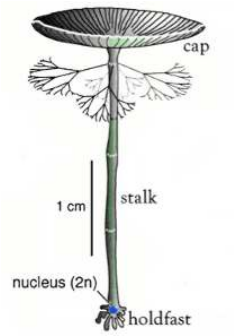
[277] http://www.nature.com/scitable/topicpage/developing-the-chromosome-theory-164

[278] Friedrich Miescher and the discovery of DNA: http://www.sciencedirect.com/science/article/pii/S0012160604008231

This page titled 7.1: Discovering how nucleic acids store genetic information is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.2: Locating hereditary material within the cell

Further evidence suggesting that hereditary information was located in the nucleus emerged from transplantation experiments carried out in the 1930's by Joachim Hammerling (1901-1980). He used the giant unicellular green alga *Acetabularia acetabulum*, known as the mermaid's wineglass (→). Hammerling's experiments (video **link**) illustrate two important themes in biology. The idiosyncrasies of a particular organism can be exploited to carry out studies that are difficult, prohibitively expensive, or simply impossible to perform in other organisms. At the same time, the underlying evolutionary relationships between organisms makes it possible to draw broadly relevant conclusions, something unlikely to be true if each organism represented a unique creation event. That said, there are dangers in thinking that complex human traits (such as altruism, self-consciousness, aberrant (normal, neurotic, psychotic) behaviors, or specific diseases) can be studied in evolutionarily distinct organisms.[279]



Hammerling exploited three unique features of *Acetabularia*. The first is the fact that each individual is a single cell, with a single nucleus. Through microdissection, it is possible to isolate nuclear and anucleate (without a nucleus) regions of the organism. Second, these cells are large (1 to 10 cm in height), which makes it possible to remove and transplant regions of one organism (cell) to another. Finally, different species of *Acetabularia* have mophologically distinct "caps" that regrow faithfully following amputation. In his experiments, he removed the head and stalk regions from one individual, leaving a small "holdfast" region that contained the nucleus. He then transplanted large regions of a anuclear stalk, derived from an individual of a different species with a distinctively different cap morphology, onto the smaller nucleus-containing holdfast region. When the cap regrew it had the morphology characteristic of the species that provided the nucleus - no matter that this region was much smaller than the transplanted, anucleate stalk region. The conclusion was that the information needed to determine the cap's morphology was located within the cell's nuclear region, rather than dispersed throughout the cytoplasm. It was a short step from these experimental results to the conjecture that all genetic information is located within the nucleus.

References

[279] Mice fall short as test subjects - McGlinn 2013 & False analogies & logical fallacies in animal models - Sjoberg 2016

This page titled 7.2: Locating hereditary material within the cell is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

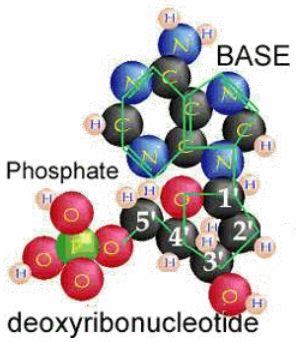
7.3: Identifying DNA as the genetic material

The molecular level mechanisms behind the storage and transmission of genetic information, remained a mystery. Two kinds of experiment led to the realization that genetic information was stored in a chemically stable form. In his studies, H.J. Muller (1890-1967) found that exposing fruit flies to X-rays, a highly energetic form of light, generated genetic changes (mutations) that could be passed from one generation to the next. One conclusion was that genetic information was stored in a chemical form, a form altered through interactions with radiation leading to changes in the molecule(s) storing the information. Moreover, once altered, the genetic information was stable again.

This page titled 7.3: Identifying DNA as the genetic material is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.4: Unraveling Nucleic Acid Structure

Knowing that DNA is the genetic material was a tremendous break through, but it left a mystery - how was genetic information stored and replicated. Nucleic acids were initially thought of as boring aperiodic polymers, that is, molecules built from a defined set of subunits, known as monomers, but without a simple overall repeating pattern. The basic monomeric units of nucleic acids are known as nucleotides (→). A nucleotide consists of three distinct types of molecules joined together, a five-carbon sugar (ribose in RNA or deoxyribose in DNA), a nitrogen-rich “base” that is either a purine, guanine (G) or adenine (A) or a pyrimidine, cytosine (C), or thymine (T) in DNA or uracil (U) instead of T in RNA, and a phosphate group. The carbon atoms of the sugar are numbered 1’ to 5’. The nitrogenous base is attached to the 1' carbon and the phosphate is attached to the 5’ carbon. The other functionally important group is a hydroxyl group attached to the 3’ carbon of the ribose/deoxyribose moiety.[283] RNA differs from DNA in that there is a hydroxyl group attached to the 2’ carbon of the ribose, the absence of this hydroxyl makes DNA a “deoxy” ribonucleic acid! We take particular note of the 5’ phosphate and 3’ hydroxyl groups of the ribose/deoxyribose because they are directly involved in the linkage of nucleotide monomers together to form nucleic acid polymers.



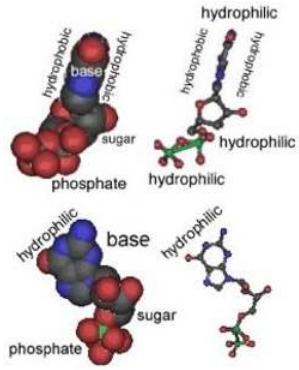
References

[283] “Moiety” defined

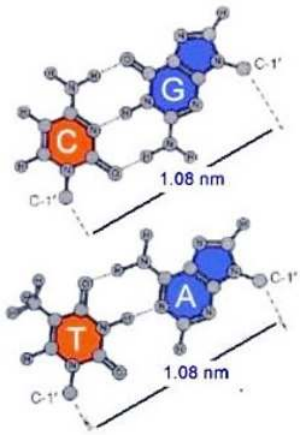
This page titled 7.4: Unraveling Nucleic Acid Structure is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.5: Discovering the structure of DNA

A critical clue to understanding nucleic acid structure came from the work of Erwin Chargaff (1905-2002). When analyzing DNA from various sources, he found that the relative amounts of G, C, T and A nucleotides present varied between organisms but were very similar for organisms of the same type or species. On the other hand, the ratios of A to T and G to C were always equal to 1, no matter where the DNA came from. Knowing these rules, James Watson (1928-) and Francis Crick (1916-2004) built a model of DNA that used structural data from Rosalind Franklin (1920-1958) and fit what was known about the structure of nucleotides. Franklin got her data by pulling DNA molecules into oriented strands; fibers of many molecules aligned parallel to one another. By passing a beam of X-rays through these fibers she obtained a diffraction pattern; a pattern that defined key structural parameters that constrain any model of the molecule's structure.[284] By making a model that was predicted to produce the observed X-ray data, Watson and Crick drew a number of conclusions about the structure of DNA molecules.[285]



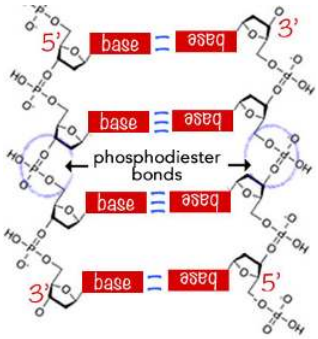
To understand their process, consider the chemical nature of a nucleotide and a nucleotide polymer (a nucleic acid) such as DNA. First the nucleotide bases in DNA (A, G, C and T) have a number of similar properties. Each nucleotide (←) has three hydrophilic regions: the negatively charged phosphate group, a sugar with a number of O–H groups, and the bases' hydrophilic edges, where the N–H and N groups lie. While the phosphate and sugar are three-dimensional moieties, the bases are flat, the atoms in the rings are all in one plane. The upper and lower surfaces of the rings are hydrophobic (non-polar) while the edge is hydrophilic. The same amphipathic factors that favor the assembly of lipids into bilayer membranes also influence nucleic acid structure. In their model Watson and Crick had the bases stacked on top of one another, hydrophobic surface next to hydrophobic surface, to reduce their interactions with water.



Each base's hydrophilic edge, with –C=O and –N-H groups can act as H-bond acceptors and donors. How were these hydrophilic groups arranged? With the two polynucleotide strands arranged in opposite orientations, that is, anti-parallel to one another: one from 5' → 3' and the other 3' ← 5'; the bases attached to the sugar-phosphate backbone interact with one another in a highly specific way (←). An A can form two hydrogen bonding interactions with a T on the opposite, anti-parallel strand, while a G can form three hydrogen bonding interactions with a C. A key feature of this arrangement is that the lengths of the A::T and G:::C base pairs are almost identical. The hydrophobic surfaces of the bases are stacked on top of each other, while the hydrophilic sugar and phosphate groups are in contact with the surrounding aqueous solution. The repulsion between negatively charged phosphate groups is neutralized (or shielded) by the presence of positively charged ions present in the solution from which the X-ray measurements were made (and in the cell). This model also provided a direct explanation for why Chargaff's rules are universal in double stranded DNA.

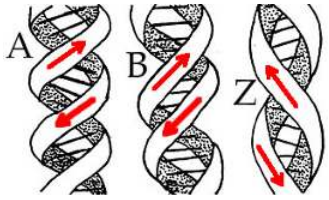
Each DNA polymer strand has a directionality to it, it runs from the 5' phosphate group of the ribose/deoxyribose at one end to the 3' hydroxyl group of the ribose/deoxyribose at the other end. Each nucleotide monomer is connected to the next through a phosphodiester linkage (→) involving its 5' phosphate group attached to the 3' hydroxyl of the existing strand. In their final model Watson and Crick depicted what is now known as B-form DNA. This is the usual structural form of DNA in a cell. Under different salt conditions DNA can form two other double helical forms, known as A and Z forms. A and B forms of DNA are "right-handed" helices, while the Z-form is a left-handed helix (←). We will not concern ourselves with these other forms of DNA, leaving that to more advanced courses, but you can imagine that they might well influence the types of intermolecular interactions that occur between DNA and other molecules, particularly proteins.

As soon as the Watson-Crick model of DNA structure was proposed its explanatory power was obvious. Because the A::T and G:::C base pairs are of the same length, the sequence of bases along the length of a DNA molecule, written, by convention in the 5' to 3' direction, has little effect on the overall three-dimensional structure of the molecule. That implies that essentially any



sequence can be found, at least theoretically, in a DNA molecule. If information were encoded in the sequence of nucleotides along a DNA strand that information would be as stable as the DNA molecule itself. This is similar to the storage of information in various computer memory devices, that is, any type of information can be stored, because storage does not involve a dramatic change in the structure of the storage material. The structure of a flash memory drive is not dramatically different whether in contains photos of your friends, a song, a video, or a textbook. What matters is how the information is "encoded", most obviously in the specific sequence of nucleotides along a DNA strand.

At the same time, the double-stranded structure of DNA and the complementary nature of base pairing (A to T and G to C) suggested a simple model for information storage and replication. Pull the two strands of the molecule apart and build two new (anti-parallel) strands using the two original strands as templates. This model of DNA replication is facilitated by the fact that the two strands of the parental DNA molecule are held together by weak hydrogen bonding interactions; no covalent bonds need be broken to separate the strands from one another. In fact, at physiological



temperatures DNA molecules often open up over short stretches and then close again, a process known as DNA breathing.[286] This makes the replication of the information stored in the molecule conceptually straightforward, even though the actual biochemical process is complex, in part because of the importance of accurate replication. The existing strands determine the sequence of nucleotides on the newly synthesized strands. The newly synthesized strand can, in turn, direct the synthesis of a second strand, identical to the original strand. Finally, the double-stranded nature of DNA molecule means that any information within the molecule is stored in a redundant fashion. If one strand is damaged, that is its DNA sequence is lost or altered, the second undamaged strand can be used to repair the damage.

Questions to answer

103. How is a DNA molecule structurally analogous to a lipid bilayer? Draw a diagram that reveals the similarities and note the most important differences.
104. Which do you think is stronger (and why), an AT or a GC base pair?
105. Why is the ratio of A to T the same in all organisms?
106. Normally DNA exists inside of cells at physiological salt concentration (~140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$ and some minor ions). What happens if you placed DNA into pure water?
107. How many general types of mutation can you think of? How would they differ in their impact on the information encoded in a DNA molecule.
108. Generate a model mechanism by which a DNA molecule could be accurately repaired, that is, without the loss of the information originally present within it.

Questions to ponder

- You isolated DNA from an organism, and you find it fails to obey Chargaff's rule; what might you predict about the structure of its DNA?

References

[284] Fiber diffraction

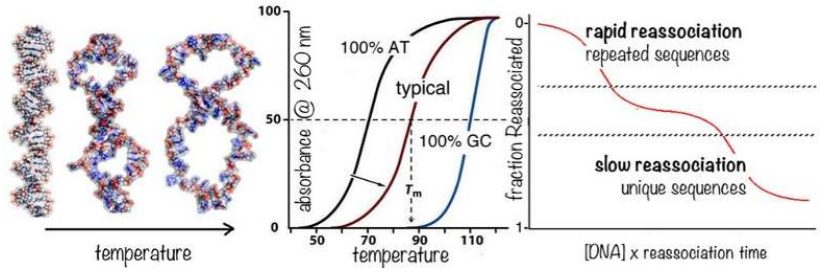
[285] An interesting depiction of this process is provided by the movie "Life Story"

[286] Dynamic approach to DNA breathing: http://www.ncbi.nlm.nih.gov/pubmed/23345902

This page titled 7.5: Discovering the structure of DNA is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.6: DNA, sequences, and information

Let us consider the types of information stored in DNA. Early students of DNA could not read DNA sequences as we can now, so they relied on indirect measurements to better understand the behavior of DNA molecules. For example, double and single stranded DNA molecules interact with light differently. The two strands of a double-stranded DNA molecule (dsDNA) are linked by hydrogen bonds; as temperature increases the two strands will separate into two single stranded molecules (ssDNA)(left panel →).



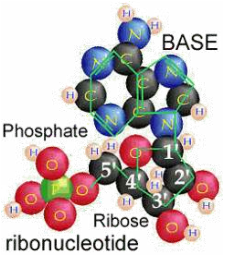
ssDNA absorbs light at 260nm (in the ultraviolet range) more strongly than does dsDNA; this makes it possible to use the absorbance of a DNA solution to determine the relative amounts of single and double stranded DNA in a sample. Such studies reveal that the temperature at which 50% of dsDNA molecules have separated into ssDNA molecules varies between organisms. This is not surprising given Chargaff's observation that the ratio of AT to GC varies between organisms and that GC base pairs involve three H-bonds and so are more stable (take more energy to break) than AT base pairs, held together by two H-bonds. One can estimate the AT:GC ratio of a DNA molecule based on dsDNA to ssDNA melting curves (middle ↑).

It quickly became clear that things were more complex than expected. Here a technical point needs to be introduced. Because of the extreme length of the dsDNA molecules found in biological systems, it is almost impossible to isolate them intact. In the course of purification, the molecules are sheared (broken) into shorter pieces, typically thousands to tens of thousands of base pairs in length compared to the millions to hundreds of millions of base pairs in intact molecules. In another type of experiment, one can look at how fast ssDNAs (the result of a melting experiment) reform dsDNA. The speed of these "reannealing reactions" depends on DNA concentration. When such experiments were carried out, it was found that there was a fast annealing population of DNA fragments and slower annealing populations (right panel ↑). How to explain this observation? Was it a function of AT:GC ratio or was something else going on? Subsequent analyses revealed that it was due to the fact that within the genomic DNA of many organisms, particularly eukaryotes, there were many (hundreds to thousands) of molecular regions that contained very similar nucleotide sequences. Because the single strands of these fragments can associate with one another, these sequences occur in much higher effective concentrations compared to regions of the DNA with unique sequences. This type of analysis revealed that much of the genome of eukaryotes is composed of families of repeated sequences and that regions of unique sequence amount to less than ~5% of the total genomic DNA. While a complete discussion of these repeated sequence elements is beyond us here, we can make a few points. As we will see, there are mechanisms that can move regions of a DNA molecule from one position to another within the genome, and mechanisms that can generate a copy of a DNA sequence and insert it into another position of the genome (leaving the original sequence behind). The end result is that the genome of a cell/organism is dynamic, a fact with profound evolutionary implications.

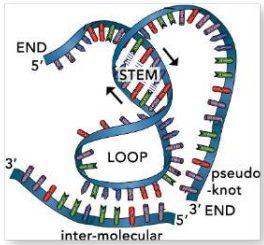
This page titled 7.6: DNA, sequences, and information is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

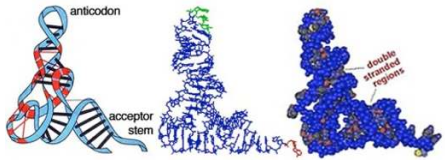
7.7: Discovering RNA- structure and some functions

DNA is not the only nucleic acid found in cells. A second class of biological nucleic acids are known as ribonucleic acids or RNAs. As noted above, RNA differs from DNA in that it contains i) the sugar ribose (with a hydroxyl group on the 2' C) rather than deoxyribose; ii) the pyrimidine uracil rather than the pyrimidine thymine found in DNA (→); and iii) RNA is typically single rather than double stranded.[287] Nevertheless, RNA molecules can associate with an ssDNA molecule with a complementary nucleotide sequence. Instead of the A::T pairing in DNA A pairs with U instead. This change does not make any significant difference when the RNA strand interacts with DNA, since the number of hydrogen bonding interactions are the same.



When RNA is isolated from cells, the major population was found to re-associate with unique sequences within the DNA. This class of RNA includes molecules, known as messenger or mRNAs, that carry information from DNA to the molecular machinery that mediates the synthesis of polypeptides, the ribosome. In addition to mRNAs there are a number of other types of RNAs in cells; in each case, their synthesis is directed by DNA-dependent RNA polymerases. These non-mRNAs include structural, catalytic, and regulatory RNAs - know generically as non-coding or ncRNAs. Some RNAs have roles regulatory and coding functions (considered in more detail later). As you may already suspect, the same hydrophobic/hydrophilic/H-bond considerations that were relevant to DNA structure apply to RNA structures, but because RNA is generally single stranded, the structures found in RNA are different and more varied. A single-stranded RNA molecule can fold back on itself, through intra-strand interactions to create local double stranded regions (→). Similarly distinct RNA molecules can interact through double-stranded regions (inter-molecular interactions). In both cases, and just as in DNA, these strands are anti-parallel to one another. This results in double-stranded regions ("stems") that end in single-stranded "loops" (or molecular ends). Regions within a stem, that can be as short as 1 base pair, that do not base pair will "bulge out". The end result is that RNA molecules can adopt a wide range of complex three-dimensional structures in solution. Transfer RNAs (tRNAs)(←), an integral component of the protein synthesis system, are one well studied example of how intermolecular interactions within an RNA molecule can produce complex three-dimensional shapes that carry out specific molecular functions (more in the next chapter).





In addition to intra- and inter-molecular interactions involving RNA molecules, RNAs can also interact with proteins to form "riboprotein" complexes. For example, the CRISPR-Cas9 system involves a double-stranded DNA endonuclease, an enzyme that generates the cleavage of both strands of a double-stranded DNA molecule, and an RNA molecule, known as a guide RNA, that direct the enzyme to the DNA sequences to be cleaved. Other RNA-protein complexes are involved in the control of RNA synthesis and stability, among a number of other functions. The classic example of a riboprotein complex is the ribosome itself, a macromolecular machine that mediates the synthesis of polypeptides. A ribosome is composed of structural and catalytic RNAs (known as ribosomal or rRNAs) and ~50 to 80 proteins (polypeptides), depending upon whether you are prokaryotic or eukaryotic; altogether it has a molecular weight of ~3.2 x 10^6 daltons.

The ability of RNA to both encode information in its base sequence and to mediate catalysis through its three dimensional structure has led to the "RNA world" hypothesis that proposes that early on various proto- (pre-LUCA) organisms relied on RNAs, or perhaps simpler RNA-like molecules, rather than DNA and proteins to store genetic information and to catalyze metabolic reactions. Some modern day viruses use single or double-stranded RNAs as their genetic material. According to the RNA world hypothesis, it was only later in the history of life that organisms developed more specialized DNA-based systems for genetic information storage and proteins for most catalytic and structural functions. While this idea is compelling, there is no reason to believe that simple polypeptides and other molecules were not also present and involved in the early stages of life's origins. At the same time, there are many unsolved issues associated with a simplistic RNA world view, the most important being the complexity of RNA itself, its abiogenic (that is, without life) synthesis, and the survival of nucleotide triphosphates in solution. What is clear is that catalytic and regulatory RNAs play key roles in modern cells and their evolution.

Questions to answer:

109. How would you calculate the probability that two DNA sequences (of length N) are identical by chance?
110. Predict how the annealing curve of genomic DNA changes as the number of repeated sequences increases.
111. Propose a plausible model for how a single-stranded RNA molecule could act as a catalyst; consider why double-stranded DNA is unlikely to act catalytically. (required thinking about what catalysts do to speed reactions).

Question to ponder:

- Where might the repeated sequences of DNA in a genome have come from?

References

[287] The exception involves viruses, where double stranded RNA is found as the genetic material

---

This page titled 7.7: Discovering RNA- structure and some functions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.8: DNA Replication

Once proposed, the double-helical structure of DNA immediately suggested a simple mechanism for the accurate replication of the information stored in DNA. Each strand contains all of the information necessary to specify the sequence of the complementary (anti-parallel) strand. The process begins when a dsDNA molecule opens to produce two single-stranded regions. Where DNA is naked, that is, not associated with other molecules (proteins), the opening of the two strands can occur easily, since they are held together only by weak H-bonding interactions. Normally, the single strands simply re-associate with one another. To replicate DNA the open region has to be stabilized and the catalytic machinery involved recruited and activated. We will consider how this is done in general terms, in practice this is a complex and highly regulated process involving a number of molecular components.

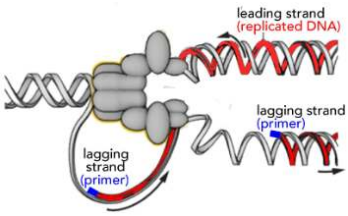
The first two issues we have to address may seem arbitrary, but they turn out to be common (conserved) features of DNA synthesis. The enzymes (DNA-dependent, DNA polymerases) that catalyze the synthesis of new DNA strands cannot start the synthesis of a new polynucleotide strand on their own or *de novo*: they must add nucleotides onto the end of a pre-existing nucleic acid polymer, they depend on an "RNA primer". The primer is synthesized by a DNA-dependent, RNA polymerases using a mechanistically similar process. RNA Primer synthesis is based on a complementary DNA sequence. Both DNA and RNA polymerases link the 5' (phosphate) end of a nucleotide triphosphate molecule to the pre-existing 3' (hydroxyl) end of a nucleic acid polymer; the polymerization reaction is said to proceed in the 5' to 3' direction; nucleotides are added to the 3' end (↓). The molecules involved in DNA replication and RNA synthesis rely on signals within the DNA that are recognized by specific proteins; together these determine where and when nucleic acid replication (synthesis) begins and ends.

After the dsDNA molecule has "opened" locally primase, a specialized DNA-dependent, RNA polymerase, binds to and synthesizes a short RNA primer. Because the two strands of the DNA molecule are anti-parallel one primase complex associates with each of the two now separated DNA strands. The result is one RNA primer on each strand. DNA-dependent, DNA polymerases then replace primase and begins to catalyze the deoxynucleotide-addition reaction. Which nucleotides are added is based on which nucleotides are present in the existing DNA strand.

Nucleotides exist in various forms within the cell, including nucleotide monophosphate (NMP), diphosphate (NDP), and triphosphate (NTP) forms. To make the nucleic acid polymerization reaction thermodynamically favorable, the reaction uses the NTP form of the nucleotide monomers, generated through the reaction:

$$5'\text{P-NMP-OH3}'_{\text{existing n-length polymer}} + 5'\text{P-NTP-OH3}'_{\text{incoming}} + \text{H}_2\text{0} \leftrightarrow 5'\ \text{P-NTP-NMP-OH3}'_{\text{n-length polymer}} + \text{diphosphate.}$$

During this reaction the terminal diphosphate of the incoming NTP is released (a thermodynamically favorable reaction) and a nucleotide mono-phosphate is added to the existing polymer through the formation of a phosphodiester [-C-O-P-O-C-] bond. This reaction creates a new 3' OH polymer end that can react with another NTP. In theory, this process could continue until the newly synthesized strand reaches the end of the DNA molecule. The strand synthesized from the original primer is known as the "leading" strand. For the process to continue, however, the double stranded region of the original DNA will have to open up further, exposing more single-stranded DNA. Keep in mind that this process is moving, through two independent complexes, in both directions along a DNA molecule. Because the polymerization reaction only proceeds by 3' addition, as new single stranded regions are opened (→) new primers must be created by RNA primase and then extended by DNA polymerase; these are known as the lagging strands. While there are two leading strands leaving a particular DNA replication start site, there are a number of lagging strands involved.



If you try drawing what this looks like, you will realize that i) this process is asymmetric in relation to the start site of replication; ii) the process generates RNA-DNA hybrid molecules; and iii) that eventually an extending DNA polymerase will run into the RNA primer part of an "upstream" molecule. But RNA regions, derived from the primers, are not found in "mature" DNA molecules, so there must be a mechanism that removes them. As it turns out, the DNA polymerase complex, like a number of enzyme systems, contains more than one catalytic activity (analogous to the ATP synthase and pump). When the DNA polymerase complex reaches the upstream nucleic acid chain it runs into an RNA primer region. An RNA exonuclease activity associated with the DNA polymerase complex removes the RNA nucleotides and replaces them with DNA nucleotides. Once the RNA region is

removed, a DNA ligase acts to link (generate a covalent phosphodiester bond between) the two DNA molecules. These reactions, driven by nucleotide hydrolysis, end up producing a continuous DNA strand that runs from one end of the chromosome to the other, or in circular chromosomes (found in prokaryotes) all the way around the circle.

Evolution considerations

At this point you might well ask yourself, why (for heavens sake) is the process of DNA replication so complex? Why not use a DNA polymerase that does not need an RNA primer, or any primer for that matter? That should be possible, particularly once you know that RNA polymerase does not need a primer. Why not have polymerases that can add nucleotides equally well to either end of a polymer? That such a mechanism is possible is suggested by the presence of enzymes in eukaryotic cells that can catalyze the addition of a nucleotide to the 5' end of an RNA molecule. The 5' capping reaction is associated with mRNA synthesis (considered later). But while apparently possible, such activities are not known to be used in DNA replication. The real answer to why DNA replication is as complex as it is an we are not completely sure why. It could be its complexity is an evolutionary relic, based on a process established within the last common ancestor of all organisms and difficult or impossible to change through evolutionary mechanisms, or not worth the effort, in terms of its effects on reproductive success. Alternatively, there could be strong selective advantages associated with the system that preclude such changes. What is clear is that this is how the system appears to function in all known organisms. For practical purposes, we need to remember a few key details, these include the direction of polymer synthesis (3' addition) and the need for an RNA primer in DNA synthesis.

This page titled 7.8: DNA Replication is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.9: Replication machines

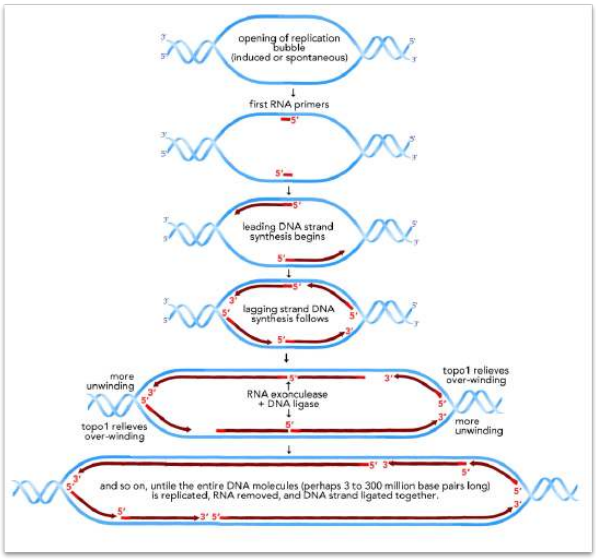
We have presented DNA replication in mechanistically simple terms, but it is worth remembering that the machinery involved is complex and still not completely understood. Complexity arises because the process is topologically constrained and needs to be highly accurate. In the bacterium *Escherichia coli* over 100 genes are involved in DNA replication and repair. To ensure that replication is controlled and complete, replication begins at specific sequences along the DNA strand, known as origins of replication or origins for short. Origin DNA sequences are recognized by sequence-specific DNA binding proteins. The binding of these proteins initiates the assembly of an origin recognition complex, an ORC. Various proteins then bind to the DNA to locally denature (unwind and separate) the two DNA strands. This leads to the formation of what is known as a replication bubble. Multiprotein complexes, known as replication forks assemble on the two DNA strands. Using a single replication origin and two replication forks, moving in opposite directions, a rapidly growing *E. coli* cell can replicate its ~4,700,000 base pairs of DNA, present in the form of a circular DNA molecule, in ~40 minutes. Each replication fork moves along the DNA adding ~1000 base pairs of DNA per second to the newly formed DNA polymer. While a discussion of the exact mechanisms involved is beyond us here, it is critical that DNA replication is complete before a cell attempts to divide - this implies that there are cellular signaling systems that monitor and coordinate the completion of DNA replication before the start of cell division. We will find such "checkpoint" systems in a number of cellular processes. In many bacteria, the signaling system is based on the fact that the chromosome is circular, that DNA replication begins at a single site (the origin), and that when replication forks collide with one another and complete replication in the "terminus" region of the chromosome signals are generated.[288]

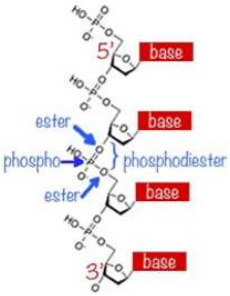
Questions to answer

112. Draw a diagram of the key steps in the replication of a circular DNA molecule. How might you adapt this system to replicate much longer linear molecules?
113. What key, non-deducible features of DNA replication do you need to remember (memorize) and why?

Accuracy and error in DNA synthesis

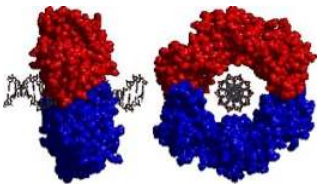
DNA replication is a highly accurate process; the DNA-dependent DNA polymerase makes about one error for every ~10,000 bases it adds. But that level of error is not low enough, and would be highly deleterious; most of these errors are recognized as mistakes. To understand how, remember that correct AT and GC base pairs have the same molecular dimensions, that means that incorrect AG, CT, AC, and GT base pairs are either too long or too short. By responding to base pair length, molecular machines can recognize errors in base pairing as mistakes leading to abnormal structures. When a mismatched base pair is formed and the mistake recognized, the DNA polymerase pauses forward synthesis, reverses direction, removes the mismatched base pair using an exonuclease activity. It then resynthesizes the region, hopefully correctly. This process is known as proof-reading; the proof-reading activity of the DNA polymerase complex reduces the total DNA synthesis error rate from ~1 in 10^4 base pairs to ~1 error per 1,000,000,000 (10^9) synthesized.[289]





Let us consider nomenclature that can seem arcane and difficult to understand, but obeys reasonably straightforward rules. An exonuclease is an enzyme that can bind to the free end of a nucleic acid polymer and remove nucleotides through a hydrolysis reaction of the phosphodiester bond (→). A 5' exonuclease cuts off a nucleotide located at the 5' end of the molecule, a 3' exonuclease cuts off a nucleotide located at the molecule's 3' end. An intact circular nucleic acid molecule is immune to the effects of an exonuclease. To break the bond between two nucleotides in the interior of a nucleic acid molecule (or in a circular molecule, which has no ends), one needs an endonuclease.

Thinking about DNA replication, you may realize that once DNA synthesis begins, it is important that it continues without interruption until completion. The interactions between nucleic acid chains are based on weak H-bonding interactions, and the enzymes involved in DNA replication could dissociate from the DNA in response to the effects of thermal motion. We can characterize how long a DNA polymerase molecule remains associated with a DNA molecule in terms of the number of nucleotides it adds before it falls off; this is known as its "processivity". Can you think of factors that would influence a polymerase's processivity? One approach to increaseprocessivity is a molecular clamp. The DNA polymerase is held onto the DNA through its strong attachment to a doughnut shaped "sliding clamp" protein complex (←) that encircles and slides along the DNA double helix (**video link**)[290]. The clamp protein is loaded onto DNA by another molecular machine, the clamp loader. Once closed around the DNA the clamp can move along the double-stranded length of the DNA molecule, driven by collisions with other molecules. The clamp is attached to the DNA polymerase complex that is adding monomers to the growing nucleic acid polymer. This moves the replication complex along the DNA in the direction of synthesis. To leave the DNA clamp has to open and the polymerase has to disengage from the clamp.



References

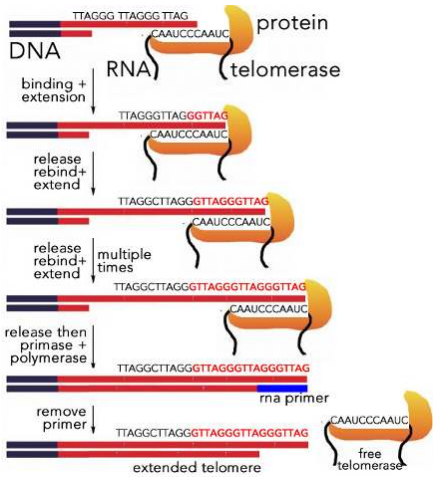
[288] Synchronization of Chromosome Dynamics and Cell Division in Bacteria

[289] Because of polymerase pausing at a mistake, proof-reading activity can be selected for, since it increases total cell division speed. see Ravasio et al, 2024. A minimal scenario for the origin of non-equilibrium order. *arXiv preprint arXiv:2405.10911*.

[290] see Clamp loader ATPases and the evolution of DNA replication machinery & DNA Clamp & Clamp Loader video

This page titled 7.9: Replication machines is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.10: Further replication complexities

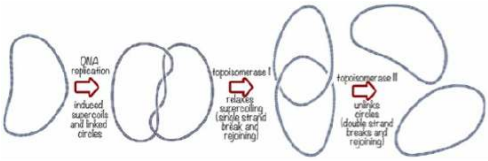


The DNA molecules found in bacteria and archaea are circular; they have no free ends.[291] Eukaryotic cells can contain more than 1000 times the DNA found in a typical bacterial cell. Instead of circles, they contain multiple linear molecules that form the structural basis of their chromosomes (more soon). The free ends of chromosomes are known as telomeres. The linearity of eukaryotic chromosomes creates problems replicating DNA ends. Left alone, more and more of the lagging strand end of the chromosome would go unreplicated and, after removal of the primer, would be single-stranded and unstable. The result? the end of the chromosome would begin to disappear with each DNA replication cycle. To address this “design limitation” eukaryotes use another RNA-protein complex, telomerase.[292]

The ends for linear DNA molecules are known as telomeres. In humans, they are composed of repeated copies of the sequence TTAGGG-3’ (→). Telomerase is composed of a polypeptide, encoded by the TERT gene, and an RNA, encoded by the TERC gene.[293] The TERC RNA does not encode a polypeptide; it contains a sequence complementary to the telomere DNA sequence and serves as the template for the synthesis of GGTTAG from the 3’ end of the telomere’s lagging strand - this process can occur multiple times, after which primase and DNA-dependent, DNA polymerase can fill in the telomere end. For more details, look to an advanced molecular biology course or follow the footnote for a further discussion of telomeres and telomerase.[294]

Topoisomerases

The circular nature of prokaryotic chromosomes creates its own issues based on molecular topology. After replication, the two double-stranded DNA circles are linked together. In eukaryotic cells extremely long linear DNA molecules can become entangled and the unwinding of DNA during replication leads to DNA supercoiling. Left unresolved, supercoiling will inhibit DNA strand DNA synthesis, perhaps you can suggest why.[295] These topological issues are resolved by enzymes known as topoisomerases; they can interconvert topologically distinct versions of a molecule. There are two generic types of DNA topoisomerases: type I topoisomerases (→) bind to the DNA, catalyze the breaking of a single bond in one sugar-phosphate-sugar backbone; the molecule "relaxes" with release of overwinding through rotation the tension is released around the bonds in the intact chain. The enzyme then catalyzes the reformation of the broken bond. Both bond breaking and formation are coupled to ATP hydrolysis. Type II topoisomerases (↓) are involved in “unknotting” and untangling DNA molecules. These enzymes bind to the DNA, catalyze the hydrolysis of both backbone chains, but hold on to the now free ends. This allows another double-stranded molecule to “pass through” the broken strand. The enzyme also catalyzes the reverse reaction, reforming the bonds originally broken.



Replication fork collisions

In addition to having typically much more DNA, the eukaryotic DNA replication enzyme complex is slower, about 1/20th as fast as the prokaryotic system. A bacterial cell can replicate its circular ~3 x 10^6 base pair chromosome in ~1500 seconds using a single replication origin. Eukaryotes use many origins of replication, scattered along the length of each chromosome. So what happens when replication forks collide with one another? In the case of a circular DNA molecule, with its single origin of replication, the replication forks resolve in a specific DNA region known as the terminator. At this point type II topoisomerases allow the two circular DNA molecules to disengage from one another, the cell division machinery forms between the two DNA molecules and they move to opposite ends of the cell. The system in eukaryotes, with their multiple linear chromosomes, is much more complex, although topoisomerases are still involved in separating replicated chromosomes (considered in greater detail in chapter 12).

Questions to answer

114. During DNA/RNA synthesis what is the average ratio of productive to unproductive interactions between an incoming nucleotide and the polymerase?

115. Why do you need to denature (melt) the DNA double-helix to copy it? How would DNA replication change if H-bonds were as strong as covalent bonds?
116. Explain how DNA polymerase might recognize a mistake associated with a mismatched base pair.

Questions to ponder:

- How would evolution be impacted if DNA were totally stable and DNA replication was error-free?

- What would be the effect if a mutation inactivated the the proof-reading function of the DNA polymerase complex?

- How might mutations in the genes encoding the clamp/clamp-loader system influence DNA replication?

References

[291] The mitochondria and chloroplasts of eukaryotic cells also contain circular DNA molecules, another homology with their ancestral bacterial parents. ,

[292] http://en.wikipedia.org/wiki/Telomerase

[293] You can explore the known genetic diseases by using the On-line Mendelian Inheritance in Man (OMIM) database:

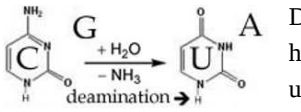
[294] more on telomerase: http://blogs.scientificamerican.com/...as-too-little/

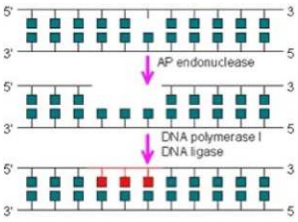
[295] see this video on DNA supercoiling and topoisomerases: http://youtu.be/EYGrElVyHnU

This page titled 7.10: Further replication complexities is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

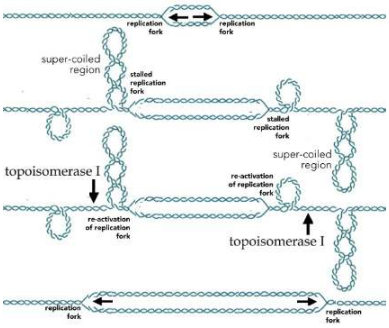
7.11: Mutations, Deletions, Duplications and Repair

While DNA is the universal genetic material of organisms, DNA is a thermodynamically unstable molecule. Eventually it will breakdowns into more stable and simpler components. In water at a temperature of ~13ºC, half of the phosphodiester bonds in a DNA sample will break after ~520 years.[296] Cytosine groups can react with water, which is present at high concentration (~54M) inside a cell. The resulting deamination reaction transforms cytosine into uracil (→). If left unrepaired the original CG base pair is replaced by an AU base pair in one strand during DNA synthesis. But, uracil is not normally found in DNA and its presence can be recognized by an enzyme that breaks the bond between the uracil moiety and the deoxyribose group.[297] The absence of a base, due either to its spontaneous loss or its enzymatic removal, acts as a signal for another enzyme system, the Base Excision Repair complex (←) that removes the section of the DNA strand containing the missing base.[298] A DNA-dependent DNA polymerase will then bind to the open DNA and, using the existing strand as a primer and the undamaged strand as a template, fill in the gap. Finally, another enzyme (a DNA ligase) will join the newly synthesized segment to the pre-existing strand. In the human genome there are over 130 genes devoted to repairing damaged DNA, which indicates the importance of DNA mutation repair.[299]



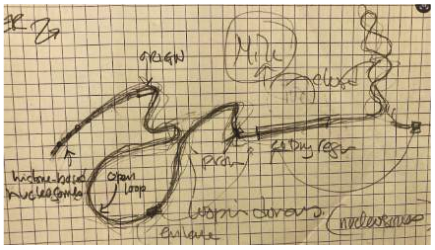


Other hydrolysis reactions that impact nucleic acid integrity include depurination: the loss of a cytosine or thymine group and depyrimidination: the loss of an adenine or guanine group. The rates of these reactions increases at acidic pH, a possible reason that cytoplasm is not acidic. How frequent are such events? A human body contains ~10^{14} cells. Each cell contains about ~10^{9} base pairs of DNA. Each cell, whether it is dividing or not, undergoes ~10,000 base loss events per day or ~10^{18} events per day per person. That's a lot! The basic instability of DNA and the lack of repair after an organism dies means that DNA from dinosaurs, the last of which went extinct ~65,000,000 years ago, has disappeared from the earth, making it impossible to accurately clone (or resurrect) a true dinosaur.[300] In addition DNA can be damaged by environmental factors, such as radiation, ingested chemicals, and reactive compounds made by the cell itself. Many of the most potent mutagens known are natural products, often produced by organisms to defend themselves against being eaten or infected by parasites, predators, or pathogens.[301]

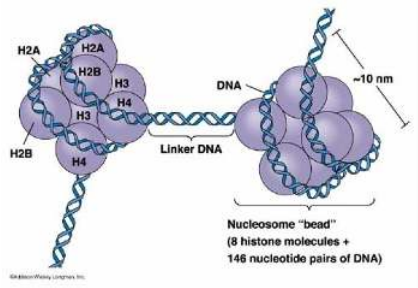


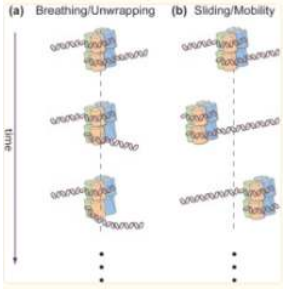
A step back before going forward: what, exactly, is a gene anyway?

We have introduced you to genes multiple times in various contexts. Often people equate DNA with genes, but reality (as usual) is rather more complex. First remember, information in DNA makessense only in the context of a living cell; the living cell includes the regulatory and structural factors, and other genes, involved in "sense making". Within the genome are DNA sequences that encode functions (→), such as acting as an origin of replication or as a telomere. Prokaryotic (circular) chromosomes have a terminator (TER) region, where DNA replication forks collide and resolve into two distinct circular chromosomes. Eukaryotic chromosomes are folded in various ways to fit into the nucleus. Some regions become inaccessible to regulatory molecules, while others are accessible. Different levels of folding are mediated by DNA sequences and regulatory proteins. There are even regions of DNA, known as mobile genetic elements or transposons that can jump from place to place in the genome. They contain the sequences encoding proteins that can copy, and insert the transposon elsewhere in the genome using recognition sequences in the transposon. Mutations can inactivate transposons. The human genome contains ~45% transposon-related (active and inactive) sequences. We will not discuss them further here, but they have been implicated in rapid evolutionary events and a range of processes.[302]



In eukaryotes, genomic DNA is typically wound around a protein complex known as a nucleosome. Prokaryotic DNA interacts what are known as nucleoid-associated proteins (NAPs).[303] Each nucleosome is composed of two sets of four (positively charged at nuclear pH) polypeptides known as histones. About 147 base pairs of DNA wrap around the nucleosome core (←). Variable lengths of "linker" DNA separate adjacent nucleosomes. The DNA-nucleosome complex can become looser or tighter; nucleosomes can "slide" along the DNA. These motions are driven by thermal collisions with solvent molecules. Non-histone proteins associate with





the DNA-histone complex to form chromatin and mediate the interactions with regulatory and structural factors. The spacing between nucleosomes can be influenced by other factors, including modifications of histone proteins, DNA modifications, such as methylation, and various regulatory molecules.

In the jargon of molecular biology, genes that direct RNA synthesis are said to be "expressed"; DNA-directed RNA synthesis is known as "transcription". RNAs that encode polypeptides are known as messenger or mRNAs. mRNAs interact with nuclear factors to be transported out of the nucleus and with cytoplasmic ribosomes to direct the synthesis of the encoded polypeptide, a process known as "translation" (next chapter). When synthesized, an mRNA contains the "coding region" that directs polypeptide synthesis and what are known as 5' and 3' untranslated regions. Untranslated regions, UTRs, are involved in mediating the interactions with ribosomes. mRNA stability is determined by interactions with RNA modifying and degrading enzymes. Unlike radioactive isotopes, an RNA's stability, its "half-life", can be regulated and a particular mRNA can have different half-lives in different cell types or under different environmental conditions.

Some gene-directed transcribed RNAs are "non-coding"; these can have a number of regulatory, structural, and catalytic functions. Non-coding RNAs include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and various small RNAs found associated with proteins. There are also long non-coding RNAs referred to as lncRNAs. An emerging complication (common in biology) is that some RNAs, originally thought to be non-coding, encode small polypeptides. These micropeptides are often involved in the regulation of various aspects of gene expression and function.[304]

Whether coding or not, each gene includes DNA sequences that, together with sequence-specific DNA binding proteins, regulate gene expression; when, where, and, what RNAs are synthesized. These DNA regulatory sequences come in two general "types": promoters and enhancers. Promoter and enhancer regions often act together to control the recruitment and activation of DNA-dependent RNA polymerases to specific sites on a DNA molecule. Promoter regions are generally upstream of the coding region. Enhancer regions can be located upstream or downstream of the gene(s) they regulate. Surprisingly, they can be up to a million base pairs away.[305] Defining all of the regulatory regions of a gene can be challenging, particularly since different promoters and enhancers may be used at different times and in the different cell types present within a single organism.

Transcribed domains can be complex, particularly in eukaryotic genes: a single gene can produce multiple, functionally distinct gene products through the use of alternative promotors and RNA splicing (coming up).[306] How differences in gene sequence influence the activity and role(s) of a gene is not simple. A critical point to keep in mind is that a gene has meaning only in cellular/organism context. Change the organism and the same, or rather, more accurately, homologous genes (genes present in a common ancestor) can have different functional roles. DNA is double stranded, so genes can be located on both strands and can be overlapping. We will return to our (necessarily) simplified consideration of gene regulation; it is often the focus of its own upper division course.[307]

References

[296] Here is the paper from which statement is derived: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

[297] UNG: uracil-DNA-N-glycosidase http://omim.org/entry/191525

[298] absent purine/absent pyrimidine endonuclease http://omim.org/entry/300773

[299] Human DNA Repair Genes – video with lots of misspelled words here: http://youtu.be/g4khROaOO6c

[300] DNA has a 521-year half-life: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

[301] Dietary carcinogens, environmental pollution, and cancer: some misconception

[302] The impact of transposable elements in adaptive evolution:

[303] Bacterial nucleoid-associated proteins, nucleoid structure and gene expression

[304] Muller's morphs: https://en.wikipedia.org/wiki/Muller's_morphs

[305] Enhancers: bridging the gap between gene control and human disease (link to pdf)

[306] Expansion of the eukaryotic proteome by alternative splicing see also Genes – way weirder than you thought

[307] Alternative ORFs and small ORFs: shedding light on the dark proteome & The mystery of the human genome's dark matter

---

This page titled 7.11: Mutations, Deletions, Duplications and Repair is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.12: Genes and alleles

A gene corresponds to specific sequences of DNA. Different versions of a gene, known as alleles, differ in their DNA sequences. Two alleles of the same gene can differ from one another by as little as one nucleotide pair out of thousands or at multiple positions. Differences between alleles can include sequence changes, deletions, duplications, and insertions. A complicating factor is that a particular gene may encode "products" with multiple functional roles, and a particular trait is generally influenced by multiple genes. A particular allele of a particular gene may influence different functional roles and traits differently, something to keep in mind in the following discussion which, for simplicity's sake, focusses on a single functional role of a gene product and its influence on a single (simple) trait.



An allele can produce a gene product with completely normal function(s) or no (apparent) functional activity at all, referred to as a null or amorphic allele.The allelic form can have less function than the "wild type" allele (hypomorphic), more function than the wild type (hypermorphic), or a new function (neomorphic). Given that many gene products function as part of multimeric complexes, encoded by multiple genes, and that many organisms (like us) are diploid, there is one more formal possibility, the product of one allele can antagonize the activity of the other - this is known as an antimorphic allele. These different types of alleles were defined genetically by Herbert Muller (1890-1967), who won the Nobel prize for showing that X-rays could induce mutations, new alleles (Appendix 1). The characterization of an allele is typically carried out with respect to how it influences a specific trait. Again, remember that most traits are influenced by multiple genes, and a single gene can influence multiple traits.

The most common version of an allele is often referred to as the wild type allele (← thanks Mercer Mayer), but that is really just because it is the most common. There are often multiple alleles of a particular gene in a population and they all may be equally "normal", although they may influence different traits differently. If there is no significant selective advantage between them, their relative frequencies within a population can drift over time. At the same time, the phenotype(s) associated with a particular allele can be influenced by the alleles present at other genetic loci, known collectively as the genetic background. Since most traits are the results of many genes functioning together, different combinations of alleles can produce different effects. The universe of variation is large. This can make identifying the genetic basis of a trait or a disease difficult, particularly when variation at any one locus may make have only a minor influence on a trait or a disease phenotype. On top of that, stochastic effects together with environmental and developmental differences can outweigh genetic influences on phenotypic outcome. Genetic background effects can lead to a particular allele producing a disease in one person and not another.[308]

Mutations are the ultimate source of genetic variation – without them evolution would not occur. A mutations can have a number of effects, in particular, they can create new activities. At the same time most mutations reduce or alter the original (and necessary) activity of a gene, an activity than might be essential. Here is an example for how a mutation can be creative. While it is common to think of a particular gene product having a single activity, when examined closely many proteins with catalytic activity can catalyze "off-target" (sometime termed promiscuous) reactions.[309] A mutation can enhance the off target activity. If that activity is useful, it can influence reproductive success and so be "selected", assuming that the mutation does not disrupt any essential function(s) of the gene product.

Genomic rearrangements, which are mutations because they change genome sequence, can occur during embryonic development.[310] In the case illustrated here (↓), imagine that an essential but multifunctional gene is duplicated and moved elsewhere in the genome. Now one copy can continue to carry out its original essential function, while the second (new) copy is free to change as long as it does not interfere with the function(s) of the essential gene. While many mutations will negatively effect the duplicated gene, some may increase and refine its favorable ancillary function. A new gene can emerge freed from the need to continue to perform its original function. We see evidence of this type of process throughout the biological world. When a gene is duplicated, the two copies are known as paralogs. Such paralogs can evolve independently.

References

[308] Genetic background effects & How do stochastic processes and genetic threshold effects explain incomplete penetrance and inform causal disease mechanisms?

[309] Shining a light on enzyme promiscuity

[310] Copy Number Variation in Human Health, Disease, and Evolution and LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes?

This page titled 7.12: Genes and alleles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.13: The origin of new (de novo) genes

A key question you might be asking yourself is where, exactly, do brand new (de novo) genes come from?[311] A hint has been found from studies of RNA synthesis. It was once thought that only the coding regions of genes were used to synthesize RNA, but higher resolution RNA sequencing and mapping techniques have revealed that a large percentage (~80%) of the genome serves to direct RNA synthesis. This includes regions that do not appear to encode polypeptides or non-coding but functional RNAs. This opens the possibility that some transcribed DNA regions produce RNAs with regulatory roles or that are translated into polypeptides. If such gene products enhance reproductive success within a population, they can be "selected" for and may become stable parts of the organisms' genome. There is evidence for such events in the evolution of fruit flies and humans.[312]

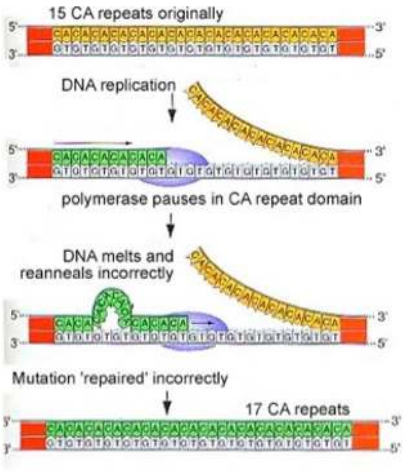
References

[311] Proto-genes and de novo gene birth [**link**] and How evolution builds genes from scratch [**link**]

[312] Origin and spread of de novo genes in Drosophila melanogaster populations [**link**], Origins of *De Novo* Genes in Human and Chimpanzee [**link**] and De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population [**link**]

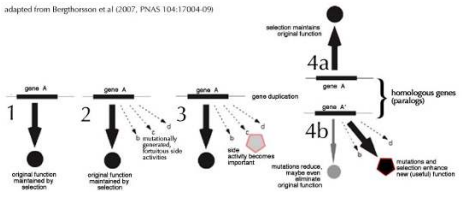
This page titled 7.13: The origin of new (de novo) genes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

7.14: DNA repeat diseases and genetic anticipation



While unavoidable and essential for evolution, defects in DNA synthesis and genomic rearrangements generally lead to inherited diseases rather than benefits. While we will return to mutational mechanisms and their effects, here we briefly consider diseases associated with DNA replication, specifically the class of genetic diseases known as trinucleotide repeat disorders (→). There are a number of such "triplet repeat" diseases, including several forms of mental retardation, Huntington's disease, inherited ataxias, and muscular dystrophies. These diseases are caused by slippage of DNA polymerase and the subsequent duplication of sequences. When these "slipable" repeats occur in a region of DNA encoding a protein, they can lead to regions of repeated amino acids. For example, expansion of a domain of CAGs in the gene encoding the polypeptide Huntingtin (OMIM:613004) leads to the neurological disorder Huntingdon's disease (or Huntingdon's chorea). OMIM stands for the "On-line Inheritance in Man" website.

A mechanistically related pathogenic syndrome is known as Fragile X (OMIM:300624). It is due to a DNA replication defect. Fragile X is the most common form of "autism of known cause". It is worth noting that notwithstanding some claims, most forms of autism have no known cause. About 6% of autistic individuals have Fragile X syndrome. Fragile X syndrome can also lead to anxiety disorders, attention deficit hyperactivity disorder, psychosis, and obsessive-compulsive disorder. The Fragile X phenotype involves the *FMR1* gene (OMIM:309550), located on the X chromosome. The disease affects mainly males, who are XY, compared to females, who have two X chromosomes. In the unaffected population, alleles of the *FMR1* gene have between 6 to 50 copies of a CGG repeat and are phenotypically normal. People with 50 to 200 repeats carry what is known as a pre-mutation; they rarely display symptoms but can transmit a "predisposition" to their children. Those with more than 200 repeats typically display symptoms and often have what appears to be a broken X chromosome – from which the disease derives its name. The pathogenic sequence in Fragile X is downstream of the *FMR1* gene's coding region. When this region expands it leads to the inhibition of the expression of the *FMR1* gene.[313] There are a number of processes that can mediate the pathogenic effects of DNA repeat diseases, some of which we will consider when we discuss the inheritance of these conditions.



Other DNA Defects

Defects in DNA repair can lead to severe diseases and often a susceptibility to cancer. A search of OMIM for DNA repair returned 654 entries! For example, defects in DNA mismatch repair lead to a susceptibility to colon cancer, while defects in transcription-coupled DNA repair are associated with Cockayne syndrome. People with Cockayne's syndrome (OMIM:216400 & 133540) are sensitive to light, are of short stature, and appear to age prematurely.[314]

Our introduction to genes has necessarily been quite foundational and we will extend it somewhat in the second half of the course. There are lots of variations and associated complexities that occur within the biological world. The key idea is that genes represent biologically meaningful DNA sequences. To be meaningful, the sequence must play a role within the organism, typically by encoding a polypeptide (which we will consider next) and/or the information needed to ensure its correct expression, that is, where and when the information in the gene and the genome is expressed. A practical problem is that most studies of genes are carried out using organisms grown in the lab or in otherwise artificial or unnatural conditions. It might be possible for an organism to exist with an amorphic allele of a gene in the lab, while organisms that carry that allele may be at a significant reproductive disadvantage in the "real world". Moreover, a particular set of alleles, a particular genotype, might have a reproductive advantage in one environment, one ecological/behavioral niche but not another. Measuring these effects can be difficult. All of which serves as a warning that we should consider skeptically pronouncements that a gene, or more accurately a specific allele of a gene, is responsible for a specific trait, particularly if the trait is complex, ill-defined, and likely to be significantly influenced by genomic context (the rest of the genome and genetic background effects) and environmental factors. Intelligence is one such complex trait. A dramatic example of the difficulty in defining a gene product's functions is illustrated by the studies of Hutchinson et al; they

produced a minimal bacterial genome containing 473 genes.[315] Of these genes, the function(s) of 149 (~32% of the total genome) were unknown, a rather surprising result.

Questions to answer

117. How does a mutation generate a new allele? How is a mutation different from an allele?
118. What would be a reasonable way to determine that you had defined an entire gene?
119. Is it possible to build a system (through evolutionary mechanisms) in which mutations do not occur?

Questions to ponder:

- How could removing information from the genome enhanced reproductive success?

- How might you go about defining the function of a “gene with unknown function”?



Short Chapter Summary

- *DNA is the primary hereditary material; structure explains replication fidelity and limits.*
- *Replication uses coordinated molecular machines with proofreading and repair; errors fuel evolution.*
- *“Gene” is context-dependent; telomeres, topoisomerases, repeats, and repair shape genomes*

Reference

[313] - Molecular mechanisms of fragile X syndrome: a twenty-year perspective.

[314] Cockayne syndrome: http://omim.org/entry/278760

[315] Design and synthesis of a minimal bacterial genome. https://www.ncbi.nlm.nih.gov/pubmed/27013737

This page titled 7.14: DNA repeat diseases and genetic anticipation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

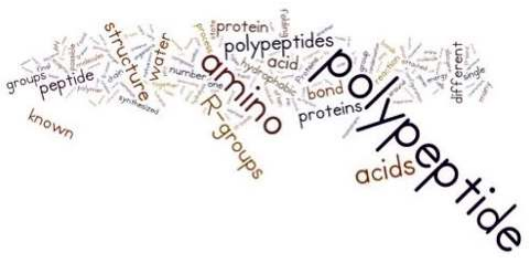
8: Peptide Bonds, Polypeptides and Proteins

8.0: Introduction
8.1: Specifying a polypeptide’s sequence
8.2: The origin of the genetic code
8.3: Protein synthesis- transcription (DNA to RNA)
8.4: Protein synthesis- translation (RNA to polypeptide)
8.5: The translation (polypeptide synthesis) cycle
8.6: Effects of point mutations on polypeptides and proteins
8.7: Getting more complex- gene regulation in eukaryotes
8.8: Turning polypeptides into proteins
8.9: Regulating protein activity, concentrations, and stability (half-life)
8.10: Allosteric and post-translational regulation
8.11: Diseases of folding and misfolding
8.12: Molecular machines

This page titled 8: Peptide Bonds, Polypeptides and Proteins is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.0: Introduction

In which we consider the nature of proteins, how they are synthesized and assembled, how they get to where they need to go within the cell and within the organism, how they function, how their activities are regulated, and how mutations can influence their expression, stability, activity, and evolution.



We have mentioned proteins many times, since there are few biological processes that do not rely on them. Proteins act as structural elements, signals, regulators, and catalysts in a wide range of molecular machines. We have yet to say much about what they are, how they are made, and how they come to do what they do. The first scientific characterization of what are now known as proteins was published by the Dutch chemist, Gerardus Mulder (1802–1880).[316] After an analysis of a number of different substances, he proposed that all proteins contain a common chemical core, with the molecular formula $C_{400}H_{620}N_{100}O_{120}P_1S_1$, and that the differences between proteins were primarily in the numbers of phosphate (P) and sulfur (S) atoms they contain. The name “protein”, from the Greek word πρώτα (“prota”), meaning “primary”, was suggested by Jons Berzelius (1779–1848) based on the presumed importance of these compounds in biological systems.[317] As you can see, Mulder’s molecular formula is not very informative, it suggests that all proteins are fundamentally similar, which while true is confusing since they have so many different roles. Subsequent studies revealed that proteins could be dissolved in water or dilute salt solutions but formed insoluble aggregates when the solution was heated. As we will see this heat-induced aggregation was due to changes in protein structure. Mulder was able to break down proteins into amino acids through an acid hydrolysis reaction. Amino acids get their name from the fact that they contain both an amino ($–NH_2$) and a carboxylic acid (–COOH) group. While there many thousands of possible amino acids, only twenty (or rather twenty two, as we will see) were identified in hydrolyzed samples of proteins. Since their original characterization as a general class of compounds, we now understand that while proteins share a common basic polymer structure, they are remarkably diverse. Proteins are involved in roles from the mechanical strengthening of skin to the movement and control of the body, the building of shells and claws, the regulation of genes, the transport of oxygen, the capture of energy, the release of light, and the catalysis and regulation of essentially all of the chemical reactions that occur within cells and organisms.

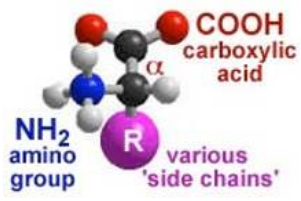
With the introduction of various methods, it was discovered that different proteins were composed of distinct sets of polymers. Each polymer was an unbranched chain of amino acids, linked by -C-N- peptide bonds, with a specific amino acid sequence. These polymers are known generically as polypeptides. Specific proteins are composed of specific sets of polypeptides and each distinct polypeptide is encoded by a distinct gene. In addition many proteins contain other molecular components, known as co-factors or prosthetic groups. These co-factors can range from metal ions to various small molecules. A protein is a fully assembled and functional entity.

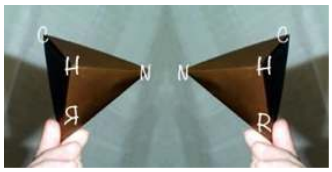
As you may recall a carbon atom (C) typically takes part in four covalent bonds. We can think of an amino acid as a modified form of methane (CH_4), with the C referred to as the alpha carbon (C_α). In an amino acid there is an H, an amino group ($-NH_2$), a carboxylic acid group (-COOH), and a variable (R) group attached to the alpha C_α atom. Different amino acids differ from one another by their R-groups, often referred to as "side-chains". Some R-groups are larger, some are small, some are hydrophobic, some are hydrophilic and some hydrophilic R-groups contain weak acidic or basic groups. The extent to which these weak acidic or basic groups are positively or negatively charged changes in response to environmental pH. Changes in side-group charge can (as we will see) influence the structure of the polypeptide/protein. The different R-groups provide proteins with a broad range of chemical properties that are further extended by the presence of co-factors.[318]

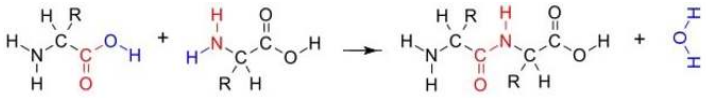
The four groups attached to the α-carbon in an amino acid are arranged at the vertices of a tetrahedron (→). If four different groups attached to the α-carbon, as they are in all biological amino acids except glycine, the resulting amino acid can exist in two forms, known as L- and D- form enantiomeric stereoisomers. Enantiomers are mirror images of one another. Only L-type amino acids are found in proteins, even though there is no obvious chemical reason for why proteins could not have also been made using both types or only D-amino acids.[319] It appears that the universal use of L-type amino acids in biological polypeptides found is an example of the evolutionary relatedness, a homologous trait presumably established in LUCA. Even though there are hundreds of

amino acids known, only 22, the 20 common amino acids and selenocysteine and pyrrolysine, are found in proteins and presumably were present in LUCA.

As we noted for nucleic acids, a polymer is a chain of subunits. In a polypeptide, amino acid are linked together by peptide bonds. Under the conditions that exist inside the cell, this is a thermodynamically unfavorable dehydration reaction, and so polypeptide synthesis must be coupled to a thermodynamically favorable reaction, a nucleotide triphosphate hydrolysis reaction. A molecule formed from two amino acids, joined together by a peptide bond, is known as a dipeptide. A dipeptide has an N-terminal (amino) and a C-terminal (carboxylic acid) end. To generate a polypeptide, new amino acids are added sequentially to the polymer's C-terminal end – a reaction analogous to the synthesis of a polynucleotide, in which monomers are added only to one end of the growing polymer. A peptide bond forms between the amino group of the added amino acid and the carboxylic acid group of the polymer. The formation of a peptide bond is associated with the release of a water molecule (↓). The addition of an amino acid to the C-terminus of a polypeptide generates a new C-terminal carboxylic acid group. While some amino acids have a carboxylic acid group as part of their R-group, new amino acids are not added there. The result is that polypeptides are synthesized as unbranched, linear polymers.







The process of amino acid addition can continue, theoretically without limit. Biological polypeptides range from the short (5-10) to the very long – hundreds to thousands of amino acids in length.[320] The Titin polypeptide found in muscle cells can be more than 30,000 amino acids in length.[321] Because there is no theoretical constraint on which amino acid occurs at a particular position within a polypeptide, there is a enormous number of possible polypeptides that can exist. In the case of a 100 amino acid long polypeptide, there are more than 20^{100} possible different polypeptides that could, in theory, be formed.

Questions to answer:

120. How does a polypeptide chain resemble and how does it differ from a nucleic acid molecule?

Question to ponder:

- Why does it make sense to think that the presence of a common set of amino acids in organisms is a homologous trait?

References

[316] From ‘protein’ to the beginnings of clinical proteomics: http://www.ncbi.nlm.nih.gov/pubmed/21136729

[317] The original claim that proteins get their name from “the ancient Greek sea-god Proteus who, like your typical sea-god, could change shape. The name acknowledges the different properties and functions of proteins.” seems more satisfying to us.

[318] Bioengineers are working to go Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon & to incorporation of non-canonical amino acids into proteins in yeast; something made possible due to the redundancy of the genetic code.

[319] It is not that D-amino acids do not occur in nature, or in organisms, they do. They are found in biomolecules, such as the antibiotic gramicidin, which is composed of alternating L-and D-type amino acids. Gramicidin is, however, synthesized by a different process than that used to synthesize proteins.

[320] Short polypeptides, or rather the genes that encode them, can be difficult to recognize since short “open reading frames” are difficult to identify unambiguous: see Peptidomic discovery of short open reading frame–encoded peptides in human cells

[321] OMIM entry for TITIN: http://omim.org/entry/188840

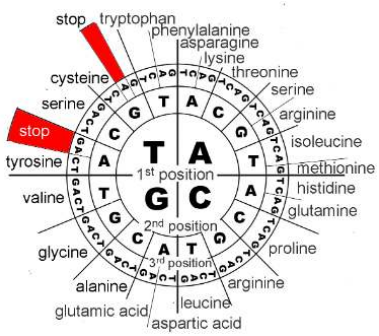
This page titled 8.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.1: Specifying a polypeptide's sequence

You may be asking yourself, if there are so many different possible polypeptides and no inherent bias favoring the addition of one amino acid over another, what determines the sequence of amino acids within a polypeptide - presumably it is not random. Here we connect the specification of polypeptide sequence to the information stored in DNA. We begin with a description of the process in bacteria and then extend it to archaea and eukaryotes. We introduce them in this order because, while basically similar (homologous), the system is somewhat simpler in bacteria, although you might find it complex enough for your liking. We will leave most of the complexities for subsequent courses. We will consider the network dynamics of these systems and ask you to make plausible predictions about their behavior in response to various perturbations, mutations and such. A point to keep in mind is that the system is continuous. The machinery required for protein synthesis is inherited by the cell; new copies of it are synthesized as the cell grows. Each new polypeptide is synthesized in an environment full of pre-existing proteins and ongoing metabolic processes.

A bacterial cell synthesizes thousands of different polypeptides. The sequence of these polypeptides, the exact amino acids from the N-terminal start to the C-terminal end, is encoded within the organism's DNA. The bacterial genome is a double-stranded circular DNA molecule that is (typically) millions of base pairs in length. Each polypeptide is encoded by a specific region of this DNA molecule. So, how are specific DNA regions recognized? transcribed into RNA and how is the information in DNA **translated** into a polypeptide?

To address the first question think back to the structure of DNA. Perhaps the one-dimensional sequence of a polypeptide could be encoded in the one-dimensional sequence of a polynucleotide chain.[322] The question was how to translate the language of nucleic acids, sequences of four different nucleotides, into the language of polypeptides, sequences of the 22 different amino acids. As pointed out by the physicist George Gamow (1904-1968)[323] the minimum set of nucleotides needed to encode all 20-22 amino acids is three; a sequence of one nucleotide (4^1) could encode at most four different amino acids, a two nucleotide sequence could encode (4^2) or 16 different amino acids (not enough), while a three nucleotide sequence (4^3) could encode 64 different amino acids (more than enough).[324] Although the actual coding scheme that Gamow proposed was wrong, his thinking about the coding capacity of DNA influenced those who set out to experimentally determine the actual rules of the "genetic code".



The genetic code is not the information itself, but the algorithm by which nucleotide sequences are "decoded" to determine polypeptide sequences. A polypeptide is encoded by the sequence of nucleotides. This nucleotide sequence is read in groups of three nucleotides, known as a codon. Each codons is read in a non-overlapping manner, with no spaces (that is, non-coding nucleotides) between them. Since there are 64 possible codons but only 22 different amino acids used in organisms, the code is redundant, that is, certain amino acids are encoded for by more than one codon. In addition there are three codons, UAA, UAG, and UGA, that (in most organisms) do not encode any amino acid but are used to mark the end of a polypeptide, they encode "stops" or periods (→). The amino acids selenocysteine and pyrrolysine are encoded for by stop codons that occur in a specific "context" of surrounding nucleotides.[325]

The region of the nucleic acid that encodes a polypeptide begins with what is known as a "start" codon and continues with a number of In-frame codons until one or more of the three stop codons is reached.[326] This "coding" sequence is known as an open reading frame, an ORF. While the information encoding a polypeptide is present in the DNA, this information is not used directly to specify the sequence of a polypeptide. The process involves an intermediate. The information in the DNA is first copied (transcribed) into an mRNA molecule. The mRNA moleculedirects polypeptide synthesis. As noted, the process of copying information from a DNA into an RNA molecule is known as transcription because DNA and RNA use the same nucleotide sequence language. In English, as opposed to molecular biology, transcription is the process of making a written copy of what someone says - the language of both is the same. In contrast polypeptides are written in a different language, amino acid sequences. The process of RNA-directed polypeptide synthesis is known as translation; it involves changing between languages, from nucleic acid-ese to polypeptide-ese.

References

[322] Nature of the genetic code finally revealed!: http://www.nature.com/nrmicro/journal/v9/n12/full/nrmicro2707.html

[323] - when he was a professor at UC Boulder

[324] The Big Bang and the genetic code: Gamow, a prankster and physicist, thought of them first

[325] In the case of Selenocysteine a stop codon is recognized by a specific tRNA.

[326] There are situations in which non-start codons occur: see repeat-associated non-ATG translation (RAN translation)

This page titled 8.1: Specifying a polypeptide's sequence is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.2: The origin of the genetic code

How exactly the genetic code arose is unknowable, but a number of plausible scenarios have been proposed. One is the frozen accident model in which the code used in modern cells is the result of an evolutionary accident, a bottleneck event associated with the appearance of LUCA. Early in the evolution of life on Earth, there may have been multiple types of proto-organisms, each using a different genetic code. The common genetic code found in all existing organisms reflects the fact that only one of these proto-organisms gave rise to all modern organisms. Alternatively, the code could reflect specific interactions between RNAs and amino acids that played a role in the initial establishment of the code. Unless we find forms of life that arose independently, we cannot be sure. What is clear is that the code is not fixed, there are examples in which certain codons are “re-purposed” in various organisms.[327] In fact there are efforts to re-engineer codons to produce proteins made using a range of more than 100 “unnatural” amino acids.[328] What these variations in the genetic code illustrate is that evolutionary mechanisms can change the genetic code.[329] Since the genetic code does not appear to be predetermined, the conservation of the genetic code among organisms is taken as strong evidence that all organisms, even the ones with minor variations in their genetic codes, are derived from a single common ancestor. The genetic code appears to be homologous trait shared by all organisms and viruses.

References

[327] Organisms with alternative genetic codes resolve unassigned codons via mistranslation and ribosomal rescue. (link) see also Knight et al., 2007. Rewiring the keyboard: evolvability of the genetic code. (link)

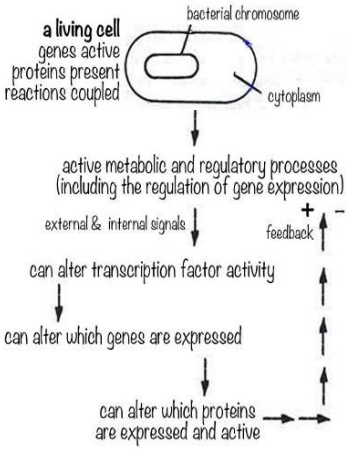
[328] Designing logical codon reassignment – Expanding the chemistry in biology

[329] The genetic code is nearly optimal for allowing additional information within protein-coding sequences & Stops making sense: translational trade-offs and stop codon reassignment:

This page titled 8.2: The origin of the genetic code is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.3: Protein synthesis- transcription (DNA to RNA)

Having introduced DNA, mRNA, and the genetic code, we turn to the process by which a polypeptide, specified by a DNA sequence, is synthesized in a cell. First how is the specific DNA region that encodes a specific polypeptide recognized. We are looking for a relatively short region within millions (in prokaryotes) or billions (in eukaryotes) of base pairs of DNA. This information will be present in one strand of a double-stranded region of DNA. From the point of view of polypeptide synthesis, the other strand is effectively nonsense. This means that a gene's regulatory sequence must specify which strand is used to generate a mRNA, as well as where, when and how often RNA synthesis starts.



To "find" a gene a specific nucleotide sequences is used, together with something (a protein) that recognizes and binds to that specific nucleotide sequence. Let us consider a particular form of the problem; we want to uniquely specify one gene (one sequence) within the ~3,000,000 base pairs of an *E. coli* cell's genomic DNA. For simplicity let us assume that the A:T ratio equals the G:C ratio. Clearly a one base pair sequence will not work, since we might expect that half of the base pairs will be recognized, either directly as T or indirectly as A bonded to T on the other strand. To be unique the sequence we want must occur once in 3,000,000 base pairs (1/3,000,000 = 3.33… x 10^{-7} = 0.000000333). If we use a two base sequence, it will occur 1/4 x 1/4 = 1/16 = 0.0625, a four base sequence 0.0039, an eight base sequence 0.00001523, but a 16 base sequence has a probability of occurring purely by chance of ~2.32 x 10^{-10}, which is less than once per genome.[330]

There are DNA sequence specific binding proteins that acts to control transcription, they are known generically as transcription factors. Positively acting transcription factors act to recruit DNA-dependent, RNA polymerase to a specific site and strand on the DNA, upstream of a gene's coding region. A negatively acting transcription factor can block the activity of positively acting factors. An positively acting transcription is the first step in gene expression (→). The combination of transcription factors present and their activity in a cells will determine gene activity.

Gene specific RNA synthesis is mediated by DNA-dependent, RNA polymerase, encoded by specific genes. As in the case of DNA-dependent DNA synthesis, RNA synthesis is also subject to errors; estimates for the error rate range from 10^{-3} to 10^{-6} errors per nucleotide incorporated.[331] Where, and in which orientation, the polymerase binds to the gene's DNA is determined by the gene's regulatory sequence(s) and the orientation of the proteins bound to it. Which genes are expressed is determined by which transcription factors are present and active, their concentrations, and their binding affinities for specific regulatory sequences, as well as the presence of competing binding sites. Polymerase binds and is activated through binding to a gene's DNA-transcription factor complex. Since there are many genes in the genome, the stability of the DNA-Transcription Factor-Polymerase complex will impact the number of RNAs synthesized per unit time. A similar process is used for genes that encode "non-coding" RNAs.

At this point, it is useful to explicitly recognize some common aspects of biological systems. These systems are highly regulated, adaptive, and homeostatic; they adjust their behaviors to changes in their environment (both internal and external) in order to maintain the living state. These behaviors are based on various forms of feedback regulation. In bacterial gene expression, there are genes that encode specific transcription factors. Which of these genes are expressed, together with the half-lives of transcribed RNAs and encoded proteins determines the concentrations of the various transcription factors present and, in turn, the pattern of gene expression. Of course, the gene(s) encoding a specific transcription factor are regulated. Transcription factors can act positively or negatively; they can lead to the activation of transcription by recruiting and activating the RNA polymerase or by blocking its recruitment and/or activation. In addition the activity of a particular transcription factor can be regulated. Please note that most transcription factors are involved in the regulation of multiple (often hundreds or more) genes, can act either positively or negatively, and that their binding affinity for the regulatory sequences of different genes can differ.

All organisms are complex. A "simple" bacterium contains thousands of genes and different sets of genes are used in different environments and situations, and in different combinations to produce specific behaviors. In some cases, these behaviors may be mutually antagonistic. For example, a bacterium facing a rapidly drying out environment might turn off genes encoding proteins involved in growth and turn on genes to prepare to survive in a more hostile environment. Our goal here is not to generate perfectly accurate predictions about the behavior of an organism in a particular situation, something that may not be possible without specific

(provided, memorized) information, but rather to help you consider plausible models for how gene expression might change in response to various perturbations. This requires us to consider, although at a elementary level, the regulatory processes active in cells.

For a transcription factor to regulate a gene, whether positively or negatively, it must be able to bind to specific DNA regulatory sequences with a reasonable affinity, meaning that it stays bound long enough to recruit (or inhibit) the binding and activation of other factors, concluding with RNA polymerase, before being knocked off by thermal collisions. Groups of genes that are expressed together, in response to common cellular, developmental, or environmental conditions, are likely to be regulated by common sets of transcription factors. The activation or inactivation of a transcription factor can involve a number of mechanisms, including its modification or interactions with other proteins that alter its ability to interact with its target DNA sequence, other proteins, or RNA polymerases.

Transcription factors are synthesized in the cytoplasm. In prokaryotic cells, without a nucleus, it can bind to its DNA targets. In eukaryotic cells it first has to enter the nucleus before it can bind to its target DNA sequences. When an RNA polymerase collides with the DNA-transcription factor complex, it can remain bound and initiate the process by which the RNA polymerase is activated and begins RNA synthesis. Once RNA polymerase has been activated, it will move away from the transcription factor-DNA complex. The DNA bound transcription factor complex can then bind another polymerase or it can disassemble, releasing the transcription factors from the DNA (in response to molecular level collisions). Once released transcription factors can diffuse away, interact with other regulatory factors, or rebind to other sites in the DNA. Clearly the number of copies of a particular transcription factor protein present (its concentration), its target site affinity, and the frequency of its various DNA binding sites will impact the behavior of the system, as will the concentrations of ancillary factors that interact with the transcription factor/DNA complex to recruit and activate the polymerase.

RNA synthesis is a thermodynamically unfavorable reaction, so for it to occur it must be coupled to a thermodynamically favorable reaction, in particular nucleotide triphosphate hydrolysis reactions. The RNA polymerase moves along the DNA to generate an RNA molecule (the transcript). Other signals within the DNA, and recognized by proteins associated with the transcription machinery, lead to the termination of transcription and the release of the RNA polymerase. Once released, the RNA polymerase returns to its inactive state. It can act on the same or another gene if the RNA polymerase interacts with transcription factors bound to the gene's promoter. Since multiple types of transcription factor proteins are present within the cell and RNA polymerase can interact with all of them, which genes are expressed within a cell will depend upon the relative concentrations and activities of specific transcription factors and their regulatory and associated proteins, together with their binding affinities for specific DNA sequences compared to their general low-affinity binding to DNA.

References

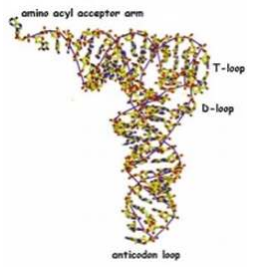
[330] As we will return to, the CRISPR-Cas9 system for mutagenesis uses a 22-base "guide RNA" to direct an endonuclease; this, in theory at least, would be expected to guarantee one target per genome.

[331] Gout et al., 2017 The landscape of transcription errors in eukaryotic cells [link]

This page titled 8.3: Protein synthesis- transcription (DNA to RNA) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.4: Protein synthesis- translation (RNA to polypeptide)

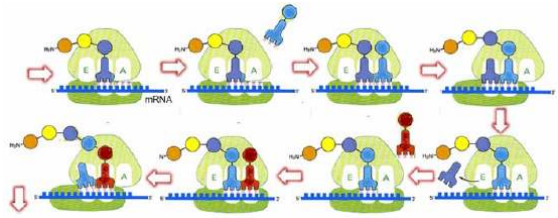
Translation involves a complex cellular organelle, the ribosome that works together with a number of accessory factors to read the code in an mRNA molecule and direct the synthesis of the corresponding polypeptide.[332] The ribosome holds the various components, the mRNA and accessory factors, including translation-associated or tRNAs, in appropriate juxtaposition to one another, to catalyze polypeptide synthesis. But perhaps we are getting ahead of ourselves. For one, what exactly is a tRNA?



Transcription generates a number of types of RNAs beside mRNAs. Of these non-mRNAs, two are central to mRNA-directed polypeptide synthesis. The first are transfer RNAs (tRNAs). These small single-stranded RNA molecules fold back on themselves to generate a compact L-shaped structures (←). In *E. coli*, there are 87 genes that encode tRNAs; there are over 400 tRNA encoding genes in humans. For each amino acid and each codon there are one or more tRNAs. The exceptions are the so called stop codons, for which there are no tRNAs. A tRNA specific for the amino acid phenylalanine would be written $tRNA^{Phe}$. Two parts of a tRNA molecule are functionally linked: the anti-codon loop, which recognizes the codon through base pairing interactions within the ribosome-bound mRNA complex and the amino acid acceptor stem, which is where an amino acid is covalently attached to the tRNA. The rest of the tRNA molecule mediates interactions with enzymes known as amino acyl tRNA synthetases. There is a distinct amino acyl tRNA synthetase for each amino acid: there is a phenylalanine-tRNA synthetase and a proline-tRNA synthetase, etc. An amino acyl tRNA synthetase binds the appropriate tRNA and the appropriate amino acid and, through a reaction coupled to a thermodynamically favorable nucleotide triphosphate hydrolysis reaction, catalyzes the formation of a covalent bond between the tRNA's amino acid acceptor stem and the amino acid, to form what is known as a charged or amino acyl-tRNA. In the course of polypeptide synthesis, the amino acid group attached to the tRNA's acceptor stem will be transferred from the tRNA to the end of the growing polypeptide.

Ribosomes are composed of roughly equal amounts (by mass) of ribosomal RNAs (rRNAs) and polypeptides. An active ribosome consists of a small and a large ribosomal subunit. In the bacterium *E. coli*, the small subunit is composed of 21 different polypeptides and a 1542 nucleotide long rRNA molecule, while the large subunit is composed of 33 different polypeptides and two rRNAs, one 121 nucleotides long and the other 2904 nucleotides long.[333] Each ribosomal polypeptide and RNA is itself a gene product. The complete ribosome has a molecular weight of ~3 x 10^6 daltons. Please note, there is no reason to remember these numbers except to appreciate that ribosomes are complex molecular machines. The large subunit rRNAs is an evolutionarily conserved catalyst, known as a ribozyme, in analogy to protein based catalysts, enzymes. The ribozyme lies at the heart of the ribosome and catalyzes the transfer of an amino acid bound to a tRNA to the carboxylic acid end of the growing polypeptide chain. RNA based catalysis is a conserved feature of polypeptide synthesis and appears to be an evolutionarily homologous trait.

The growing polypeptide chain is bound to a tRNA, known as the peptidyl tRNA. When a new aa-tRNA enters the ribosome's active site (site A), the existing polypeptide is added to it, the amino acid originally associated with the incoming aa-tRNA becomes the terminal amino acid of the growing peptide (↓). The "old" peptidyl tRNA is released and tRNA-peptide complex shifts position.



The cytoplasm of cells is packed with ribosomes. In a rapidly growing bacterial cell, ~25% of the total cell mass consists of ribosomes. Although structurally similar, there are differences between bacterial, archaeal, and eukaryotic ribosomes, a point of significance since a number of antibiotics act by selectively inhibiting bacterial but not eukaryotic ribosome-mediated protein synthesis. Both chloroplasts and mitochondria have ribosomes of the bacterial type; another piece of evidence that they are descended from bacterial endosymbionts. Protein synthesis blocking anti-bacterial antibiotics are mostly benign since they do not block most of the protein synthesis that occurs in a eukaryotic cell.

References

[332] Can't stop yourself? go here for a more detailed description of translation.

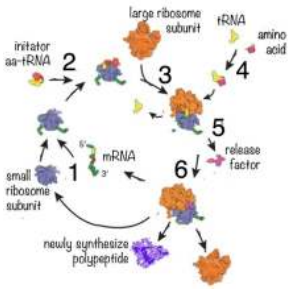
[333] In the human, the small ribosomal subunit is composed of 33 polypeptides and a 1870 nucleotide rRNA, while the large ribosomal subunit contains 47 polypeptides, and three rRNAs of 121, 156, and 5034 nucleotides in length.

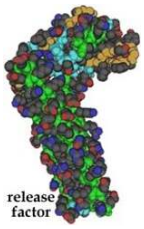
This page titled 8.4: Protein synthesis- translation (RNA to polypeptide) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.5: The translation (polypeptide synthesis) cycle

As we walk through the process of protein synthesis, keep in mind that at each step there are accessory factors involved in regulating the process and coupling it to the thermodynamically favorable reactions that make it possible. While important to consider if you want to re-engineer or manipulate the translation system, we think they obscure an understanding of the underlying process and are best considered in more advanced courses. It is important to recognize that mRNA-directed polypeptide synthesis (translation) occurs only because all the components needed already exist in the cell. The interactions between these components are based on stochastic, thermally driven collisions. For example, consider the addition of an amino acid to a tRNA, the formation of an amino acyl-tRNA or aa-tRNA. Random motions bring the correct amino acid and the correct tRNA to their binding sites on the appropriate amino acyl tRNA synthetase. Once the aa-tRNA is formed, (usually) only the correct amino acid charged tRNA will bind productively to the ribosome-mRNA-nascent polypeptide complex. If the wrong amino acid is inserted, the ribosome stalls and catalyzes its removal.(footnote 343) Generally, many unproductive collisions occur before a productive (correct) one occurs, since there are more than 22 different amino acid/tRNA molecules bouncing around in the cytoplasm. The stochastic and error correcting aspects of the peptide synthesis process are, however, rarely illustrated.

The first step in polypeptide synthesis is the synthesis of the mRNA (←) that encodes the polypeptide. That mRNA (1) contains a sequence, located near the 5’ end of the mRNA, that mediates its productive binding to the small ribosomal subunit.[334] (2) The mRNA-small ribosome subunit complex now interacts with and binds to a complex containing an initiator (start) amino acid:tRNA. In bacteria, archaea, and eukaryotes the start codon is generally an AUG codon and encodes the amino acid methionine, although other, non-AUG start codons are possible.[335] This start codon-tRNA complex defines the beginning of the polypeptide as well as the coding region’s reading frame. (3) The met-tRNA:mRNA:small ribosome subunit complex can now interact with a large ribosomal subunit to form the functional mRNA:ribosome complex. (4) Charged amino acyl tRNAs can interact with the mRNA:ribosome complex. Based on the mRNA sequence and its start codon-defined reading frame, amino acids will be added sequentially. With each new amino acid added, the ribosome moves along the mRNA. An important point, that we will return to when we consider the folding of polypeptides into their final three-dimensional shapes is that the newly synthesized polypeptide is threaded through a molecular tunnel within the ribosome. Only after the N-terminal end of the polypeptide emerges from this tunnel does the nascent polypeptide begin to fold. (5) The process of polypeptide polymerization continues until the ribosome reaches a stop codon, that is a UGA, UAA or UAG.[336] Since there are no tRNAs that recognize these codons, the ribosome pauses, waiting for a charged tRNA that will never arrive. Instead, a polypeptide known as release factor, with a shape something like a tRNA (→), binds to the polypeptide:mRNA:ribosome complex instead. (6) This leads to the release of the polypeptide, the disassembly of the ribosome into small and large subunits, and the release of the mRNA.[337] When associated with the ribosome, the mRNA is protected against interactions with proteins (ribonucleases) that could catalyze its degradation. Upon its release from the ribosome, an mRNA may interact with a new small ribosomal subunit, and begin the process of polypeptide synthesis again or it may interact with a ribonuclease and be degraded. Where it is important to limit the synthesis of a particular polypeptide, the relative probabilities of these two events, new translation versus RNA degradation, will be skewed in favor of degradation. An RNA's stability is regulated by the binding of specific proteins to nucleotide sequences within the mRNA. The relationship between mRNA synthesis and degradation will determine the half-life of the population of mRNA molecules, the steady state concentration of the mRNA in the cell, and indirectly, the level of the encoded polypeptide present.





In eukaryotes, RNA synthesis and processing occur primarily in the nucleus, while polypeptide synthesis occurs in the cytoplasm. In contrast, in bacteria and archaea, there is no barrier between the cell’s DNA and the cytoplasm, which contains the ribosomal subunits together with the other components involved in polypeptide synthesis. The newly synthesized 5' end of an mRNAs is deposited directly into the cytoplasm, where they can interact with ribosomes. In fact, the process of protein synthesis (translation) can begin before mRNA synthesis (transcription) is complete; the two processes can interact in interesting ways.[338] Similar interactions can occur in eukaryotic mitochondria and chloroplasts, not surprising as they seem to be derived (evolutionarily) from bacteria.

Questions to answer:

121. Speculate on how the presence of L- and D-amino acids in proteins would influence the DNA code used to specify the amino acid sequence of a polypeptide.
122. How might the concentration of various tRNAs and the frequency of various codons influence the rate of polypeptide synthesis?
123. What is the minimal number of different tRNA-amino acid synthetases in a cell?
124. In a prokaryote, how might transcription and translation physically interact (provide a drawing)?

Question to ponder:

- How might a ribosome shift its reading frame while translating an mRNA? what would be effect of such a shift?

- *A ribosome can make a mistakes in amino acid incorporation or polypeptide termination; what would be value to a cell if a translating ribosome could recognize and correct its mistakes? What would be the cost(s) to the cell?*

References

[334] Known as the Shine-Delgarno sequence for its discovers

[335] Hidden coding potential of eukaryotic genomes: nonAUG started ORFs: http://www.ncbi.nlm.nih.gov/pubmed/22804099

[336]

[337] Interested in learning more, check out eukaryotic translation termination factor 1

[338] Molecular bumper cars (RNA polymerase-ribosomal interactions

This page titled 8.5: The translation (polypeptide synthesis) cycle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.6: Effects of point mutations on polypeptides and proteins

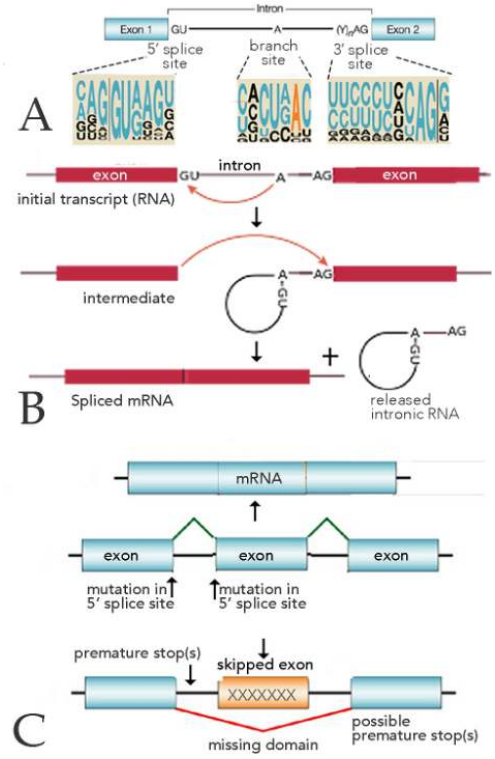
Mutations in a gene's regulatory region can alter the gene's expression by regulating the frequency of transcription. Mutations that change non-coding regions of an mRNA can influence its half-life (stability) and/or interactions with the ribosome, while mutations in a gene's coding region can, but generally do not influence transcription rate, unless of course regulatory regions are located within the coding region. Mutations can influence the sequence of the encoded polypeptide. We define three coding region mutations, involving a single base pair change as synonymous, mis-sense, or non-sense mutations. Because of the semi-redundant nature of the genetic code, it is possible that a single nucleotide change in a coding region will have no effect on the amino acid encoded – this is referred to as a synonymous mutation. That said, when different codons for the same amino acid are recognized by different tRNAs, these tRNAs may be present at different concentrations in the cell. The efficiency of translation is influenced by the rate of aa-tRNA binding. Different organisms can differ in the codons they use to encode particular amino acids, a fact that leads to what is known as "codon bias". Codon bias influences the efficiency of mRNA translation. Ribosome "stalling" can occur if the particular tRNA needed is present at low concentration. When genetically engineering the synthesis of a mRNA from one organism in another, translational efficiency can be increased by altering the gene that encodes the mRNA to reflect the codon bias of the host, rather than the codon bias of the donor.[339]

A second possibility is that the change of a single nucleotide will change the encoded amino acid; this is known as a mis-sense mutation. The effect of a mis-sense mutation depends upon where in the polypeptide it occurs and which amino acid is substituted. We can compare homologous polypeptides found in various organisms. Regions that are similar in terms of amino acid sequence and structure are referred to as conserved regions, compared to regions that are more variable, known (happily) as variable regions.[340] A mis-sense mutation that replaces an amino acid in a conserved region of a polypeptide is likely to have a more drastic effect on the polypeptide's function than a similar change in a variable region. Similarly, a mutation that replaces a large hydrophobic amino acid with a highly hydrophilic amino acid is more likely to perturb polypeptide structure and function than replacing a large hydrophobic amino acid with a smaller one. Finally, a single nucleotide mutation can replace a codon that specifies an amino acid with a stop codon; this is known as a non-sense mutation. The result of a non-sense mutation is a truncated polypeptide. As a first guess, the effect of a non-sense mutation will be more severe the closer it is to the beginning of the coding region, compared to its effect near the end of the coding region – although other factors contribute to any particular mutation's effect.

Mutations that influence splicing

While ignoring many details, a final class of point mutations are worth noting; these influence the "splicing" of a newly synthesized RNA molecule. Eukaryotic genes are generally broken up into coding regions, known as exons, and interspersed non-coding regions, known as introns. When a polypeptide-encoding gene is expressed, the initially synthesized RNA molecule contains both introns and exons. But ribosomes cannot distinguish between exon and intron sequences, probably one reason that prokaryotes do not have introns. In eukaryotes, introns are removed before the mature mRNA is exported across the nuclear envelope and into the cytoplasm (coming up). So the obvious question is, how exactly are introns recognized and removed, what molecular machines are used? As you may have guessed, there must be information present in the sequence of the newly synthesized RNA that identifies the intronic regions to be removed and the exons to be linked: where an exon ends and an intron starts, known as the 5' splice site and where the intron ends and the next exon starts, the 3' splice site. Finally, there is information within the intron known as the branch site (part A→). We can visualize this information through what are known as "sequence logo" plot.[341] Such a plot indicates the information associated within a sequence; where there is no preference, that is, where any of the four nucleotides are acceptable, the information present at that site is 0. Where either of two nucleotides are acceptable, the information is 1, and where only one particular nucleotide is acceptable, the information content is 2.

Splicing is carried out by a polypeptide-RNA complex known as the spliceosome. The spliceosome can recognize intron-exon boundary sequences and, using endonuclease and ligase activities, can cut out the intron and join the 3' end of one exon to the 5' of the next (B→), releasing the intron sequence in a looped form. Point mutations that disrupt the normal intron-exon boundary sequences (C→) can inhibit splicing, so that the intron remains in the final mRNA. Since introns do not encode polypeptides, there is little or no selection against the presence of stop codons in their sequence. A ribosome reading along a non-spliced RNA will likely add a series of inappropriate amino acids to the growing polypeptide before it encounters a stop codon, leading to the termination of polypeptide synthesis. Alternatively if, for example, the 3' splice site is disabled, a "down-stream" exon may be used for splicing. The result can be that an exon normally included is lost from the spliced mRNA; the polypeptide sequence it encodes



will be missing from the final polypeptide. The down-stream reading frame may also be altered, leading to the synthesis of incorrect amino acid sequences and the creation of stop codons. The result is that mutations that disrupt splicing can often have complex effects, and such mutations (alleles) have been associated with a number of human diseases.[342]

The complexity of eukaryotic genomes is greatly increased by the fact that most genes contain multiple exons and introns. Different sets of exons can be spliced together, a regulateable process known as alternative splicing, in different cells and within a single cell to produce mRNA molecules that encode variants of the "same" gene with different activities. These processes can lead to a range of complex behaviors that can muddy the interpretation of experimental manipulations.[343]

Insertions and deletions

Another class of mutations involves the insertion or deletion of nucleotide regions. Such insertions or deletions (known generically as indels) can disrupt or alter the binding of proteins to a gene's regulatory region, influencing gene expression. If they occur within the coding region, they can alter both the polypeptide sequence and reading frame so that the sequence of the polypeptide downstream of the insertion site will be altered. If the reading frame is preserved, the final polypeptide may have a region of new or deleted amino acids, but retain the original sequence downstream of the insertion/deletion.

Questions to answer:

125. What do the terms "up-stream" and "down-stream" mean in terms of gene structure.
126. What effects on polypeptide synthesis arise from neglecting codon bias?
127. Why doesn't release factor cause the premature termination of translation at non-stop codons?
128. What might happen if a ribosome starts translating an mRNA at the "wrong" place?
129. When analyzing the effects of a particular non- or mis-sense mutation, what factors would you consider first?

Question to ponder:

- How would you go about reengineering an organism to incorporate non-biological amino acids into its proteins.

References

[339] To learn more look at Codon Bias as a Means to Fine-Tune Gene Expression.

[340] A polypeptide assumes a 3D-dimensional that shape can be conserved.

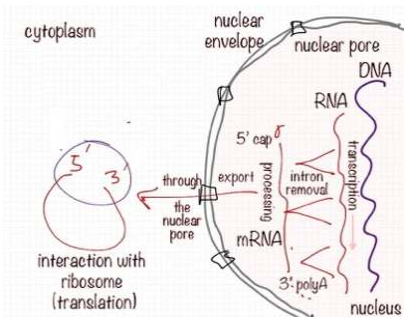
[341] Sequence logos: a new way to display consensus sequences: http://www.ncbi.nlm.nih.gov/pubmed/2172928

[342] The pathobiology of splicing: https://www.ncbi.nlm.nih.gov/pubmed/19918805

[343] See Biological plasticity rescues target activity in CRISPR knock outs

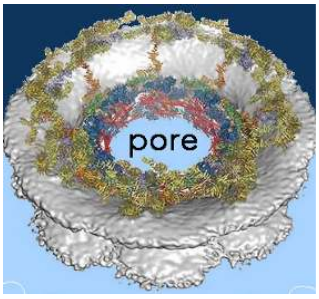
This page titled 8.6: Effects of point mutations on polypeptides and proteins is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.7: Getting more complex- gene regulation in eukaryotes



We will briefly reiterate a few points on how gene expression and polypeptide synthesis differ between prokaryotes and eukaryotes. The first and most obvious difference in gene expression between prokaryotes and eukaryotes (us) is the presence of a nucleus, a distinct domain within the eukaryotic cell that separates the bulk of the cell's genetic material, its DNA, from the cytoplasm, where the ribosomes are located (→). Aside from those within mitochondria and chloroplasts, the DNA molecules of eukaryotic cells are located within the nucleus. The barrier between nuclear interior and cytoplasm is known as the nuclear envelope. No similar barrier exists between DNA and ribosomes in prokaryotes. In both bacteria and archaea the DNA is in direct contact with the cytoplasm. In eukaryotes, a newly synthesized mRNA molecule undergoes a number of modifications before it comes into contact with the cytoplasm - it can be "spliced" and modified (processed) at both its 5' and 3' ends. Only after RNA processing has occurred will the "mature" mRNA be exported out of the nucleus , through a nuclear pore, and into the cytoplasm, where it can interact with ribosomes. Prokaryotic mRNAs are generally not processed.

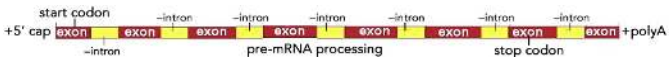
The nuclear envelope complex, considered in greater detail in cell biology courses, consists of two lipid bilayer membranes punctuated by nuclear pores. These pores are macromolecular complexes (protein machines) of ~125,000,000 daltons. Molecules



of less than ~40,000 daltons can generally move passively through a nuclear pore; larger molecules must be actively transported through a process coupled to the a thermodynamically favorable reaction, in this case the hydrolysis of guanosine triphosphate (GTP). The movement of larger molecules into and out of the nucleus through nuclear pores is regulated by what are known as nuclear localization and nuclear export sequences, located within polypeptides. These sequences are recognized by proteins (receptors) associated with the pore complex (→). A protein with an active nuclear localization sequence (NLS) will be found primarily in the nucleus while a protein with an active nuclear exclusion sequence (NES) will be found primarily in the cytoplasm. By controlling NLS and NES activities, a protein can come to accumulate, in a regulated manner, in either the nucleus or the cytoplasm, or can be found in both cellular regions. As we will see later on, the nuclear envelope breaks down during cell division (mitosis) in many but not all eukaryotes. Tares in the nuclear envelope have been found to occur when migrating cells try to squeeze through small openings.[344] Once the integrity of the nuclear envelope is re-established. Proteins with NLS and NES sequences moved back to their appropriate locations through active, that is energy dependent, coupled reaction-based transport processes.

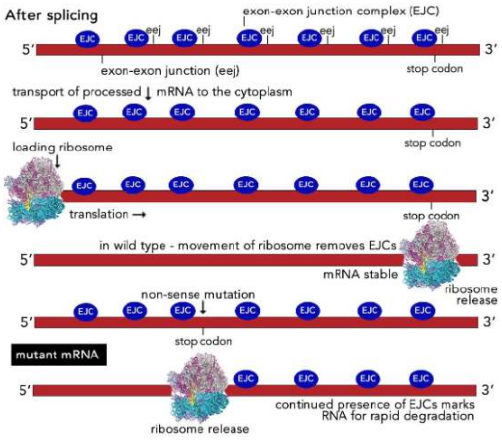
Nonsense mediated RNA decay

An originally surprising observation was that the truncated polypeptide generated by a non-sense mutation could produce phenotypic effects that were more severe than those associated with the failure to produce any polypeptide at all. To protect against the negative effects of non-sense mutations, particularly those that occur well "up-stream" of the normal stop codon, eukaryotic



organisms have a defense mechanism known as non-sense mediated decay. In a typical gene, the "normal" stop codon is generally within an exon located near the 3' end of the newly synthesized "pre-mRNA". During mRNA processing, introns are recognized and removed by the splicing system (↓); the 5' end is"capped" and the 3' end processed and (generally) a stretch of A nucleotides, a "polyA tail", is added. Typically, all of these modifications are completed before the processed transcript, now an mRNA, is transported through the nuclear pore complex into the cytoplasm.

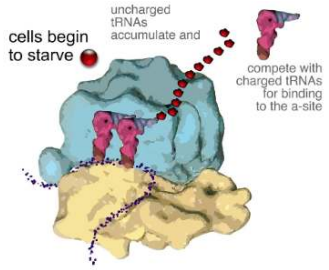
The removal of an intron leads to the formation of an exon-exon junction and the association of an exon-exon junction protein complex (EJC) "upstream" of each exon-exon junction (↓).[345] When a ribosome engages with the 5' end of the mRNA and moves down the mRNA during translation it displaces the EJCs, so when the first ribosome reaches the end of the mRNA's coding region all of the EJCs have been removed. The stability of the EJC-free mRNA is regulated by signals located primarily in its 5' and 3' untranslated regions.



The situation is different when a non-sense mutation generates a stop codon within an upstream exon ("mutant mRNA" →). The ribosome engages with the mRNA and continues until it reaches this stop codon, upon which release factor binds and the ribosome disengages. The EJCs downstream of the mutation-generated stop codon remain associated with the mRNA. The failure to remove the EJCs marks the mRNA as aberrant and triggers the non-sense mediated decay response.[346] Non-sense mediated decay activation leads to the degradation of mRNAs containing out-of-context non-sense codons and dramatically reduces the synthesis of potentially toxic polypeptides. In a further weird twist, it has been reported that RNA fragments generated from the degraded mRNA can re-enter the nucleus and regulate other genes - which further complicate the already complicated molecular relationship between mutation, genotype, and phenotype.[347]

Alarm generation: another secondary effect of disrupting gene expression

The translation system is a major consumer of energy within the cell.[348] When a cell is starving, it does not have the energy to generate amino acid charged tRNAs. The result is that uncharged tRNAsaccumulate (→). Since uncharged tRNAs can fit into the amino-acyl-tRNA binding sites on the ribosome, their presence increases the probability of unproductive tRNA interactions with the mRNA-ribosome complex, a situation that can lead to the premature termination of translation. When this occurs the stalled ribosome generates a signal that can lead to adaptive changes in the cell that enable the cell to survive for long periods in a “dormant” state.[349]



Another response that can occur is a more social one. Some cells in the population can “sacrifice” themselves for their closely related neighbors, remember kin selection and inclusive fitness. By shutting down transcription and translation, an addiction module can undergo programmed cell death. That mechanism is based on a toxin and anti-toxin system that differ in their stability. Just as ribonucleases degrade mRNAs, proteases degrade proteins and polypeptides. How stable a protein/polypeptide is depends upon its structure, which we will consider soon, and the presence of proteases that degrade it. Interrupting protein synthesis leads to the rapid disappearance (turn-over) of the anti-toxin while the toxin persists, leading to cell death, which in turn leads to the release of the cell’s nutrients, nutrients that can be used by its neighbors, in part to maintain active gene expression and protein synthesis. Of course, sacrificing for ones neighbors makes evolutionary sense only if one has neighbors and those neighbors are close relatives.

Questions to answer:

130. A gene has many introns - provide a model for how it might encode multiple functionally distinct polypeptides.
131. How can a mutation in splice site sequence influence gene expression and protein function?
132. How does non-sense mediated decay (NMD) protect against potentially deleterious mutations (alleles)?
133. Why would a cell want to stop (rather than continue) polypeptide synthesis when it is starving?

Question to ponder:

- How might the presence of uncharged tRNAs lead to the termination of translation? what would be benefit to a cell?

References

[344] Tearing the nuclear envelope: http://www.sciencemag.org/news/2016/...r-nuclei-burst

[345] The exon junction complex as a node of post-transcriptional networks

[346] Mechanism and regulation of the nonsense-mediated decay pathway

[347] Wilkinson, M. F. (2019). Genetic paradox explained by nonsense,

[348] Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources

[349] Characterization of the Starvation-Survival Response of Staphylococcus aureus & Bacterial Ribosome Rescue Systems

This page titled 8.7: Getting more complex- gene regulation in eukaryotes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.8: Turning polypeptides into proteins

Early genetic studies on the effects of mutations led George Beadle (1903-1989) to put forward the one gene one protein model, a model that needs some revision.[350] A protein is a functional entity, typically composed of one or more polypeptides.[351] These polypeptides can be the same or different, that is encoded by different genes. While polypeptides are synthesize in a linear manner they fold into three dimensional objects. In a protein composed of multiple polypeptides, these polypeptides must find and interact with one another and assume a functional conformation, the protein's structure. When we think about how a polypeptide folds, we have to think about the directionality of synthesis, the environment that the newly synthesized polypeptide comes to inhabit, and how it interacts with itself and with other molecules present. In the case of a protein composed of multiple polypeptides, each is synthesized independently, so we have to consider how these polypeptides come to interact with one another and avoid "inappropriate" interactions.

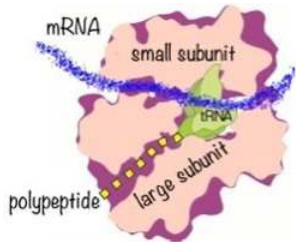
In the context of protein structure the terms primary, secondary, tertiary, and quaternary are often used (video link). The primary structure of a polypeptide is the sequence of amino acids along the polypeptide chain, written from its N- to its C- terminus. The polypeptide's secondary structure involves local folding motifs: α-heIical, β-sheet, and connecting domains. The polypeptide's tertiary structure is its overall three dimensional shape, which includes how its amino acid side-chains are oriented. Quaternary structure refers to how various polypeptides and co-factors are arranged to form a functional protein. In a protein that consists of a single polypeptide and no co-factors, tertiary and quaternary structures are the same; they may be different when a polypeptide is part of a protein. As a final complexity, a particular polypeptide can be part of a number of different proteins – the universe of proteins that a polypeptide is a part of could be considered another level of structure. Some of these interactions are relatively stable, others more ephemeral and regulative. This is one way in which a gene can play a role in a number of different processes and be involved in the generation of a number of different phenotypes.

Polypeptide synthesis (translation), like most all cellular processes, is stochastic, based on random collisions between molecules. In the specific case of translation, the association of the mRNA with ribosomal components occurs stochastically and involves a competition between different mRNAs for ribosome binding. This competition reflects relative mRNA concentration and their ribosome binding affinities. mRNAs of different gene differ in their translational efficiency. Similarly, the addition of a new amino acid to a growing polypeptide involves a productive collision between the appropriate amino acid-charged tRNA and the RNA-ribosome complex. In bacterial cells from 10 to 20 amino acids are added to the end of a growing polypeptide chain per second, the rate is about half that in mammalian cells.[352] This noisiness is rarely illustrated in presentations of polypeptide synthesis.

You may wonder whether there are errors in polypeptide synthesis similar to those in DNA or RNA synthesis. There are! Translation errors can lead to an in-frame stop codon that terminates translation and the release of an aberrant polypeptide that is (generally) rapidly degraded.[353] There are also cases that are "programmed" such that at certain positions along an mRNA the ribosome can "slip back" one nucleotide (a −1 frame shift) or skip forward one nucleotide (a +1 frame shift), leading to a different sequence of amino acids added from the point of the frame shift to the end of the polypeptide.[354] If the wrong amino acid is inserted at a particular position and it disrupts normal folding, the polypeptide may disrupt normal cellular functions.[355] In some cases the ribosome recognizes a mistake and corrects it.(see footnote 343) There are molecular machines that recognize mis-folded proteins and mark them for degradation. What limits the effects of mistakes made during translation is that most proteins (unlike DNA molecules) have finite and relatively short half-lives; that is, the time an average polypeptide exists before it is degraded. Normally this limits the damage that a mis-translated polypeptide can do to the cell and organism.

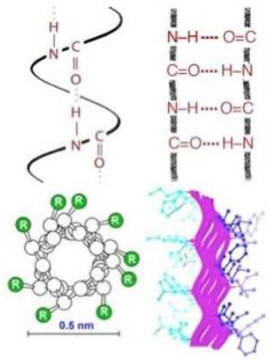
Factors influencing polypeptide folding and structure

Polypeptides are synthesized, and fold, in a directional manner. Both occur in the N- to C- direction after the newly synthesized polypeptide exits the ribosome's ~10 nm long and ~1.5 nm diameter tunnel. This tunnel is narrow enough to block the folding of the newly synthesized polypeptide. As the polypeptide emerges from the tunnel folding begins (video link)(→). At the same time it encounters the crowded cytoplasmic environment. The newly synthesized polypeptide needs to avoid low affinity, non-specific, and non-functional interactions with components that could interfere with normal folding.[356] If the polypeptide is part of a multi-subunit protein, once synthesis is complete it must "find" its correct partner(s), another diffusion-driven stochastic process. If the polypeptide does not fold correctly, it will not function correctly and may even damage the cell or the organism. A number of degenerative neurological disorders appear to be due, at least in part, to the accumulation of mis-folded polypeptides (see below).



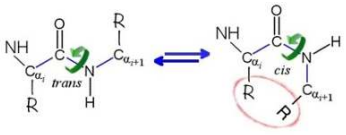
We can think of the folding process as a “drunken” walk across an energy landscape, with movements driven by intermolecular interactions and molecular collisions. The goal is to find the energy minimums of the system, generally assumed to be the native and functional state. That said, the native state is not necessarily static; folded polypeptide and the final protein will be subject to thermal fluctuations. It may move between states with similar but not identical stabilities.[357] Calculating the final folded state of a polypeptide is a complex computational problem. Multiple approaches have been used to characterize the structure of a protein. The structure of the protein can be determined directly by X-ray crystallography, cryo-electron microscopy, or nuclear magnetic resonance (NMR) spectroscopy, which we are not going to explain here, but which you may encounter in a biophysics class). Alternatively if the structure of a homologous (evolutionarily-related) protein is known, it can be used as a framework to model the structure of a previously unsolved protein. Finally, on-line tools, such as AlphaFold 3 use generative AI and are becoming increasingly accurate at their structure predictions (as confirmed by other methods).

Peptide bonds, H-bonds, bond rotation, proline and R-group effects

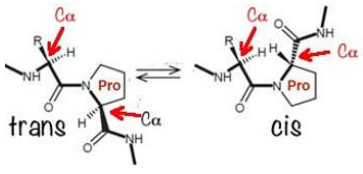


A number of constraints influence the folding of a polypeptide. The first is the peptide bond itself. Polypeptides consist of a string of peptide bonds. Not surprising then that there are common patterns in polypeptide folding. The first of these patterns to be recognized, the α-heIix (left →), was discovered by Linus Pauling (1901-1994) and Robert Corey (1897-1971) in 1951. This was followed by their description of the β-sheet (right →). The forces that drive the formation of α-helix and β-sheet will be familiar, they are the same forces that underlie water structure, namely H-bonding interactions.

In both an α-helix and a β-sheet, all of the possible H-bonds involving the peptide bond's donor and acceptor groups (–N–H and O=C– with “...” indicating a H-bond) are formed within the polypeptide (→). In an α-helix these H-bond interactions run parallel to the polypeptide chain. In a β-sheet these H-bonding interactions occur between polypeptide chains. The interacting strands within a β-sheet can run parallel or anti-parallel to one another; they can occur within a single polypeptide chain, folded back on itself or between different polypeptide chains. In an α-helix, the R-groups point outward from the helix axis. In β-sheets the R-groups point in an alternating manner either above or below the plane of the sheet. While all amino acids can take part in α-helix or β-sheetstructures, the imino acid proline cannot - the N-group coming off the α-carbon has no H. The presence of a proline in a polypeptide chain leads to a break in the pattern of intrachain H-bonds. It is worth noting that some polypeptides can adopt functionally different structures: for example one form of the prion protein PrPC contains a high level of α-helix (~42%) and essentially no β-sheet (~3%), while in an alternative form, PrPSc, associated with the disease scrapie, contains high levels of β-sheet (~43%) and ~30% α-helix.[358] The result is two very different 3-dimensional protein structures, even though the primary sequences of the two are identical.



Although typically drawn as a single bond, the peptide bond behaves more like a double bond, or rather a bond and a half. In the case of a single bond, there is free rotation around the bond axis in response to molecular collisions. In contrast, rotation around a peptide bond requires more energy to move from the trans to the cis configuration and back again (→). It is more difficult to rotate around the peptide bond because it involves the breakage of the "partial" bond. In addition, in the cis configuration the R groups of adjacent amino acids are on the same side of the polypeptide chain. If both R groups are large they can bump into each other and, if they get too close, repel each other. The result is that usually the polypeptide chain will tend to be in the trans arrangement. In both α-helix and β-sheet configurations, the peptide bonds are in the trans configuration because the cis configuration disrupts their regular organization.



Peptide bonds involving a proline residue have a different problem arising from the fact that proline is an imino rather than a amino acid. Its amino group is “locked” into a particular shape by the ring and therefore inherently destabilizes both α-helix and β-sheet structures. Peptide bonds involving prolines (←) are found in the cis configuration ~100 times as often as those between other amino acids. This cis configuration leads to a bend or kink in the polypeptide chain. The energy involved in the rotation around a peptide bond involving a proline is much higher than that of a

standard peptide bond; so high, in fact, that there are protein catalysts, peptidyl proline isomerases such as PIN1 (OMIM:601052) that facilitate the cis-trans rotation.

Hydrophobic R-groups

Many polypeptides and proteins are found primarily in the cytoplasm, an aqueous (water-based) environment. Yet, a number of amino acid R-groups are hydrophobic. Interactions between hydrophobic groups and water decrease the entropy of the system by the forced organization of water molecules around the hydrophobic group, a thermodynamically unfavorable situation. A typical polypeptide, with large hydrophobic R groups along its length will, in aqueous solution, tend to collapse onto itself so as to minimize, although not always completely eliminate, the interactions of its hydrophobic residues with water. This is very much like the effect that drives the assembly of bi-polar lipids into micelles and bilayers. In practice this means that the first step in the folding of many newly synthesized polypeptides, after they leave the ribosomal tunnel, is to collapse onto themselves so that the majority of their hydrophobic R groups are located internally, out of contact with water. In contrast, where there are no (or few) hydrophobic R groups in the polypeptide, the polypeptide will tend to adopt an extended configuration. If a protein comes to be embedded within a membrane then the hydrophobic R-groups will tend to be located on the surface of the folded polypeptide that interacts with the lipid bilayer's hydrophobic interior. Hopefully this makes sense, thermodynamically.

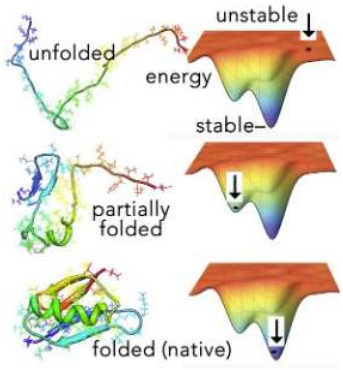
Acidic and basic R-groups

Some amino acid R-groups contain carboxylic acid or amino groups and so act as weak acids or bases. Depending on the pH of their environment these groups may be uncharged, positively, or negatively charged. Whether these groups are charged or uncharged can influence the overall structure, and therefore the activity, of a protein. By regulating pH in specific cellular compartments, cells and organisms can modulate protein activity. There are compartments within eukaryotic cells that are maintained at low pH in part to influence protein structure and activity. As an example, the internal regions of the vesicles associated with endocytosis become acidic through the ATP-dependent pumping of H^+ ions across their membranes. This in turn activates a number of enzymes located within them. Some of these proteins mediate the hydrolytic breakdown of ingested compounds.

Subunits and prosthetic groups

Many proteins contain non-amino acid-based components, known as co-factors. A protein without its cofactors is known as an apoprotein, with its cofactors it is known as a holoprotein. The apoprotein form is generally inactive and often unstable. Cofactors can range in complexity from a single metal ion to complex molecules, such as vitamin B12. The retinal group of bacteriorhodopsin and the heme group (with its central iron ion) are co-factors. In general, co-factors are synthesized by various enzymatic pathways, and so dependent on the activities of a number of genes. A functional protein can therefor be the direct product of a single gene, many genes, or (indirectly) entire metabolic pathways. Some cofactors cannot be synthesized and must be ingested.

Chaperones



The path to the native, that is, stable, functional state of a polypeptide or protein is not necessarily a smooth or predetermined one. A folding polypeptide can get "stuck" in a local energy minimum, with not enough environmental energy for it to get out again (→). Because a misfolded polypeptide or protein may be toxic, and certain not useful, active mechanisms exist to recognize it, unfold it, and let the folding proceed again are present. The process of unfolding misfolded polypeptides is carried out by proteins known as chaperones; we call them folding/re-folding chaperones to distinguish them from other types of chaperones. Given that the process of unfolding a misfolded protein is thermodynamically unfavorable it depends upon coupling to a favorable reaction, such as ATP hydrolysis. Once unfolded, the polypeptide has a second (or third or …) chance to fold correctly. A more bio-economical process than resynthesizing it. The “simple” eukaryote, the yeast *Saccharomyces cerevisiae*, has at least 63 distinct molecular chaperones.[359]

Chaperones do not determine the native state of a polypeptide–that is a function of the polypeptide's primary amino acid sequence. Rather, they suppress the formation of misfolded alternative structures. Consider, for example, the effect of a mis-sense mutation. Such a mutation can change the pattern of folding of a polypeptide; it may get caught more frequently in a mis-folded form. A folding/refolding chaperone can recognize a mis-folded polypeptide, unfold it, either totally or partially, and release it to fold again, enabling the polypeptide to reach a functional structure, even in the presence of a destabilizing mutation. Chaperones can serve as a buffer against mutational effects.

If most proteins are composed of multiple polypeptides but polypeptides are synthesized individually, how do polypeptides come to be correctly assembled into functional proteins in a cytoplasm crowded with other molecules? Protein assembly often involves specific chaperones that interact with specific polypeptides as they are synthesized; they act to reverse unproductive folding and inter-molecular interactions. Assembly chaperones can stabilize folding or hold them until they interact with correct partners to form the final, functional protein.[360] When proteins are synthesized *in vitro*, the absence of appropriate chaperones can make it difficult to produce multi-subunit functional proteins.

Another class of chaperones are known as "heat shock proteins." The genes that encode these proteins are expressed in response to increased, potentially lethal temperatures or other cellular stressors. At higher temperatures collisions with surrounding molecules can lead a protein to unfold, misfold, or aggregate with other proteins. The protein is said to be "denatured". Once expressed, heat shock proteins recognize denatured polypeptides, couple ATP hydrolysis reactions to unfold them, and then release the unfolded protein, giving them a chance to refold correctly.

Heat shock proteins help an organism adapt.[361] In classic experiments, when bacteria were briefly exposed to a temperatures sufficient to activate the expression of the genes that encode heat shock proteins, they had a higher survival rate at elevated temperatures compared to those grown continuously at lower temperature. Heat shock response-mediated survival is an example of the ability of an organism to adapt to its environment - it is a physiological response. The presence of the heat shock system itself, however, is a selectable trait, encouraged by temperature variation in the environment. It is the result of evolutionary factors.

How do chaperones recognize unfolded or abnormally folded proteins? In the case of a water soluble protein, while hydrophobic R-groups are normally located in the protein's interior, in the denatured (unfolded) protein they may be found exposed on its surface. The presence of these surface hydrophobic residues will favor aggregation; interacting hydrophobic regions will minimize hydrophobic-water interactions. Chaperones for water-soluble proteins recognize and interact with surface hydrophobic regions. For assembly chaperones, we can expect that specific sequences or structures in the target protein are recognized, which presumably is one reason that there are so many chaperone-like proteins, and specific chaperones for specific polypeptides and proteins.

Questions to answer

134. Why does it matter that rotation around a peptide bond is constrained?
135. How can changing the pH of a solution alter a protein's structure and activity?
136. How might the presence of a folding/refolding-chaperone mitigate the effects of a mis-sense mutation?
137. How do assembly-chaperones facilitate the assembly of multi-polypeptide proteins?
138. How does entropy drive protein folding & assembly? How do molecular collisions destabilize protein structure?

Questions to ponder

- How might surface hydrophobic R-groups facilitate protein-protein interactions.

- Develop a plausible model for how the expression of heat shock genes is regulated in response to temperature.

References

[350] One gene one protein & One gene one enzyme + When is a gene product a protein when is it a polypeptide?

[351] see also: When is a gene product a protein when is it a polypeptide?

[352] see http://bionumbers.hms.harvard.edu/default.aspx

[353] Quality control by the ribosome following peptide bond formation

[354] Ketteler 2012. On programmed ribosomal frameshifting: the alternative proteomes

[355] The evolutionary consequences of erroneous protein synthesis

[356] Remember, all molecules interact with each other via LDF-mediate interactions.

[357] folding video: from YOUTUBE - Stoneybrook: https://youtu.be/YANAso8Jxrk

[358] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC47901/ and prion disease: https://en.wikipedia.org/wiki/Prion

[359] An atlas of chaperone–protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways

[360] Assembly chaperones: a perspective

[361] The heat shock response: life on the verge of death

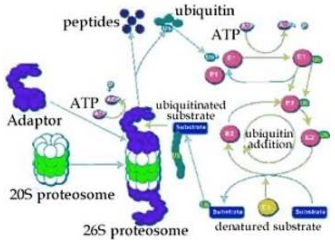
This page titled 8.8: Turning polypeptides into proteins is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.9: Regulating protein activity, concentrations, and stability (half-life)

Proteins act through their interactions with other molecules. Catalytic proteins (enzymes) interact with substrate molecules; these interactions lower the activation energy of the reaction's rate limiting step, resulting in an increase in the overall reaction rate. At the same time, cells and organisms are not static. They must regulate which proteins they produce, the final concentrations of those proteins, how active those proteins are, and where those proteins are located. It is primarily by altering proteins that in turn influence gene expression, that cells and organisms adapt to changes in their environment.

A protein's activity can be regulated in a number of ways. The first and most obvious is to control the total number of protein molecules present within the system. Let us assume that once synthesized a protein is fully active. With this assumption, the total concentration of a protein, and the total protein activity in a system [P_{sys}] is proportional to the rate of that protein's synthesis (dSynthesis/dt) minus the rate of that protein's degradation (dDegradation/dt), with dt indicating per unit time. The combination of these two processes, synthesis and degradation, determines the protein's concentration in the cell. Both the rate of protein's synthesis and degradation can be regulated. These processes can influence the rate at which a cell (or organism) can respond to various perturbations.

The degradation of proteins is mediated by a special class of enzymes known as proteases. Proteases cleave peptide bonds via hydrolysis (adding water) reactions. Proteases that cleave a polypeptide chain internally are known as endoproteases - they generate two polypeptides. Those that hydrolyze polypeptides from one end or the other, generally release one or two amino acids at a time, and are known as exoproteases. Proteases can also recognize and remove specific parts of a protein in order to activate or inactivate it, or to control where it is found in a cell. For example, nuclear proteins become localized to or excluded from the nucleus because they contain a NLS or an NES (see above). For these sequences to work they have to interact with the transport machinery associated with nuclear pores; but a protein may be folded so that its NLS/NES sequences are hidden, that is not visible on its surface. Changes in a protein's structure can reveal or hide such sequences, thereby altering the protein's distribution within the cell and therefore its activity. As an example, a transcription factor located in the cytoplasm is, in terms of its effects on gene expression, inactive; it becomes active if it enters the nucleus. Similarly, many proteins are originally synthesized in a longer and inactive "pro-form". When the pro-peptide is removed, cut away by an endoprotease, the processed protein becomes active. Proteolytic processing itself can be regulated.



As with all molecular level processes, the final breakdown of a protein is stochastic, based on collisions. Degradation can be regulated, signals within or added to a molecule can influence whether a collision with a degrading complex will be productive, that is, whether the molecule will be broken down. Protein degradation is particularly important for controlling the levels of "regulated" proteins, whose concentration within the cell may lead to unwanted effects. The rate of molecular degradation can be regulated, generally through the presence or addition of a signal that serves to influence the outcome of collisions with the degradative machinery (→). Degradation is an active and highly regulated process, involving ATP hydrolysis and multi-subunit complexes. One of these, involved in protein degradation, is known as the proteosome. The proteosome degrades the polypeptide into small peptides and amino acids that can be reused. As a mechanism for regulating protein activity, however, degradation has a serious drawback, it is irreversible.

This page titled 8.9: Regulating protein activity, concentrations, and stability (half-life) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.10: Allosteric and post-translational regulation

Allosteric regulation is a reversible way to control a protein's activity; a regulatory molecule binds to the protein altering its structure, its activity, its location within the cell, and/or its stability. When an allosteric effector binds to a protein, it interacts through van der Waals interactions - it is not covalently bound to the protein. Such interactions are reversible, influenced by thermal factors. Allosteric regulators can act either positively or negatively. The nature of such factors is broad, they can be a small molecule or another protein. What is important is that the allosteric binding site is distinct from the enzyme's catalytic site. In fact allosteric means “other site”. Because allosteric regulators do not bind to the same site on the protein as the substrate, changing substrate concentration generally does not alter their effects. That said, binding of substrate could, in theory, change the affinity of the protein for its allosteric effectors, just as binding of the allosteric effector can change the binding affinity of the protein for the substrate.

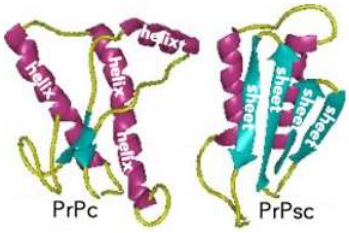
There are other types of regulation as well. A molecule may bind to and block the active site of an enzyme. If this binding is reversible, then increasing the amount of substrate can over-come the inhibition. An inhibitor of this type is known as a competitive inhibitor. In other cases, if the inhibitor reacts with the enzyme, forming a covalent bond. This type of inhibitor is essentially irreversible; increasing substrate concentration does not overcome inhibition. These are therefore known as non-competitive inhibitors.

Proteins may be modified after their synthesis, folding, and assembly - this process is known as post-translational modification. A number of different types of post-translational modifications occur within cells. Here we consider post-translational modification only generically. In general they involve the formation of a covalent bond linking a specific chemical group to specific amino acid side chains in the protein - these groups can range from a phosphate group (phosphorylation), an acetate group (acetylation), the attachment of lipid/hydrophobic groups (lipid modification), carbohydrates (glycosylation) and others. In general where a protein can be modified that modification can be reversed, except, of course, when the modification involves protein degradation or proteolytic processing. One type of enzyme catalyzes the addition of the modifying group while another catalyzes its removal. For example, proteins are phosphorylated by enzymes known as protein kinases, while protein phosphotases remove phosphate groups. Post-translational modifications act in much the same way as do allosteric effectors, they modify the structure and, in turn, the activity of the polypeptide or protein. They can also modify a protein’s interactions with other proteins, the protein's localization within the cell, and its stability.

This page titled 8.10: Allosteric and post-translational regulation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.11: Diseases of folding and misfolding

If a functional protein is in its native (or natural) state, a dysfunctional mis-folded protein is said to be denatured. It may not take much of a perturbation to unfold or denature a protein. In fact, under normal conditions, proteins often become partially denatured spontaneously, these are either refolded, often with the help of chaperones, or degraded through the action of proteases. A number of diseases, however, arise from irreversible protein mis-folding. Kuru was among the first of these protein mis-folding diseases to be identified. Beginning in the 1950s, D. Carleton Gadjusek (1923–2008)[362] studied a neurological disorder common among the Fore people of New Guinea. The symptoms of kuru, which means "trembling with fear", are similar to those of scrapie, a disease of sheep, and variant Creutzfeld-Jakob disease (vCJD) in humans. Among the Fore people, Kuru was linked to the ritual eating of the dead. Since this practice has ended the disease has disappeared. The cause of kuru, scrapie, and vCJD appears to be the presence of an abnormal form of a normal protein, known as a prion. We can think of prions as a type of anti-chaperone. The idea of proteins as infectious agents was championed by Stan Prusiner (b. 1942), who was awarded the Nobel Prize in Medicine in 1997.[363]



The protein responsible for Kuru and Scrapie is encoded by the PRP gene (OMIM:176640). The resulting PrPc protein normally exists in a largely α-helical form. Its pathogenic form, PrPsc (the "sc" indicates scrapie) contains a high level of β-sheet (→). PrPsc has a pathogenic activity, it catalyzes the transformation of PrPc into PrPsc. Once introduced, PrPsc initiates a chain reaction leading to the accumulation of more and more PrPsc. As it accumulates PrPsc assembles into rod-shaped aggregates that appear to damage cells. In the central nervous system these aggregates lead to neuronal cell death, dysfunction, and severe neurological defects. There is no natural defense, since the protein responsible is a normal protein.

When the Fore ate brains they inadvertently introduced PrPsc protein into their bodies. Genetic studies indicate that early humans evolved resistance to prion diseases, suggesting that cannibalism might have been an important selective factor during human evolution. Since cannibalism is uncommon today, how does one get such diseases in the modern world? There are rare cases of iatrogenic transmission, that is, where the disease is caused by faulty medical practice, for example through the use of contaminated surgical instruments or when diseased tissue is used for transplantation. But where did people get the disease originally? Since the disease is caused by the formation of PrPsc, any event that leads to PrPsc formation might cause the disease. Normally, the formation of PrPsc from PrPc is rare. We all have PrPc but few of spontaneously develop Kuru-like symptoms. There are, however, mutations in the *PRP* gene that greatly increase the frequency of PrPc → PrPsc conversion. Such mutations may be inherited (genetic) or may occur during the life of an organism (sporadic). Fatal familial insomnia (FFI)(OMIM:600072) is due to the inheritance of a mutation in the *PRP* gene, a mutation that replaces the aspartic acid normally found at position 178 of the PrPc protein with an asparagine. When combined with a second mutation in the PRP gene at position 129, the FFI mutation leads to Creutzfeld-Jacob disease (CJD).[364] If one were to eat the brain of a person with FFI or CJD, one might well develop a prion disease.

So why do PrPsc aggregates accumulate? To cut a peptide bond, a protease must position the target peptide bond within its catalytically active site. If the target protein's peptide bonds do not fit into the active site, they cannot be cut. Because of their structure, PrPsc aggregates are highly resistant to proteolysis. They gradually accumulate over many years, a fact that may explain the late onset of PrP-based diseases.

Questions to answer

139. How is the post-translational modification of a protein analogous to allosteric regulation? how is it different?
140. Assuming that synthesis rate decreases by 50% what happens to steady state polypeptide concentration? What happens if degradation rate increases by 50%? Generate predictive graphs of these (and other) scenarios .
141. How is the proteolytic processing of a polypeptide like and unlike an allosteric effector or a post-translational modification.
142. Why do post-translational modifications (and their reversals) require energy?

Questions to ponder

- Why is a negative allosteric regulator not considered a "competitive" inhibitor?

- How might the concentration of an allosteric effector influence the activity of the target protein?

- How would a cell recover from the effects of exposure to an irreversible, non-competitive inhibitor?

- What are the advantages of allosteric & post-translational modification based regulation compared to protein degradation.

References

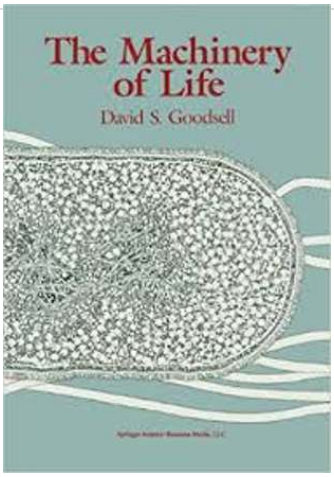
[362] Carleton Gajdusek: http://www.theguardian.com/science/2009/feb/25/carleton-gajdusek-obituary

[363] Stanley Prusiner: 'A Nobel prize doesn't wipe the skepticism away' & http://youtu.be/yzDQ8WgFB_U

[364] OMIM entry for Creutzfeld-Jacob diease: http://omim.org/entry/123400

This page titled 8.11: Diseases of folding and misfolding is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

8.12: Molecular machines



Drew Berry - video Molecular Machines

This page titled 8.12: Molecular machines is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

9: Genomes, Genes and Regulatory networks

9.0: Introduction

9.1: Locating information within DNA

9.2: Enhancers and transcription start and stop sites

9.3: Interaction networks and model systems

9.4: E. coli as a model system

9.5: Adaptive behavior and gene networks- the lac response

9.6: Types of regulatory interactions

9.7: Final thoughts on (molecular) noise

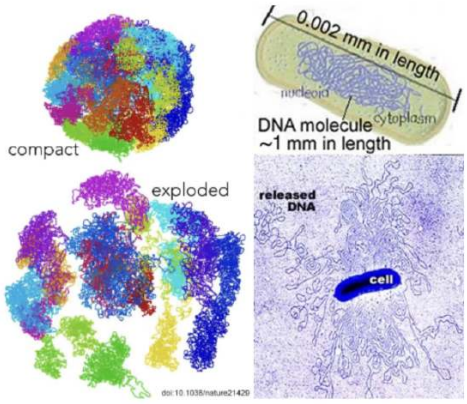
This page titled 9: Genomes, Genes and Regulatory networks is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

9.0: Introduction



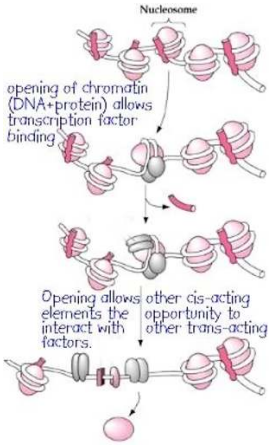
In which we consider how DNA molecules, and the genes that reside within them, are organized, how genes are recognized, and how their expression is controlled to form regulatory networks.

An important part of our approach to biology is to think concretely about the molecules we are considering. Nowhere is this more important than with DNA. DNA molecules are very long and cells, even the largest cells, are small. A typicalcylindrical bacterium (↓) is ~2 µm in length and ~1 µm in circumference. Based on the structure of DNA, each base pair is ~0.34 nm in length. A region of DNA that is 1000 (10^3) base pairs long is therefore ~0.34 µm in length. A bacterium, like *E. coli*, has ~3 x 10^6 base pairs of DNA – that's a DNA molecule almost a millimeter in length or about 500 times the length of the cell in which it finds itself, and of course there are two of them after DNA replication. That implies that at the very least the DNA has to be folded back on itself many times (←). A human cell has ~6000 times more DNA, resulting in a total length of greater than 2 meters of DNA per cell (before DNA replication). These DNA molecules have to fit into a nucleus that is typically ~10 µm in diameter. In both cases, the DNA has to be packaged in ways that allow it to fit within the cell or nucleus and still be accessible to the various proteins involved in the regulation of gene expression and DNA replication. To accomplish this, the DNA molecule is associated with specific proteins; the resulting DNA:protein complex is known as chromatin.



Left: A diagram of a bacterial cell showing its DNA molecule; disrupting the cell membrane (below) allows the DNA molecule within the cell to unfold.
Right: The (color coded) chromosomes within the nucleus of a human cell in compact & "exploded" views.

How DNA structure is regulated and how the information stored in DNA is used is the general topic of epigenetics (on top of genetics). Genetics refers to the information itself, encoded in the sequence of DNA molecules. A mutation will effect the sequence of DNA, it may or may not effect a gene, what a gene encodes and/or its expression. For a particular gene to be expressed, transcription factors must be able to find (by diffusion) and bind, through various intermolecular interactions, to specific nucleotide sequences within the gene's regulatory region(s). But the way the DNA is organized into chromatin, particularly in eukaryotic cells, can dramatically influence the ability of transcription factors to interact with and bind to their regulatory sequences. For example, if a gene's regulatory regions are inaccessible to protein binding because of the structure of the chromatin, the gene will be "off" (unexpressed) even if the transcription factors that would normally turn it on are present and active. As with essentially all



biological systems, the interactions between DNA and various proteins can be regulated. It is also worth remembering that there are typically one to two copies of a particular gene within a cell. so there may be stochastic effects to consider.

Different types of cells can often have their DNA organized differently; different chromosomal regions may be open(accessible) or closed (inaccessible)(→). Accessible means that proteins, and various molecular machines, can diffuse into and interact with specific regions of the DNA. Accessible, transcriptionally active regions of DNA are known as euchromatin. DNA packaged so as to be inaccessible to regulatory factors is known as heterochromatin. Given that only so many RNA polymerase complexes can move along a DNA molecule at a time; each ribosome assembles a single polypeptide as it moves along an mRNA molecule.

An example of the impact of gene copy number involves sex-determination. In most mammals, and humans, wild type females have two copies of the X chromosome, one from Mom and one from Dad.[365] Males are XY - with a maternal X and a paternal Y. The genes on the X and Y. chromosomes are different from one another. The X chromosome contains ~1,000 genes, the Y chromosome contains ~80 genes.[366] There are number of plausible mechanisms by which expression of genes on the X could be the same or similar in males and females. The one used in most mammals is the random (stochastic) transformation, early in embryonic development, of one of the two X chromosomes in each cell into a heterochromatic state, known as a Barr body. The genes on this chromosome are no longer expressed. X-activation is a inheritable event; the epigenetic changes found in the inactivated X persist through DNA replication and cell division. The X chromosome inactivation event is inherited vertically.[367] The result is that XX females are epigenetic mosaics composed of distinct clones of cells in which either one or the other of their X chromosomes have been inactivated. There is the possibility of evolutionary selection if the expression of one X chromosome leads to a reproductive advantage for one clone over the other. A particular tissue may end up preferentially expressing genes on the maternal or the paternal X chromosome.

An analogous process involving differential DNA (chromatin) accessibility is what is known as monoallelic gene expression, in which only one or the other of the two genes present in a diploid cell (on other chromosomes) is expressed. Monoallelic expression can lead to phenotypic differences between cells.[368] A question remains whether epigenetic states can be transmitted through the generation of sperm and egg and into the next generation.[369] Most epigenetic information appears to be reset during the process of gamete (sperm and egg) formation and embryonic development.

References

[365] Human Genome Project: Chromosome X: http://www.sanger.ac.uk/about/history/hgp/chrx.html

[366] The uncertainties in gene number are associated with identifying what exactly is a gene.

[367] X Chromosome: X Inactivation: http://www.nature.com/scitable/topicpage/x-chromosome-x-inactivation-323

[368] Monoallelic Gene Expression in Mammals - Chess, 2016

[369] Identification of genes preventing transgenerational transmission of stress-induced epigenetic states

This page titled 9.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

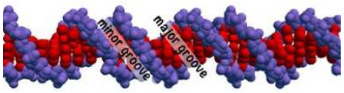
9.1: Locating information within DNA

For genes to be useful there needs to be mechanisms by which specific genes can be recognized and expressed at specific times, at specific levels, and in multicellular organisms, in specific cell types.[370] As noted, recognizing genes involves a two-component system. The first involves nucleotide sequences that provide a molecular address; this address identifies a specific region of a DNA molecule as well as the strand of the DNA to be used to direct RNA synthesis (transcribed). The second component are the proteins that recognize and specifically bind to such "regulatory" DNA sequences. The regulatory region of a gene can be simple and relatively short or long and complex. In some human genes, the regulatory region is spread over many thousands of base-pairs of DNA, located “up-stream” and/or "down-stream", within introns and coding regions.[371] The DNA within a chromosome can fold back on itself, allowing widely separated regions to interact.

Transcription factor proteins bind to gene regulatory sequences.[372] Multiple transcription factors and transcription factor binding sites can be involved in the regulation of a gene’s expression. In early genetic studies, two general types of mutations were found that influenced a gene's expression. “Cis” mutations are located within a gene’s regulatory region located near a gene’s coding region. “Trans” mutations mapped to other, more distant sites within the genome – often sites located on different chromosomes. Such trans mutations turned out to alter genes that encode transcription factors and other molecular components involved in gene expression. A transcription factor protein binds specifically (with high affinity) to sequences within the target gene’s regulatory region. A particular transcription factor can influence the expression of many hundreds of genes. Transcription factors can act either positively to recruit and activate DNA-dependent, RNA polymerase or negatively, to block polymerase binding and activation. Post-translational modifications and the binding of allosteric factors can alter a transcription factors activity, while interactions with other proteins can alter binding specificity and down-stream effects on gene expression.

Genes that efficiently recruit and activate RNA polymerase are highly expressed. Generally, high levels of mRNA lead to high levels of the encoded polypeptide. Mutations in a gene’s regulatory sequence (a cis mutation) will directly effect only that gene's expression, unless of course the gene encodes a transcription factor or its activity influences the regulatory circuitry of the cell. Genes are organized in interacting systems often associated with feed-back mechanisms involved in homeostatic, adaptive, and developmental processes. An experimental point is to determine whether the expression of a particular gene is directly or indirectly influence by a mutation or an environmental factor.

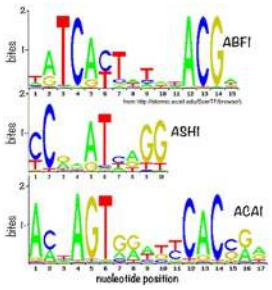
Transcription regulatory proteins recognize specific DNA sequences by interacting with the edges of base pairs accessible through



the major and/or minor grooves of the DNA helix (↓). There are anumber of different types of structurally distinct DNA binding domains. Transcription factor proteins can be grouped in structurally and evolutionarily related, families.[373] The binding affinity of a particular transcription factor to a particular regulatory sequence will be influenced by the DNA sequence as well as the binding of other proteins in the molecular neighborhood. We can compare affinities of different proteins for different binding sites by using an assay in which short DNA molecules containing particular nucleotide sequences are mixed in a 1:1 molar ratio, equal numbers of protein and DNA molecules:

$$DNA_{\text{sequence}} + \text{protein} \leftrightarrows DNA: \text{protein}.$$

After the binding reaction reaches equilibrium we measure the percentage of the DNA bound to the protein. If the protein binds with high affinity and on its own to a sequence the value will be close to 100%. If the transcription factor binds with low affinity to



the sequence, the ratio of bound to unbound protein will be close to 0%. Given DNA molecules of specific length and sequence and purified proteins, we can empirically determine the relative binding affinities of various proteins for particular sequences.[374] What we discover is a transcription factor does not recognize a single, unique nucleotide sequence, but rather has a range of affinities for related sequences. This binding preference is a characteristic of each transcription factor protein; it involves both the length of the DNA sequence recognized and the pattern of nucleotides within that sequence. A simple approach to this problem considers the binding information present at each nucleotide position as independent of all others in the binding sequence, which is not accurate but close enough for most situations. As noted before, the data is presented as a “sequence logo” that represents the amount of binding information at each position along the length of the binding site (→).[375] A value of 0 indicates no preference any of the four nucleotides, a value of 2 indicates that only a specific nucleotide is compatible with protein binding. The fewer the number of

nucleotides that are acceptable the more information is present. Different transcription factor proteins produce different preference plots. Mutations that influence a transcription factor's DNA binding site can have dramatically different effects; they can abolish site-specific DNA binding altogether or they may alter binding affinity and strength, leading to changes in the pattern of gene expression (addressed later on). Changes in the sequence recognized by a transcription factor can range from having little effect to completely abolishing transcription factor binding.

This is not to say that proteins cannot be functionally specific in their binding to nucleic acid sequences. There are classes of proteins, known as restriction endonucleases and site specific DNA modification enzymes (methylases and acetylases) that bind to unique nucleotide sequences. For example the restriction endonuclease EcoR1 binds to (and cleaves) the nucleotide sequence GAATTC; change any one of these bases and there is no significant binding and no cleavage of the sequence. The CRISPR-Cas9 system for genetic manipulation is also highly specific, using a 22 nucleotide RNA to target a double-stranded endonuclease to a specific site in the genome.[376] So the fact that the binding specificities of transcription factors are more flexible suggests that there is a reason for such flexibility, presumably involving differential regulation of a number of genes.

To be effective in recruiting a functional RNA polymerase complex to specific sites along a DNA molecule, the binding of a protein to a specific DNA sequence must be relatively long lasting. A common approach to achieving this outcome is for the transcription factor to be multivalent, that is, so that it can bind to multiple (typically two) sequence elements at the same time. This has the effect that if the transcription factor dissociates from one binding site, it can remains tethered to the other. Since the molecule is held by this binding close to the DNA it is more likely to rebind to its original site. In contrast, a protein with a single binding site is more likely to diffuse away before rebinding can occur.

It is worth noting that most transcription factor proteins bind weakly (with low affinity) to generic DNA sequences. Non-sequence specific binding is transient; rapidly broken by thermal motion. That said, since there are astronomical numbers of such non-sequence specific binding sites within a cell’s DNA, much of the time transcription factors are found transiently and non-specifically associated with DNA. A functionally meaningful behavior arising from the low affinity binding of proteins to DNA is their one-dimensional diffusion along the length of the DNA molecule.[377] Common, low energy, collisions are more likely to move the protein along rather than away from the DNA molecule. This enables a transcription factor protein weakly bound to DNA to move back and forth along the DNA molecule until it either interacts with, and binds to, a high affinity site or dissociates completely. This type of “facilitated target search” behavior can greatly reduce the time it takes for a protein to find a high affinity binding site among the millions of low affinity sites present in the genome.[378]

References

[370] As an aside, are many transcribed DNA sequences that do not appear to encode a polypeptide or regulatory RNAs. It is not clear whether this transcription is an error, due to molecular level noise or whether such RNAs play a physiological role..

[371] Regulatory regions located far from the gene’s transcribed region are known as enhancer elements.

[372] In prokaryotes transcription factors are often referred to as sigma (σ) factors.

[373] Determining the specificity of protein-DNA interactions: http://www.ncbi.nlm.nih.gov/pubmed/20877328

[374] Of course we are assuming that physiologically significant aspect of protein binding involves only the DNA, rather than DNA in the context of chromatin, and ignores the effects of other proteins, but it is a good initial assumption.

[375] Sequence logos: a new way to display consensus sequences: http://www.ncbi.nlm.nih.gov/pubmed/2172928

[376] The CRISPR-Cas9 system involves targeting a double-stranded DNA exonuclease to a specific site in a DNA sequence; it uses a RNA molecule to achieve very high levels of specificity. see CRISPR/Cas9 and Targeted Genome Editing

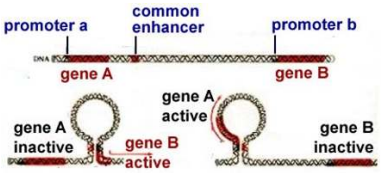
[377] As illustrated in the PhET applet:http://phet.colorado.edu/en/simulation/gene-expression-basics

[378] Physics of protein-DNA interactions: mechanisms of facilitated target search

This page titled 9.1: Locating information within DNA is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

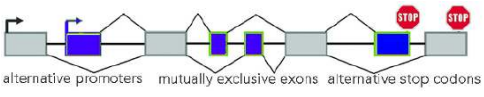
9.2: Enhancers and transcription start and stop sites

As the conditions in which an organism lives get more complex, the more dynamic gene expression needs to be. This is particularly the case in multicellular eukaryotes, where different cell types need to express different genes and / or different versions (splice variants) of genes. One approach is to use multiple different gene regulatory regions, regions that bind different sets of transcription factors. Regulatory factors not only bind to DNA, they interact with one another. We can imagine that the binding affinity of a particular transcription factor will be influenced by the presence of other transcription factors already bound to neighboring or overlapping sites on the DNA. Similarly the structure of a protein can change when it is bound to DNA, and such a change can lead to interactions with DNA:protein complexes located at more distant sites, known as enhancers. Such regulatory elements can form a multi-part regulatory systems.



Consider the following situation. Two genes share a common enhancer, depending upon which interaction occurs, gene A or gene B but not both could be active (→). The end result is that combinations of transcription factors are involved in turning on and off gene expression. In some cases, the same protein can act either positively or negatively, depending upon molecular context, that is, the specific gene regulatory sequences accessible, the other transcription factors present and their various post-translational modifications. Here it is worth noting that the organization of regulatory and coding sequences in DNA imposes directionality on the system. A transcription factor bound to DNA in one orientation or at one position may block the binding of other proteins (or RNA polymerase), while bound to another site it may stabilize protein (RNA polymerase) binding. Similarly, DNA binding proteins can interact with other proteins to control chromatin configurations that can facilitate or block accessibility to regulatory sequences. While it is common to see a particular transcription factor protein labelled as a transcriptional activator or repressor, in reality the activity of a protein reflects the specific gene under consideration, and its interactions with other factors, all of which can influence gene expression outcomes.

The exact position on the DNA where RNA polymerase starts transcribing an RNA molecule is known as the transcription start site.



Different regulatory sequences can lead to the use of different transcription start sites. In genes with introns, where transcription starts can determine which exons are included in the final mRNA molecule. Other factors influence splicing, and so determine which exons are included and which are excluded from the final mRNA (→). Where the RNA polymerase falls off the DNA, and so stops transcribing RNA, is known as the transcription termination site.

Once transcription initiates, the RNA polymerase moves down the DNA. As it clears the transcription start site region there is now room for another polymerase complex to associate with the DNA. Assuming that the factors associated with the regulatory region remain intact and active, the time to load a new polymerase on an existing regulatory complex can be faster than the time it takes to build up a new regulatory complex from scratch. The result is that transcription is often found to occur in bursts, a number of RNAs are synthesized from a gene in a short period of time followed by a period of transcriptional silence associated with the disassembly and subsequent reassembly of a new transcription start complex. A similar bursting behavior is observed in polypeptide synthesis. The onset of translation begins with the small ribosomal subunit interacting with the 5' end of the mRNA; the assembly of this initial complex involves a number of components, and takes time but once formed can persist. While this complex exists multiple ribosomes can interact with the mRNA, each synthesizing a polypeptide, leading to bursts, multiple translational events. Once the translation initiation complex dissociates, it takes time, more time than just colliding with another small ribosomal subunit, for a new complex to form. The combination of transcriptional and translational bursting contributes to noisy protein synthesis. Since cellular behavior can be influenced by changes in gene expression, these processes can lead to phenotypic differences between genetically identical cells.

Questions to answer:

143. How might a transcription factor determine which DNA strand will be transcribed?
144. A mutation inhibits the expression of a gene, how might you determine whether the mutation altered a transcription factor or the DNA sequences that regulate gene expression.
145. What factors are likely to influence the length of a gene's regulatory region?
146. How might you tell which X chromosome was inactivated in a particular cell of a female person?

Questions to ponder:

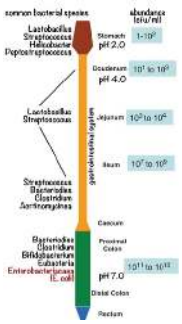
- What factors might drive the evolution of overlapping genes?
- How can over-lapping genes, or genes on different DNA strands influence each others' expression?
- How might you determine which allele is expressed in a cell displaying monoallelic gene expression?

This page titled 9.2: Enhancers and transcription start and stop sites is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

9.3: Interaction networks and model systems

Interaction networks are a universal feature of biological systems, from the molecular to the mental and social. These are organized in a hierarchical and bidirectional manner, involving various forms of "feedback". So what exactly does that mean? Most obviously, at the macroscopic level, the behavior of ecosystems depends upon the interactions between organisms. As we move down the size scale the behavior of individual organisms is based on the interactions between the cells and tissues formed during the process of embryonic development or the building of social communities. Gene expression also involves interaction networks; genes express polypeptides and regulatory RNAs that influence the expression of other genes, often including themselves. Multiple gene products are involved in the regulation of any particular gene. Since many of the interactions involved have a stochastic nature, chance plays a role in generating variation and outcomes. At the same time there are regulatory interactions and feedback loops that act to control stochastic effects and to ensure that biological behaviors are robust and (often) predictable. All of the interactions and processes that underlie biological systems are the result of evolutionary processes and historical situations, including past adaptations and non-adaptive events in ancestral populations.

Scientific studies of biological systems are driven by the desire to understand how it is that such systems came to be and how they behave the way they do. Such knowledge is helpful, particularly in the age of precision medicine and genetic engineering, in order to identify, treat and/or avoid a disease. But there are a number of reasons that some questions cannot be answered directly; one is that it may not be possible (or ethical) to carry out necessary experiments. But here the evolutionary relationships between organisms come to our aid. We can choose organisms that are easier to study, develop faster, or are “simpler” in some way. By studying various “model” organisms we can hope to identify common and relevant mechanisms. At the same time, it is important to recognize that the various “types” of experimentally useful organisms are each adapted to their own environmental niche, evolving independently of others for millions to hundreds of millions of years. Even the most closely related of organisms, such as the great apes, a group that includes humans, display multiple functionally significant differences. Once isolated, and maintained in the laboratory, we put organisms in an unnatural situation, a situation that subjects them to different adaptive and selective pressures. At the same time, isolated organisms are often maintained under conditions that reduce genetic variation - they become inbred. Inbreeding can be desirable (for science), since it reduces variability and makes experiments more interpretable; it also makes the organisms studied less relevant to “real” (wild) organisms. An example of this is the fact that different in-bred lines of mice can respond differently to different experimental and therapeutic interventions.[379]



Notwithstanding the complexities of biological systems, we can approach them at various levels of resolution through a systems perspective, using specific organisms to study specific processes andbehaviors. At each level, there are objects that interact with one another in various ways to produce specific behaviors. Many of these systems are conserved, related to one another evolutionarily. To analyze a system we need to define, identify, and appreciate the nature of the objects involved, how they interact, and the behaviors that emerge from such interactions; how such interactions influence the system. Does the system move to a new state or does it return, after a perturbation, to its original state? There are many ways to illustrate this way of thinking. We will get concrete by looking at a (relatively) simple system and consider how it behaves at the molecular, cellular, and social levels. The model system we consider here is the bacterium *Escherichia coli* (*E. coli*), how it behaves in isolation, in social groups, and how it metabolizes the milk sugar lactose.[380] Together these illustrate a number of common regulatory principles that apply more or less universally to biological systems at all levels of organization.

References

[379] Löscher, 2024. Of Mice and Men: The Inter-individualVariability of the Brain’s Response to Drugs

[380] The Lac Operon: A Short History of a Genetic Paradigm

This page titled 9.3: Interaction networks and model systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

9.4: E. coli as a model system

Every surface of your body harbors a flourishing microbial ecosystem. This is particularly true of yourgastrointestinal system that runs from your mouth and esophagus (with a branch leading to your nose), through the stomach, into the small and large intestine and the colon (→).[381] Each region supports its own unique microbial community, known as its microbiome. These environments differ in terms of a number of properties, including differences in pH and O_2 levels. Near the mouth and esophagus O_2 levels are high and microbes use aerobic (O_2 dependent) respiration to maximize the extraction of energy from food. Moving through the system O_2 levels decrease until anaerobic (without O_2) mechanisms are necessary. Microbes with different ecological preferences and adaptabilities are found at different positions along the length of the gastrointestinal track.[382]

One challenge associated with characterizing the complexity of the microbiome present at various locations is that the organisms present are often dependent upon one another for growth and survival. When isolated from one another and their normal environment they do not grow. Since the standard way to count bacteria is to dilute samples so that single bacteria land in isolation from one another on an agar plate surface, this can be a problem. When they grow and divide, they form macroscopic (visible) colonies. We count the number of "colony forming units" (CFUs) per original sample volume; this number provides a measure of the number of bacteria capable of growing and dividing. If an organism cannot form a colony under the assay conditions, it will appear to be absent from the population. Many bacteria are dependent on others and do not grow in isolation. Recent studies, however, have found ways to culture more of such organisms.[383] To avoid many issues, molecular methods use DNA sequence analyses to identify which organisms are present without having to grow them.[384] The result of these molecular analyses reveal, for the first time, the true complexity of the microbial ecosystems living on and within us.[385]

Much of the early work in molecular biology was carried out using a relatively minor component of this microbial community, *E. coli*, a member of the Enterobacteriaceae family of bacteria. *E. coli* is found in the colons of birds and mammals.[386] *E. coli* is what is known as a facultative aerobe, it can survive in both anaerobic and an aerobic environments. This flexibility, as well as *E. coli*'s generally non-fastidious nutrient requirements make it easy to grow in the laboratory. Moreover, the commonly used laboratory strain of *E. coli*, known as K12, does not cause disease in humans. That said, there are strains of *E. coli*, such as *E. coli* O157:H7, that are pathogenic (disease-causing). *E. coli* O157:H7 contains 1,387 genes that are not found in *E. coli* K12 and it is estimated that the two strains diverged from a common ancestor ~4 million years ago. What makes *E. coli* O157:H7 pathogenic? Sorry, that is beyond our scope here.[387]

References

[381] The gut microbiome: scourge, sentinel or spectator?

[382] The Gut Microbiome: Connecting Spatial Organization to Function and Gut biogeography of the bacterial microbiota

[383] See Lopez-Garcia & Moreira, (2020) Cultured Asgard Archaea Shed Light on Eukaryogenesis

[384] Application of sequence-based methods in human microbial ecology: http://www.ncbi.nlm.nih.gov/pubmed/16461883

[385] The human microbiome: our second genome

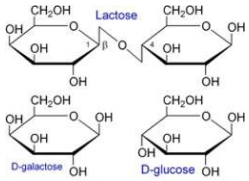
[386] Evolutionary ecology of E.coli

[387] Enterohemorrhagic E. coli (EHEC) pathogenesis: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3417627/

This page titled 9.4: E. coli as a model system is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

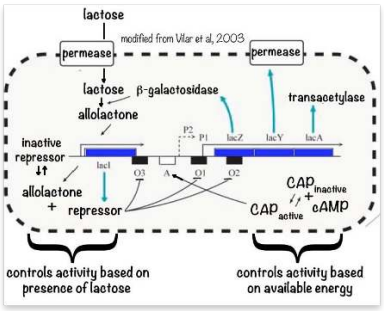
9.5: Adaptive behavior and gene networks- the lac response

Lactose is a disaccharide (a sugar) composed of D-galactose and D-glucose (←). It is synthesized, biologically, exclusively by female mammals. Mammals use lactose in milk as a source of calories (energy) for infants. One reason, it is thought, is that lactose is not easily digested by most microbes. The lactose synthesis system is derived from an evolutionary modification of an ancestral gene that encodes the enzyme lysozyme. Through a gene duplication event and subsequent mutations, a gene encoding the protein α-lactoalbumin was generated. α-lactoalbumin is expressed in mammary glands, where it forms a macromolecular complex with a ubiquitously expressed protein, galactosyltransferase, to form the protein lactose synthase.[388]



E. coli is capable of metabolizing lactose, but only when there are no better (easier) sugars to eat. If glucose or other compounds are present in the environment, the genes required to metabolize lactose are not expressed. Two genes are required for *E. coli* to metabolize lactose. The first encodes lactose permease. Lactose, being large and hydrophilic, cannot pass through *E. coli* cell membranes. Lactose permease is a membrane protein that allows lactose to enter the cell, moving down its concentration gradient. The second gene, enzyme β-galactosidase, catalyzes the reaction that splits lactose into D-galactose and D-glucose, both of which can be metabolized by proteins expressed constitutively, that is, all of the time. So how exactly does this system work? How are the lactose utilization genes turned off in the absence of lactose and turned on when lactose appears and energy is needed? How does the cell "know" when lactose is present in the environment? The answers illustrate general principles of the interaction networks controlling gene expression.

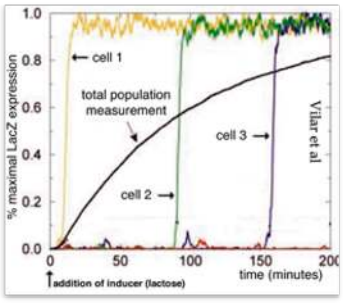
In *E. coli*, like many other bacteria, multiple genes are organized into what are known as operons. In an operon a single regulatory region controls the expression of multiple genes, often encoding proteins involved in the same metabolic pathway. A powerful approach to the study of genes is to look for mutations that abolish a specific process, and so produce a specific and discernible phenotype. As we said, wild type *E. coli* can grow on lactose as their sole energy source. To understand lactose utilization, we look for mutant *E. coli* that cannot grow on lactose.[389] We start by checking that the mutant *E. coli* can grow on glucose. Why? Because we are not really interested (in this case) in mutations that disrupt standard metabolism, such as the ability to use glucose. We seek to identify the genes and gene products involved in a specific process, lactose utilization. Such an analysis, by François Jacob (1920–2013) and Jacques Monod (1910–1976) revealed a number of distinct classes of mutations: some led to an inability to respond to the presence of lactose in the medium, others led to the de-repression, that is the constant expression of the genes involved in the ability to metabolize lactose, lactose permease and β-galactosidase. In such mutant strains both genes are expressed whether or not lactose is present.



By mapping where these mutations are in the genome (using the Hfr horizontal gene transfer system described in chapter 12) and a number of other experiments, the following model was generated (→). The genes encoding lactose permease (lacY) and β-galactosidase (lacZ) are part of the lac operon. The lac operon is regulated by two distinct factors. The first is the product of a constitutively active gene, lacI; the lacI-encoded polypeptide assembles into a tetrameric protein that acts as a transcriptional repressor. A typical cell contains ~10 lac repressor molecules and one or two copies of the lac operon. The lac repressor binds to sites in the lac operon's promoter. The binding of the repressor blocks the expression of the operon. The repressor's binding sites within the lac operon promoter appear to be its' only functionally significant binding sites in the entire *E. coli* genome. The second regulatory sequence element is known as the activator site. It can bind the cyclic AMP receptor protein (CRP) also known as catabolyte activator protein (CAP).[390] CAP/CRP is encoded by a gene located outside of the lac operon. CAP/CRP is a homodimer composed of two identical polypeptides. The DNA binding activity of CAP/CRP is regulated by the binding of an allosteric co-factor, cyclic adenosine monophosphate (cAMP). cAMP accumulates in the cell when nutrients, specifically free energy delivering nutrients like glucose, are low. An increase in cAMP concentration [cAMP] acts as a signal that the cell needs energy. In the absence of cAMP, CAP/CRP does not bind to or activate expression of the lac operon. In its presence, when energy is needed, CAP/CRP-cAMP is active, binds to a site in the lac operon promoter, and recruits and activates RNA polymerase, leading to the synthesis of lactose permease and β-galactosidase RNAs and proteins. Active CAP/CRP also acts to inhibit the expression of many other genes. Even if energy levels are low and [cAMP] is high, the lac operon

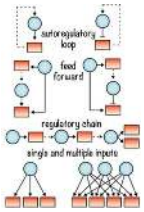
will remain inactive (not expressed) if lactose is absent because binding of lac repressor protein to sites (labeled O1, O2, and O3) in the lac operon's regulatory region blocks polymerase recruitment.

What happens when lactose appears in the environment? Initially nothing. Cells are expressing the lac repressor, blocking activation of the lac operon. In the absence of lac operon expression lactose permease is not present and lactose does not enter the cell. This conclusion is based on a model in which the system acts deterministically, but biological systems are not deterministic, they are stochastic, noisy and probabilistic. Given the small number, about 10, of lac repressor molecules per cell there is a small but significant, non-zero chance that periodically and stochastically, the repressor will release from the lac operon. If this occurs



when CAP/CRP is active, β-galactosidase and lactosepermease will be expressed independently of the presence of lactose. If lactose is present, there is a positive feedback loop (←).[391] Cells that have, by chance, expressed lacY (lactose permease) and lacZ (β-galactosidase) genes will respond. The permease will enable lactose to enter these cells. This lactose willbe converted to allolactone, in a reaction catalyzed by β-galactosidase.Allolactone binds to, and inhibits the lac repressor protein. Unrepressed, there is a increase (~1000 fold) in the rate of expression of the lacZ and lacY genes. Inaddition to generating allolactone from lactose, β-galactosidase catalyzes the hydrolysis of lactose into D-galactosidase and D-glucose, which are used to drive cellular metabolism. Through this process, the cell goes from essentially no to full expression of the lac operon, which enables it to metabolize lactose. At the same time, cells that did not (by chance) express the lactose operon are unable to metabolize lactose, even though lactose is present outside of those cells. So even though all of the *E. coli* cells present in a culture may be genetically identical, they express different phenotypes due to the stochastic nature of gene expression.[392] In the case of the lac system, over time the noisy nature of gene expression leads more and more cells to activate their copy of the lac operon. Also cells that can metabolize lactose have energy for growth. The offspring of cells inherit lactose permease and β-galactosidaseproteins, so will be able to use lactose. Once "on", the operon will be expressed as long as lactose is present, since allolactone binds to and inactivates the lac repressor protein.

What happens if (and when) lactose disappears from the environment? what determines how long it takes for the cells to return to the state in which they no longer express the lac operon? The answer is determined by the effects of cell division and regulatory processes. In the absence of lactose, the allolactone concentration falls and the lac repressor protein returns to its active (repressive)



state, inhibiting lac operon expression. No new lactose permease and β-galactosidase are synthesized and their concentrations will fall based on the rate of their dilution by growth and cell division and their degradation (proteolysis). In the absence of lactose, each cell division will reduce the concentration of the lactose permease and β-galactosidase by ~50%. As the proteins are diluted and degraded, the cells return to their initial state, that is, with the lac operon off and no copies of either lactose permease or β-galactosidase present.

References

[388] Molecular divergence of lysozymes and alpha-lactalbumin: http://www.ncbi.nlm.nih.gov/pubmed/9307874

[389] The basic experimental approach involves a technique known as replica plating

[390] Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing

[391] Modeling network dynamics: the lac operon, a case study

[392] An example of such behavior here: http://www.elowitz.caltech.edu/publications/Noise.pdf

This page titled 9.5: Adaptive behavior and gene networks- the lac response is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

9.6: Types of regulatory interactions

A comprehensive analysis of the interactions between 10^6 transcription factors and (many more) regulatory sequences in the baker's yeast *S. cerevisiae* revealed the presence of a number of common regulatory motifs.[393] These include (↓→):

• **Auto-regulatory loops**: A transcription factor binds to sequences that regulate its own transcription. Such interactions can be positive (amplifying) or negative (squelching).

• **Feed forward interactions**: A transcription factor regulates the expression of a second transcription factor; the two transcription factors then cooperate to regulate the expression of a third gene.

• **Regulatory chains**: A transcription factor binds to the regulatory sequences in another gene and induces expression of a second transcription factor, which in turn binds to regulatory sequences in a third gene, etc. The chain ends with the production of some non-transcription factor products.

• **Single and multiple input modules**: A transcription factor binds to sequences in a number of genes, regulating their coordinated expression. In most cases, sets of target genes are regulated by sets of transcription factors that bind in concert.

In each case the activity of a protein involved in an interaction network can, like the lac repressor, be regulated through interactions with other proteins, allosteric factors, and post-translational modifications. It is through such interactions that signals from inside and outside the cell can control patterns of gene expression leading to maintenance of the homeostatic state or various emergent adaptations.

References

[393] Transcriptional regulatory networks in *Saccharomyces cerevisiae*: http://www.ncbi.nlm.nih.gov/pubmed/12399584

This page titled 9.6: Types of regulatory interactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

9.7: Final thoughts on (molecular) noise

When we think about the stochastic behaviors of cells, we can identify obvious sources of molecular and cellular level noise. First, there are generally only one or two copies of a particular gene within a cell. The probability that those genes are accessible and able to recruit transcription factors, associated proteins, and RNA polymerase is determined by the frequency of productive collisions between regulatory sequences and relevant transcription factors together with their dissociation rates. Cells are small, and the numbers of different transcription factors can vary quite dramatically. Some transcription factors are present in high numbers (~250,000 per cell) while others (like the lac repressor) may be present in less than 10 copies per cell. The probability that particular molecular interactions occur will be a function of their relative concentrations, diffusion, binding and kinetic energies, and targets. This will influence the probability that a particular gene regulated by a particular transcription factor is active or not. Once on, transcriptional and translational bursting will produce gene products that can alter the state of the cell so that secondary, down-stream changes occur in gene expression and other cellular processes. These changes may, like the lac operon system be reversible once the stimulus (lactose) is removed or they may be more or less irreversible, as occurs during cellular differentiation and embryonic development.[394]

Questions to answer:

147. How might you design a regulatory network to produce a steady level of product?
148. How might you design a regulatory network that oscillates like a clock?

Question to ponder:

- Design a gene regulatory system that acts as an irreversible switch between states?



Short Chapter Summary

- Regulatory DNA + transcription factors + chromatin state = conditional, context-specific expression.
- Simple motifs (repressors, activators, feedback) combine into networks with emergent behavior.
- Noise is real and often functional; cells bet-hedge.
- Classic operons (e.g., lac) teach universal regulatory logic

References

[394] A single molecule view of gene expression: http://www.ncbi.nlm.nih.gov/pubmed/19819144

This page titled 9.7: Final thoughts on (molecular) noise is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

10: Social Systems

10.0: Introduction

10.1: Targeting membrane proteins to where they need to be

10.2: Nuclear targeting and exclusion

10.3: Intercellular signaling- signals, receptors and responses

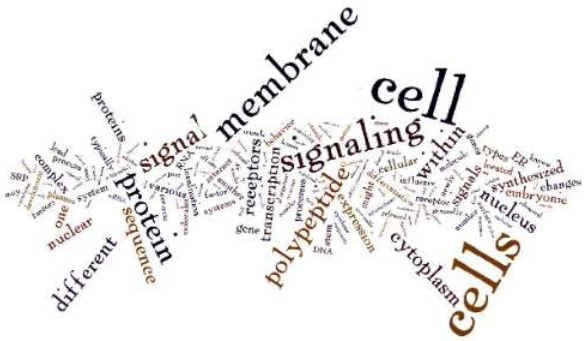
10.4: Signaling molecules and receptors

10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells

---

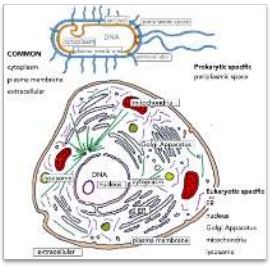
This page titled 10: Social Systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

10.0: Introduction



In which we consider the signals, receptors, and molecular machinery that control how proteins come to be where they are needed within cells and organisms, and how cells interact with one another through various signaling systems.

As noted earlier, each cell is a bounded non-equilibrium system. The plasma membrane forms an unambiguous boundary between the rest of the universe and the cell. In prokaryotes, the cell is typically surrounded by a cell wall, a semi-rigid structure that protects the cell from osmotic effects among other things. As discussed, the cell's metabolic activities occur primarily within the space defined by its cell membrane, the cytoplasm.[385] A polypeptide synthesized within the cytoplasm has a number of places it might end up, and like other biological processes these outcomes are controlled by signals and receptors. In a prokaryote (←), a newly synthesized polypeptide can remain in the cytoplasm where it can interact with the organism's genetic material, a circular DNA molecule also located in the cytoplasm. Alternatively, a newly synthesized polypeptide can end up embedded within the plasma membrane (an integral membrane protein) or it can pass through the membrane and be secreted. Secreted proteins (in a prokaryote) can remain within the periplasmic space, can become part of the cell wall, or can pass through the cell wall and into the external environment.



Eukaryotic cells (←) are more topologically complex and contain a distinctive double membrane structure, the nucleus. The cell's genetic material, its DNA, is organized into linear chromosomes located within the nucleus. The synthesis of RNA molecules, occurs within the nucleus. The membranes of the nucleus are elaborated within the cytoplasm into a network known as the endoplasmic reticulum or ER. There are a number of other intracellular membranes, including the Golgi apparatus and various types of small vesicles, involved in moving molecules to and from the plasma membrane and between the ER and other intracellular membrane systems. Finally, there are the mitochondria and (in plants) chloroplasts; these are double-membrane structures with their own genomic (circular) DNAs, derived from apparent endosymbiotic events early in the history of eukaryotes (discussed earlier). Complex signal-receptor interaction mechanisms serve to maintain the topological details of cells as they give rise to new cells. Our focus here is on the general rules by which specific proteins are "targeted" to specific cellular locations and compartments.

References

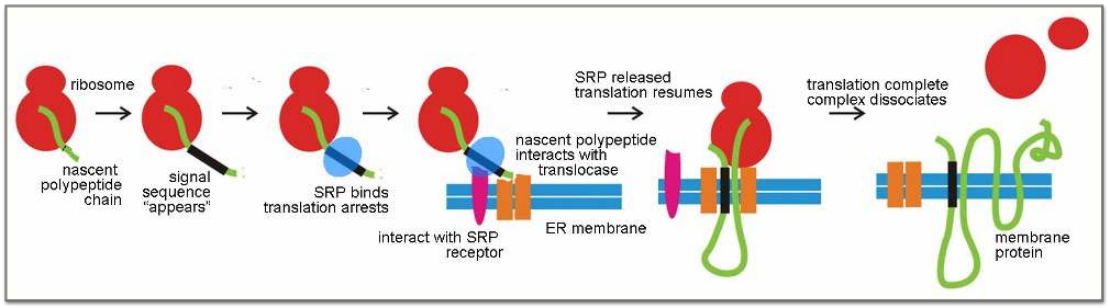
[385] In prokaryotes, there is a space between the cell membrane and the cell wall known as the periplasmic space; a number of reactions occur within this region. A similar space exists in plants and fungi that, unlike animal cells, have evolutionarily distinct cell walls.

This page titled 10.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

10.1: Targeting membrane proteins to where they need to be

What determines where a polypeptide (protein) ends up? As you might suspect, there are signals and receptors involved. Signals are typically part of the polypeptide's primary (amino acid) sequence and receptors are proteins encoded by other genes. The receptors that are present in the living cell, they are part of what is inherited from a cell's progenitor, part of the continuity of life. We begin our description of polypeptide targeting with prokaryotes, because they are simpler. We will consider how a newly synthesized polypeptide comes to end up in the cytoplasm, the plasma membrane, or outside of the plasma membrane.

In prokaryotes, the genomic DNA is located in the cytoplasm; there is no barrier between a newly synthesized RNA molecule and the ribosomes, tRNAs, and the other components involved in RNA-dependent polypeptide synthesis. Newly synthesized mRNA molecules can interact with the small and large ribosomal subunits, assemble with them to form a functional ribosome and direct polypeptide synthesis. For a water-soluble cytoplasmic polypeptide, as opposed to a polypeptide that resides in, or passes through the membrane, no further "signals" are necessary. The ribosomal complex moves along the mRNA, the polypeptide is synthesized, passes through the ribosomal channel, and emerges into the cytoplasm. When the ribosome reaches a stop codon, release factor binds, leading to the disassembly of the ribosomal-mRNA-polypeptide complex. The ribosomal components, as well as the mRNA can then initiate a new mRNA-ribosome complex, to produce another polypeptide. The released (newly synthesized) polypeptide may fold on its own or associate with other polypeptides to form a functional protein. Some of these folding steps may involve interactions with chaperones.



What is going on with a polypeptide destined for insertion into a membrane? Clearly it has a different structure than a water-soluble protein; differences you should be able to predict. The first step in delivering a membrane protein to or through a membrane is to recognize that it as a membrane protein or a protein that needs to pass through the cell membrane. The general mechanism (and the only one we will consider) involves what is known as a signal sequence (↓). A signal sequence is composed primarily of hydrophobic amino acids; the typical signal sequence is between 8 to 12 amino acids in length and generally located near the polypeptide's N-terminus, the first part of the polypeptide to be synthesized. The presence of such a signal sequence marks the polypeptide as a membrane protein. As the nascent polypeptide emerges from the ribosomal tunnel, the signal sequence is recognized through its binding of a cytoplasmic receptor, the signal recognition particle (SRP). SRP is composed of polypeptides and a structural RNA. The binding of a SRP to a signal sequence causes translation to halt, although the mRNA-ribosome-nascent polypeptide-SRP complex remains intact. The complex diffuses within the cell until it engages an SRP-receptor located on the cytoplasmic surface of the plasma membrane. The SRP receptor is associated with a transmembrane polypeptide translocator protein (↑). When the mRNA-ribosome-nascent polypeptide-SRP+SRP Receptor complex forms, SRP disassociates from the ribosome-nascent polypeptide complex, translation resumes and the nascent polypeptide interacts with the translocator protein and either folds to become embedded within the membrane, or passes through the membrane, and is released (secreted) on the other side. Typically, if the polypeptide is secreted, the signal sequence is removed by proteolysis during the process.

Now consider the situation in eukaryotic cells. Although topologically more complex the same basic process applies. The difference is that the SRP receptor is not located in the plasma membrane, it is located in the ER membrane. A protein with a signal sequence will be delivered to the ER membrane or released into the lumen of the ER. From there other signals will determine whether the protein stays in the ER, moves to the Golgi apparatus, where it is post-translationally modified, and may then move to the plasma membrane, or to some other membrane compartment within the cell. A protein in the lumen of the ER is effectively outside of the cytoplasm, and can be retained within a membrane compartment (such as the ER) or secreted from the cell. At this point, we will not concern ourselves with further details, except to say that whenever a protein is targeted to a specific cellular compartment, we can assume that the protein contains signals that are recognized by receptors that lead to its localization.

This page titled 10.1: Targeting membrane proteins to where they need to be is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

10.2: Nuclear targeting and exclusion

All cellular, that is non-mitochondrial / chloroplast polypeptides are synthesized in the cytoplasm, but can be assembled in any of the cell's topologically distinct compartments. So, what happens if the protein needs to be assembled and functions in the nucleus, say as part of the DNA replication, DNA repair, RNA transcription, or RNA processing machinery? And what about a cytoplasmic protein that might interfere with such processes if it were to find its way into the nucleus? Again we find the same pattern, there must be signals, typically amino acid sequences that indicate the protein should be located within or excluded from the nucleus. Such signals exist, and are referred to as nuclear localization (NLS) or nuclear exclusion (NES) sequences. Such sequences interact with receptors, that is, molecular machines associated with the nuclear pore complex that mediate the polypeptide's (protein's) translocation into or out of the nucleus.

It is worth noting that a protein can contain both NLS and NES sequences. Their "activities" can be regulated by allosteric effector binding or post-translational modifications. NLS and NES sequences may be accessible or inaccessible, that is unable to interact with the nuclear pore machinery. Where a protein is within a cell, that is, the percent of the protein in a cell located in the nucleus, the cytoplasm, or both, can be controlled. The extent to which a protein, such as a transcription factor or kinase, is within the nucleus will influence its functional impact on the cell. Nuclear localization of a positively acting transcription factor can lead to the activation of a gene, as can the nuclear exclusion of a negatively acting transcription factor. Changing the intracellular distribution of a transcription factor, whether positively or negatively acting, can influence the expression of the genes the transcription factor regulates. The situation is different from that found in membrane targeting (the signal sequence-SRP system), which is essentially irreversible - once a protein is inserted into a membrane or excreted from the cell, and its signal sequence removed, the protein cannot return to the cytoplasm. In contrast, many proteins can shuttle back and forth between nucleus and cytoplasm.

Questions to answer:

149. How is a water soluble protein different from a protein that resides in a membrane?
150. What are the components needed to insert of polypeptide/protein into or through a membrane? How might mutations in these proteins influence a polypeptide’s localization within a cell?
151. redict what would happen if a signal sequence were mutated.
152. How might you activate a NLS or NES sequence within a protein? How might such a sequence be rendered inactive?

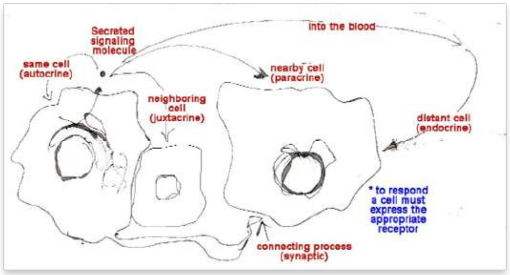
Question to ponder:

- How might a cytoplasmic protein be inserted into a membrane?

This page titled 10.2: Nuclear targeting and exclusion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

10.3: Intercellular signaling- signals, receptors and responses

The ability of cells to place proteins on their surface and to secrete proteins into the extracellular space, opens up the possibility of various forms of signaling between cells. Intercellular signaling enables cells to influence each other in various ways.[396] Here we consider only the basics of such processes, more details will be added later on. Intercellular signaling system involves the synthesis of a signaling molecule. This involves expressing the gene(s) encoding the signaling molecule or the metabolic machinery needed for its synthesis, followed by its processing, and secretion from or localization to the cell surface (next↓page). Similarly, for a cell to respond to a signal, whether from another cell or from itself, a cell has to express a receptor for the signal molecule. Such receptor proteins are generally located on the responding cell's surface. When the signal binds to the receptor it acts as an allosteric effector, changing the structure and behavior of the receptor. Different signal-receptor combinations produce different types of changes in the receptor, changes that initiate a cascade of events leading to changes in cell behavior, gene expression, or (often) both.



When signaling molecules are released from a cell they are (generally) free to diffuse. In autocrine signaling (←), the cell that released the signal also has receptors for the signal; in a sense the cell talks to itself.[397] In paracrine signaling, cells located near the signal secreting cell can respond, if they have the appropriate receptors on their surfaces. Endocrine signaling occurs when the released signaling molecules are transported widely throughout an organism, typically through the blood stream. Juxtacrine signaling occurs when the signaling and receiving cells need to "touch one another" through surface membrane proteins. Altogether such interactions underlie the coordination of the behavior of groups of cells; they are the basis for multicellularity, cellular differentiation, organ formation and coordination, and the formation and function of metabolic, immune, and nervous systems. The effects of intercellular signaling can be transient, for example, as in muscle contraction, or they can lead to irreversible changes in gene expression, cell morphology, and/or behavior. Signaling induced cascades in changing gene expression and cellular behaviors underlie embryonic development and disease progression.

References

[396] Antebi et al. 2017. An operational view of intercellular signaling pathways

[397] as an example, see Glucagon regulates its own synthesis by autocrine signaling

This page titled 10.3: Intercellular signaling- signals, receptors and responses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

10.4: Signaling molecules and receptors

Molecules that provoke a signaling response are typically called agonists. Different agonists interact with agonist-binding receptors, typically composed of one or more integral membrane proteins. Their interactions produce distinct “down-stream” molecular cascades that exploit post-translational modification or allosteric effects to activate or inactivate various enzymes and transcription factors. In general for each component of a signaling system, there are molecules (generally proteins) that can act as antagonists; they inhibit the signaling process. Antagonists (or inhibitors) may bind agonists, receptors, or "downstream" effectors and so block signaling. Moreover, any one particular cell may express a number of different signaling pathway components; cells of different types will express different combinations of signaling systems, so they may respond to different incoming signals differently.

In cases where signaling leads to changes in gene expression, these changes can modify the behavior of the cell, and lead to changes in cellular phenotype. As a general rule, any particular signaling input will generate both direct and indirect effects. For example, activation of a signaling system may lead to the activation (or repression) of a specific set of transcription factors. These can directly regulate the expression of a set of target genes. Some of these genes may themselves encode transcription factors, or polypeptides that regulate transcription factor activity and gene accessibility. The expression of these genes will, in turn, regulate other genes – these are considered indirect or secondary targets of the signaling system. Since which genes will be turned on or off will be influenced by the total set of transcription factors and associated proteins that are expressed and active in a cell, the response of different cepptypes of cells to the same signal can be different, and characteristic of the cell type. For example, a muscle cell might respond differently from a kidney cell to the same signal. Similarly, once a cell response to a signal, changes in the patterns of gene expression and protein activity can lead to subsequent changes in cell morphology and behavior, including evolving changes in patterns of gene expression. It may differentiate, that is become different from what it was originally. The process of embryonic development consists of a series of signals and cellular responses that lead to the specialization of cells, the development of tissues and organ systems. Normally, this process of signal-driven differentiation is irreversible. It proceeds in one and only one direction. The process results in what is known as terminal differentiation. Only recently have strategies been developed that can reverse these effects.

This page titled 10.4: Signaling molecules and receptors is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells

A question asked by early developmental biologists was why exactly do cells differentiate? Was it the due to the loss of (differentiation inhibiting) genetic information? Is the genomic DNA of a neuron different from that of a skin or a muscle cell? To answer this question Briggs and King (in the 1950s) carried out nuclear transfer experiments in frogs. These experiments were extended by Gurdon (1933-) and McKinnell (1926-2025) in the early 1960s. They were able to generate adult frogs from fertilized eggs in which the original nucleus was replaced by the nucleus from a differentiated cell.[398] The process was inefficient - in only a small percentage of the eggs with a transplanted differentiated cell nucleus was normal embryonic development observed. The ability of somatic nuclei to be "reprogrammed" by the egg differed between different somatic cell types. This suggested the presence of effectively irreversible changes in differentiation-associated DNA/chromatin modifications.[399] It was also possible that stochastic variations between cells of the "same type" influenced the ability of transplanted nuclei to support normal development. Nevertheless, these experiments suggested that it was changes in gene regulation rather than the loss of genetic information that was important in embryonic differentiation.

In 1996 Wilmut et al used somatic cell nuclear transplantation to clone the first mammal, a sheep named Dolly. Since then many different species of mammal have been cloned. In 2004, cloned mice were derived from the nuclei of olfactory neurons using a method similar to that used by Gurdon. These neurons came from a genetically engineered mouse that expressed the fluorescent protein GFP in most cell types. After the nuclei of a mature (haploid) oocyte was removed, a neuronal (diploid) nucleus derived from the GFP-mouse was introduced. Blastula derived from these cells were then used to generate totipotent embryonic stem cells from cells of the inner cell mass. A totipotent cell is capable of producing, through cell division and differentiation, all of the different types of cells in the adult. It was the nuclei from these cells that were then transplanted into enucleated eggs. The resulting embryos were able to develop into fully grown fluorescent mice, proving that neuronal nuclei retained all of the information required to generate a complete adult animal.
The process of cloning from somatic cells is inefficient – many attempts had to be performed, each using an egg, to generate an embryo that is apparently normal (most embryos produced this way were abnormal). There are serious ethical issues associated with the entire process of reproductive cloning, particularly given the persistent inequalities in modern society.[400] For example the types of cells used, embryonic stem cells, are derived from the inner cell mass of mouse or human embryos - their isolation involves destroying the original embryo.

In a breakthrough series of studies, Takahashi and Yamanaka (2006) determined that introducing a set of four transcription factors Oct3/4, Sox2, c-Myc, and Klf4 into terminally differentiated cells led some of the transfected cells to reverse their differentiation, and return to a more pluripotent state, that is a state that can subsequently differentiate into many different cell types.[401] This process of dedifferentiation has been found to be robust, and the dedifferentiated cells produced are known as “induced pluripotent stem cells” or iPSCs. iPSCs behave much like embryonic stem cells. The hope is that patient-derived iPSCs can be used to generate tissues or even organs that could be transplanted back into the patient, and so reverse and repair disease-associated damage.

Questions to answer:

153. What is the benefit of (essentially) irreversible cellular differentiation in multicellular animals?
154. Based on your understanding of the control of gene expression, outline the steps required to reprogram a nucleus so that it might be able to support embryonic development.

Questions to ponder:

- Why, if differentiation is normally uni-directional and irreversible, is it possible to artificially reprogram somatic cells to an “earlier” state? Why doesn’t this happen all the time in your body?
- What are the main ethical objections to human cloning? What if the clone were designed to lack a brain, and destined to be used for "spare parts”? Does that change anything, or does it make things worse?



Short Chapter Summary

- Proteins end up in the "right" places (nucleus, membranes, organelles) via targeting signals.
- Signals + receptors → cascades → gene and behavior changes; specificity comes from wiring and dose.
- Cell identity can be reprogrammed; gene regulatory landscapes are plastic.

References

[398] The egg and the nucleus: a battle for supremacy: http://www.nobelprize.org/mediaplayer/?id=1864

[399] see: Individual neurons may carry over 1,000 mutations

[400] J. Gray. 2017. A History of the Future: how writers envisioned tomorrow's world

[401] Takahashi & Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.

This page titled 10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

11: Cellular reproduction and horizontal gene transfer in prokaryotes

11.0: Introduction

11.1: What counts as sex in prokaryotes

11.2: Other naturally occurring horizontal gene transfer mechanisms

---

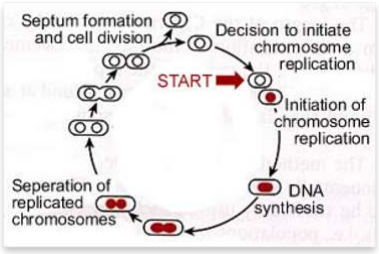
This page titled 11: Cellular reproduction and horizontal gene transfer in prokaryotes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

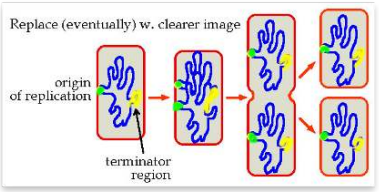
11.0: Introduction



In which we consider how prokaryotes replicate asexually and how they can (under specific conditions) pass genetic information to one another and acquire such information from their environment.

The simplest types of biological (cellular as opposed to viral) reproduction are probably the asexual processes found in prokaryotes.[402] In bacteria and archaea, the genome consists of a single large circular double-stranded DNA molecule, the bacterial chromosome. In some cases, the cell also contains smaller circular DNA molecules, known as plasmids. For the moment we will ignore plasmids and focus on the chromosome, although the replication processes are similar.[403] The chromosome contains two important DNA sequence elements, the origin of replication (ORI) and the terminator region (TER). When conditions are appropriate, a cell will pass through a decision point, a molecular switch, known as "start" (→). This switch activates proteins that bind to the ORI region of the chromosome, initiating the assembly of the DNA replication complex. A replication bubble (a region of the DNA in which the two DNA strands have separated) forms, and replication forks begin to move in both directions around the molecule, making a copy. As the ORI sequence is replicated, the two ORI sites remain associated with the plasma membrane. The replication forks (↓) move around the DNA molecule, and collide in the TER region.The collision of the DNA replication forks generates a signal that indicates that DNA replication is complete. During this period the cell is growing, adding mass and volume. The division of one cell into two is mediated by the formation of a septum, an extension of the plasma membrane and the cell wall. Septum growth initiates between the two membrane-bound ORI sequences that ensures that each daughter cell receives one complete chromosome, one total genome.





In the chromosome the order of genes around the circular molecule is generally conserved between organisms of the same species. The genes along the chromosome constitute a syntenic linkage group, the same genes in the same order along a chromosome (discussed further below). In the standard asexual mode of replication, all of the alleles are inherited together, the result is that a mutation in any particular gene (generating a new allele) acts in concert with the other alleles (in the other genes) present. Over time, each organism produces a clone, and various clones interact with the environment and each other. Clones can display different levels of reproductive success; some clones can take over the population, while others may become extinct. In the case of studies on the evolution of bacterial antibiotic resistance (see below), each clone has to develop antibiotic resistance independently of every other clone; a similar situation was observed in long term bacterial evolution studies.[404] There is no cross talk between lineages in such situations. Of course, if DNA is passed from clone to clone, as occurs within Griffith's (previously considered) transformation experiments, things get more complex. The movement of genes between lineages is known as horizontal gene transfer. We consider the three versions of horizontal gene transfer found in prokaryotes.

References

[402] While viruses may seem simpler, they act as molecular parasites and rely on cellular systems to replicate. They are probably best discussed in a more advanced course. We ignore them here.

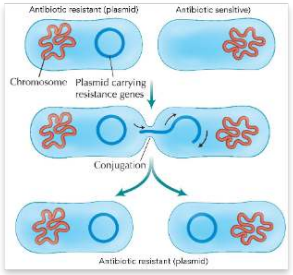
[403] Noirot-Gros et al., 2002. An expanded view of bacterial DNA replication

[404] see A cinematic approach to drug resistance and E. coli Long-term Experimental Evolution Project

This page titled 11.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

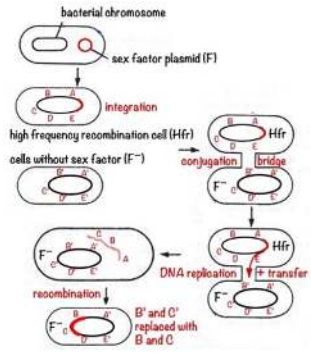
11.1: What counts as sex in prokaryotes

Conjugation is a major pathway for horizontal gene transfer in bacteria.[405] In contrast to transformation, conjugation "forces" DNA into what may be a reluctant recipient cell. In the process of conjugation, we start by distinguishing between two types of bacterial cells (of the same species). F^+ cells contain a DNA sequence known as the fertility (or



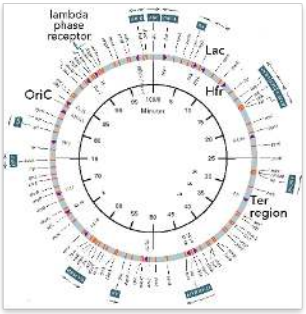
sex) factor F, cells without the F factor are referred to as F^- cells. The F factor can exist independently of the host chromosome as a plasmid. Discovered by Esther Lederberg (1922–2006), the F plasmid can be integrated into the host chromosome. When integrated the cells become what are known as high frequency recombination (Hfr) cells. The ~100 kilobase F plasmid contains about 100 genes that encode the proteins needed to transfer a single-stranded copy of its DNA into a cell that lacks an F-plasmid.[406] F plasmids typically contain genes that encode an addiction system (discussed earlier). Such systems encode a stable toxin and an unstable (rapidly degraded) anti-toxin. Once the plasmid enters a cell, both toxin and anti-toxin proteins are synthesized. If the plasmid is lost, the cell dies because of the anti-toxin disappearing before the toxin, leading to toxin activation and cell death.

Here we consider a population of F^- cells into which we insert an F^+ cell. The F-plasmid contains two distinct origins of replication - one, known as oriV, is involved in normal plasmid replication during cell growth and division. The second, known as oriT, is involved in generating the single stranded DNA molecule that can be transferred into a F^- recipient cell. To initiate conjugation, the F^+ cell makes a physical (conjugation) bridge, known as a pillus, that connects to the F^- cell (→). A single stranded copy of the F plasmid is synthesized and transferred through the pillus into the recipient F^- cell. Subsequent DNA synthesis generates a double-stranded copy of the F-plasmid in the recipient cell, while the donor cell retains the original plasmid.



If the F-plasmids integrates into the host cell genome, the cell becomes an Hfr cell (←). Integration of the F-plasmid can occur at various points along the host chromosome. As with the free plasmids, the integrated F-plasmid can initiate (at its oriT site) the transfer of its own as well as linked host genes into a F^- cell. The amount of DNA transferred will be determined largely by how long the bridge between the cells remains intact. In *E. coli* it takes ~100 minutes to transfer the entire donor chromosome from an Hfr to an F^- cell. Once inside the F^- cell, the transferred donor DNA is integrated (by homologous recombination) into the recipient's chromosome, replacing the recipient's versions of the genes transferred (a process to which we will return). Using Hfr strains carrying different alleles of genes associated with recognizable phenotypes, and by controlling the duration of conjugation by breaking the conjugation bridge by shearing the cells in a kitchen blender, experimenters were able to determined the order of genes along the chromosome. The result was the discovery that related organisms often had the same genes arranged in the same order.[407] The typical drawing of the circular bacterial chromosome is like a clock going from 0 to 100, with the genes placed in their respective positions, based on the time it takes to transfer them in minutes (→).

If the entire F-plasmid sequence is transferred, the original F^- cell becomes an Hfr cell. If a Hfr cell loses the F-plasmid sequence,



it reverts to a F^- state. The end result of the conjugation process is similar to that obtained in sexual reproduction in eukaryotes, namely the original F^- cell now has a genome derived in part from itself and from a"donor" Hfr cell. The outcome of an Hfr/F^- cell interaction can lead to a cell with a different set of alleles than either of the "parental" cells, this process is often referred to as bacterial sex, although it is mechanistically quite distinct from sexual reproduction in eukaryotes. Versions of this process are involved in the transfer of plasmids from cell to cell within a community.[408]

All plasmids contain an "origin of replication". "Low copy number" plasmids exist in one to two copies per cell, while "high copy number" plasmids may be present in as many as ~700 copies per cell.[409] Plasmid copy number is determined in large part by their origin of replication sequences. Plasmids can encode genes responsible for antibiotic resistance; the rapid dispersion of the antibiotic resistance phenotype is a cause of increasing concern.[410]

Many plasmids, also known as mobile genetic elements, are selfish, that is, their presence may not directly benefit that cell and their loss may result in the death of the host cell, due to the presence of an addiction module.

Questions to answer:

155. How would mutating the origin or terminator regions influence the cell's reproduction?
156. How would the progeny of an incomplete F-factor mediated recombination event differ from its "parents"?

Questions to ponder:

- Why might F-plasmids encode addiction modules
- How might a "selfish" plasmid evolve into a virus.

References

[405] review of prokaryotic conjugation and Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria

[406] fertility factor review by S.M. Rosenberg & P.J. Hastings 2001.

[407] Synteny: http://en.wikipedia.org/wiki/Synteny

[408] Plasmids Spread Very Fast in Heterogeneous Bacterial Communities: https://www.ncbi.nlm.nih.gov/pubmed/12524329

[409] Plasmids 101: Origin of Replication

[410] Addgene: Mechanisms of Antibiotic Resistance

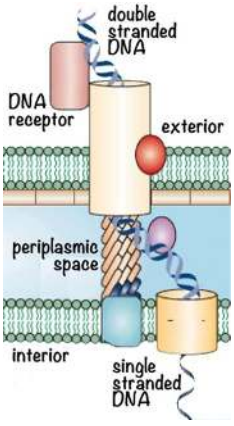
This page titled 11.1: What counts as sex in prokaryotes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

11.2: Other naturally occurring horizontal gene transfer mechanisms

Many gene transfer mechanisms are regulated by social and/or ecological interactions between organisms.[411] The mechanisms involved can be complex; one could easily imagine an entire course focused on this topic. We introduce only the broad features of these systems. Also, we want to be clear about the various mechanisms of DNA uptake. When an organism dies its DNA can be eaten by others as a source of energy, carbon, nitrogen, and phosphorus. When eaten the information in the DNA, the result of mutation and selection, is lost.[412] Alternatively, the released DNA molecule can be integrated into another organism's genome, resulting in the possible acquisition of whatever information was present in the sequence. This information might be useful, harmful, or irrelevant to the organism that acquires it. The study of these natural DNA import, as distinct from conjugation-mediated transfer systems has identified some of the specific molecular machines involved. Some organisms use a system that preferentially imports DNA molecules derived from organisms of the same or closely related type. You can probably imagine how they might do this – one way could be that they have receptor systems that recognize species-specific "DNA uptake sequences." The various mechanisms of horizontal gene transfer, unsuspected until relatively recently, have had profound influences on evolutionary dynamics, particularly within microbial communities, where they appear to be more common than among eukaryotes. The result is that, in many cases, an organism in such a population does not have to "invent" all of its own genes, it can adopt (import) genes generated by evolutionary mechanisms in other organisms in other environments for other purposes. So the question is, what advantages might such information uptake systems convey, and (on the darker side), what dangers do they make possible?

Transformation

There are well established methods, used in genetic engineering, to enhance the ability of bacteria to take up DNA from their environment.[413] Here wemfocus on natural transformation. Natural transformation is an active (energy-requiring) process that involves a number of components, encoded by genes that can be expressed or not depending upon environmental conditions. Consider a type of bacteria that can import DNA from its environment. If the density of bacteria is low, there will be little DNA to import, and it may not be worth the energetic expense associated with expressing the genes and synthesizing and assembling the proteins involved in the DNA uptake and integration machinery. Bacteria use quorum sensing systems (considered earlier) to monitor cell density and to control the expression of genes involved in the synthesis of the DNA uptake system. When present in a crowded environment, the quorum sensing system can turn on the expression of the genes involved in DNA uptake system.

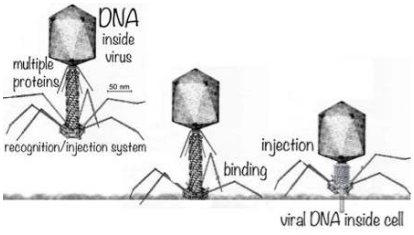


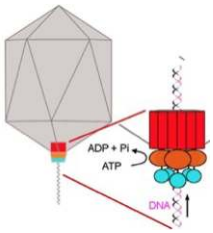
Here we outline the process in one type of bacteria but functionally similar mechanisms are used in other bacterial and archaeal species. Double-stranded DNA binds to the cell's surface through a variety of DNA receptor proteins. In some cases these receptors bind specific DNA sequences, in others they bind DNA generically, that is, any DNA sequence. As shown, Gram negative bacteria have two lipid membranes, an outer one and an inner (plasma) membrane, with a periplasmic space, between them. In an ATP-hydrolysis coupled reaction, DNA bound to the exterior surface of the bacterium is moved, through a protein pore across the outer membrane and into the periplasmic space, where it is passed to the DNA channel protein (→). Here one strand of the DNA is degraded by a nuclease while the other moves intact through the channel into the cytoplasm of the cell in a 5' to 3' direction (similar to the one-strand transfer seen in bacterial conjugation). Once inside the cell, the DNA associates with single-stranded DNA binding proteins and, by homologous recombination, can be inserted into the host genome (or degraded, depending on the system).[414] While the molecular details of this and functionally similar processes are best addressed elsewhere, what is key is that transformation enables a cell to take up foreign DNA and to add the imported DNA sequences to its own genome.

Virus moving genes: transduction

The final form of horizontal gene transfer that we will consider involves bacterial viruses. The structure and behavior of viruses is a complex topic, the details of which are largely beyond us here, but it is not unreasonable to consider viruses as nucleic acid transport machines. Viruses are dependent for their replication on the infected host cell, they have no active metabolic processes. They are not alive in any meaningful sense of the word. That said, they are infectious, they can spread through a population. Viruses cannot be killed, because they are not alive, but they can be inactivated by various treatments.

Viruses contain a nucleic acid genome and a protein-based transport and delivery system. We consider (briefly) a typical bacterial virus, known as a bacteriophage or bacteria eater. The T4bacteriophage T4, looks complex and it is (→), other viruses are somewhat simpler. The T4 phage, short for bacteriophage, has a ~169,000 base pair double-stranded DNA genome that encodes 289 polypeptides, almost as many as a minimal cell (see above).[415] The assembled virus has an icosahedral protein head that contains a DNA molecule attached to a tail assembly that recognizes and binds to target cells. Once a suitable host cell is found, based on tail binding to cell surface molecules, the tail domain attaches to the cell's surface and contracts, like a syringe, punching a hole through the bacterial cell's external wall and plasma membrane. The DNA emerges from the bacteriophage and enters the cytoplasm, infecting the cell. Genes within the phage genome are expressed, leading to the replication of the phage genome and the fragmentation of the host cell's DNA.[416] The phage DNA encodes the proteins used to build and assemble new phage. DNA is packed into these heads by a protein-based DNA pump (↓) driven by an ATP hydrolysis reaction complex.[417] In the course of packaging viral DNA, the system will, occasionally, make a mistake and package a fragment of the host cell's DNA. When such a phage particle infects another cell, it can inject that cell with a DNA fragment derived from the previous host. The mis-packaged DNA may not contain all of the genes the virus needs to make a new virus or to kill the host. If this is the case, the host cell may have to be co-infected by a wild type virus for the mutant virus to replicate. The DNA transferred by the virus to the host can be inserted into the host cell's genome, with the end result being similar to that discussed previously for transformation and conjugation. DNA from one organism is delivered to another, horizontally rather than vertically.





Because the horizontal movement of DNA via toxic viruses is common in the microbial world, a number of defense mechanisms have evolved to control it.[418] These include restriction endonuclease / DNA modification systems used widely for genetic engineering, and the CRISPR-Cas9 system that enables cells to recognize and destroy foreign (viral) DNAs. These systems, evolved as part of the prokaryotic immune systems, together with various plasmids form the tools used in modern molecular biology and genetic engineering methods. They illustrate how studying apparently arcane aspects of the biological world, bacterial viral defense mechanisms, can have dramatic impacts on modern technological, medical, and economic systems.

Questions to answer:

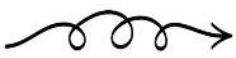
157. What is an asexual clone? How would you recognize it.
158. What is the effect of an amorphic allele / mutation on the behavior of a prokaryotic clone.
159. What are some possible (evolutionary) advantages to the ability to take up and integrate, as opposed to simply eat foreign DNA?
160. Why might the "source" of foreign DNA matter?
161. Present a plausible model that would identify host from foreign DNA
162. Propose a model by which a "selfish" plasmid might evolve into a virus.
163. How can co-infection of a cell with wild type virus "rescue" a virus that has lost some of its essential genes?
164. How might inserting a piece of DNA into a bacterium's genome be harmful

Questions to ponder:

- Describe a mechanism by which a prokaryotic organism might protect itself from invading viruses?
- How is it that "punching a hole" in a membrane (during DNA uptake or phage infection) does not kill the cell?
- How does vertical differ from horizontal inheritance?

Possible extension:

- Introduce and consider the role of the lysogenic / lytic switch in bacteriophage / bacterial interactions.
- Extend discussion to mobile genetic elements



Short Chapter Summary

- Bacteria/archaea exchange genes via transformation, transduction, and conjugation—evolution on fast-forward.

- HGT spreads innovations (e.g., antibiotic resistance); “sex” without meiosis still reshapes populations.

References

[411] DNA uptake during bacterial transformation: http://www.ncbi.nlm.nih.gov/pubmed/15083159

[412] This is of course why genes are rarely if ever transferred from food to the organism doing the eating.

[413] Making Calcium Competent (bacterial) Cells: http://mcb.berkeley.edu/labs/krantz/protocols/calcium_comp_cells.pdf

[414] Bacterial transformation: distribution, shared mechanisms and divergent control & Natural competence and the evolution of DNA uptake specificity

[415] http://en.wikipedia.org/wiki/Bacteriophage_T4

[416] An infected bacterial cell can protect is neighbors, often its clonal relatives, if it can kill itself before the virus can replicate. This is an example of a simple altruistic behavior.

[417] The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces

[418] see The phage-host arms-race: Shaping the evolution of microbes

---

This page titled 11.2: Other naturally occurring horizontal gene transfer mechanisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

12: Asexual and Sexual Reproduction in Eukaryotes

12.0: Introduction

12.1: Ploidy during the cell cycle

12.2: Monitoring cellular processes- mitosis

12.3: Sex-determination and its chromosomal basis

12.4: Steps in meiosis- from diploid to haploid

12.5: Recombination and independent segregation

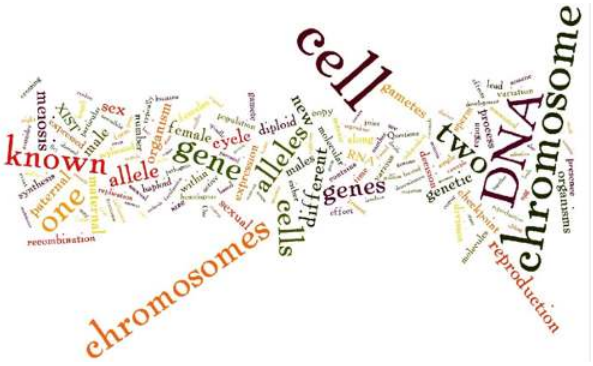
12.6: Linkage and haplotypes

12.7: X-inactivation and sex-linked traits

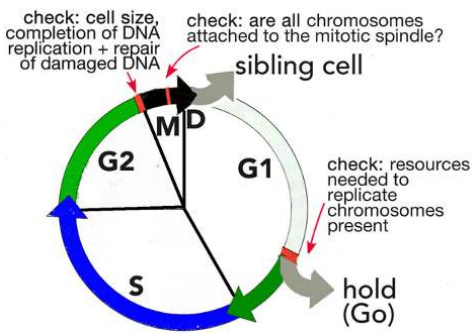
12.8: X-linked diseases and mono-allelic gene expression

This page titled 12: Asexual and Sexual Reproduction in Eukaryotes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

12.0: Introduction



In which we consider asexual and sexual reproduction, chromosome segregation (in mitosis) & cell division and how they are modified (in meiosis) to produce haploid gametes that fuse to form a new diploid organism. We consider the implications of chromosome pairing, recombination & independent segregation as well as sex determination, the dimorphism of gametes, the maternal inheritance of mitochondria, and maternal and paternal effects.



In general terms asexual reproduction in a eukaryote is similar to that in a prokaryote. As a cell grows at some point there is a molecular decision to replicate its genomic DNA and to divide into two. The complication arises from the fact that eukaryotic cells typically have multiple linear chromosomes; one copy of each has to be accurately delivered to the "new cell". Eukaryotic cells also have essential cytoplasmic organelles: mitochondria, and in algae and plants, chloroplasts. These organelles need to be replicated and delivered to the new cell. Mitochondria and chloroplasts have their own small but essential genomes. As you might guess, since they appear to be derived from prokaryotes, these organellar genomes are circular DNA molecules. In the course of asexual, sometimes termed somatic reproduction, each of the sibling cells receives a number of these cytoplasmic organelles.[419] In the eukaryotes that we will consider, most of the cells of the organism are diploid.

Somatic (asexual) reproduction involves what is known as the cell cycle. We can think of the cell cycle as beginning with cell division (D →). In somatic cells, the process of dividing one cell into two is known as cytokinesis; it produces two sibling cells, each with identical genomes, excluding rare mutations that might arise during DNA replication. Cytokinesis involves cytoskeletal and cytomuscular systems that are discussed in detail in later cell biology courses - not here! Generally, cell division by cytokinesis is usually, but not always symmetrical, so that the two sibling cells are half the volume of the parental cell and very similar. Asymmetric cell division can occur, and generally results in cells that behave differently.[420] Cytokinesis is followed by a period of cell growth, known as G_1 (→), during which energy and materials that are imported from the external environment, or have been previously stored within the parental cell, are converted into lipids, nucleic acids, proteins, and other molecules leading to the growth of the cell. As the cell grows a number of decisions need to be made: will the cell continue to grow and divide or will it stop growing and enter a steady state where it maintains itself (building and disassembling molecules, repairing DNA, etc) – a state known as G_o (↑). Generally, the majority of cells in any particular tissue are in the G_o state. In G_o there is no new DNA synthesis, so the possibility of mutation is lower than when DNA is being replicated. If or when various external and internal signals act on and within the cell, many (but not all) cells can reverse the G_o decision and resume growth and eventually divide. It is difficult to talk about these systems without personalizing them, even though they are not self-conscious "decisions" but the outcomes of molecular switches acting as the system level - what could be considered the cell's consciousness.

\start→)

The decision to start DNA synthesis is based in part on whether the cell has, or can expect to have, the resources needed to complete DNA replication. In a human cell this requires ~12 billion nucleotide addition reactions. The DNA synthesis decision point is known as “start” (↑). There are mutant alleles, originally described through genetic studies in yeast, that result in a malfunctioning molecular switch controlling start. Such mutations, called "wee" mutants by their Scottish discoverers, lead to a disconnect between growth and division and result in smaller and smaller cells and eventually cell death.[421] Once a cell passes

through the start checkpoint, the cell enters the part of the cell cycle during which DNA synthesis occurs, known as S (↑). As genomic DNA synthesis begins there are other "checkpoints".[422] Checkpoints are molecular feedback systems by which the cell monitors various aspects of its internal state and makes a decision to pause or proceed with a process, in this case DNA synthesis and later cell division.

During S phase the cell continues to grow and replicate its DNA. In contrast to the smaller circular genomes of prokaryotic cells, with their single origin of replication, the much larger eukaryotic genomes and their organization into multiple linear chromosomes involves multiple DNA synthesis start sites per chromosome. These multiple replication origins are regulated such that each is activated once and only once during S phase in order to ensure that each region of the genome is replicated once and only once. Before cytokinesis, a checkpoint monitors for the presence of unreplicated DNA and delays the cell cycle until all DNA has been replicated.[423] DNA replication can lead to mutations, so a checkpoint monitors the completion of replication-associated DNA repair processes. The presence of this checkpoint explains the observation that damaging DNA by radiation or inhibiting DNA synthesis using drugs, leads to delays in the cell cycle. Pathogens, such as the bacteria *Listeria monocytogenes*, can exploit this DNA damage checkpoint to enhance their own replication at the expense of the host.[424]

References

[419] Plants and algae, which we will not be discussing in any detail, contain a second type of intracellular, DNA-containing organelle, known as chloroplasts. Their inheritance is similar to that of mitochondria.

[420] These differences are discussed in detail in the section on developmental biology.

[421] Paul Nurse and Pierre Thuriaux on wee Mutants and Cell Cycle Control: https://www.ncbi.nlm.nih.gov/pubmed/27927897

[422] The quorum sensing systems we discussed previously is a version of a checkpoint system.

[423] DNA replication is complex process, see Can the Stalling of DNA Replication Promote Epigenetic Changes?

[424] Listeria monocytogenes induces host DNA damage and delays the host cell cycle to promote infection

This page titled 12.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

12.1: Ploidy during the cell cycle

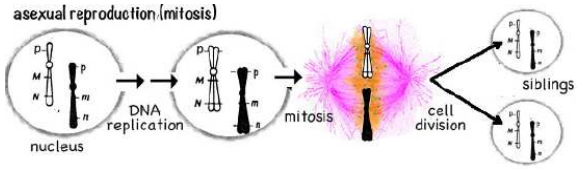
By the end of S phase DNA synthesis is complete. The cell now has two complete copies of each chromosome. At this point the cell has entered into what is known as the G_2 phase of the cell cycle. Cells can continue to grow in G_2. During the asexual reproduction cycle the ploidy, the number of copies of the genome and each chromosome, is conserved. A haploid cell gives rise to a haploid cell, while a diploid cell gives rise to a diploid cell. After S-phase and during G_2 there are twice the number of copies of the genome, and of each chromosome. While a diploid cell is diploid during G_1, it is effectively tetraploid during G_2. This can have physiological effects because the more copies of a gene the more RNA molecules can be synthesized per unit time. Based on this logic, we might expect to see changes in the rates and patterns of gene expression in G2 compared to G1 cells.

Questions to answer:

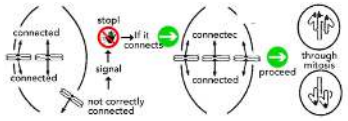
165. How many ways can you think up by which a cell could detect, and attempt to repair, damaged DNA or errors in DNA synthesis?
166. What factors limit the efficiency of DNA repair mechanisms?
167. Why, do you suppose, does a wee mutant cell eventually die?
168. What effects could arise from the local over- or under-replication of DNA during S phase?
169. How might gene expression change over the course of the somatic cell cycle?

This page titled 12.1: Ploidy during the cell cycle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

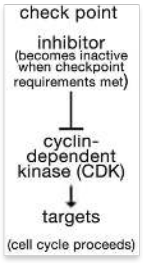
12.2: Monitoring cellular processes- mitosis



Once the DNA replication/repair checkpoint has been passed, the cell can divide. The first step of this process (in eukaryotes) is known as mitosis (→). Mitosis involves a molecular machine, the mitotic spindle, based on protein polymers, αβ-tubulin-based microtubules and actin-myosin based contractile ring. There is a molecular checkpoint that monitors the assembly of the mitotic spindle, and a checkpoint that monitors that each replicated chromosome has connected correctly to the mitotic spindle, made of organized microtubules (← see video link). Each replicated chromosome consists of two linear double stranded DNA molecules. The pair of replicated chromosomes interacts with the mitotic spindle through a specific protein structure known as the kinetocore. Kinetocores are assembled in association with specific chromosomal DNA regions known as centromeric sequences. Each replicated chromosome will have its own kinetocore and each interacts independently with the mitotic spindle



(this is different from their behavior during meiosis, as we will see). The presence of the mitotic chromosome attachment checkpoint was recognized in experiments in which chromosomes were manipulated so that they could not connect correctly to the mitotic spindle; such a manipulation caused a delay or halt in mitosis until the "mistake" was corrected.[425] The mitotic checkpoints serve to ensure that each sibling cell gets one and only one copy of each and every chromosome present in the parental cell.[426] Once these checkpoints are passed, molecular links between replicated chromosomes are severed and the mitotic spindle moves chromosomes to opposites sides (poles) of the parental cell. The parental cell then divides using the contractile ring, a actin/myosin polymer based molecular machine. While the two cells inherit the same set of alleles as were present in the parental cell, they may behave differently due to differences in different environments they find themselves in - factors returned to when we consider developmental processes.



The cell cycle decision check points are composed of multicomponent interaction networks. While we consider check point mechanisms only briefly, they play important roles in developmental processes and many diseases. A typical check point is built around a protein kinase, an enzyme that phosphorylates various targets – such post-translational phosphorylation can lead to changes in protein structure, protein-protein interactions, protein activities and stability, and intracellular localization. Cell cycle checkpoints often involve a particular class of kinases, known as cyclin-dependent kinases (CDKs)(→). The activity of these CDKs can be regulated positively by the binding of a small regulatory protein, known as a cyclin, as well as negatively by other interacting proteins and post-translational modifications. Cyclin's and other regulators are themselves the target of various forms of regulation, including proteolytic degradation, triggered by their post-translational modification. Typically cyclin-CDK complex formation and activity is inhibited by various factors (proteins). When the conditions involved in the checkpoint are met, inhibition is removed allowing for the cyclin-CDK complex to form and become active. The active kinase phosphorylates and regulates the activity and stability of its targets, allowing the cell to pass through the check point and proceed along the cell cycle. One effect of activating the CDK is often the rapid degradation (removal) of the cyclin; this makes the switch effectively irreversible until such time as cyclin protein levels increase again during the next cell cycle.

Questions to answer:

170. How do chromosomes interact with one another during mitosis/cytokinesis?
171. What does it mean that a checkpoint acts to "make a decision based on evidence"?
172. How does cyclin degradation make a checkpoint decision effectively irreversible?
173. Make a graph of CDK activity and the concentration of the cyclin regulating it, as a function of the cell cycle.
174. Predict what might go wrong if a checkpoint is ignored? (start with a cell cycle diagram)

Questions to ponder:

- Why is the decision to start a new cell cycle critical?
- When is the decision to start a new cycle made?

References

[425] Mitotic forces control a cell-cycle checkpoint

[426] Kinetochores, microtubules, and spindle assembly checkpoint signaling

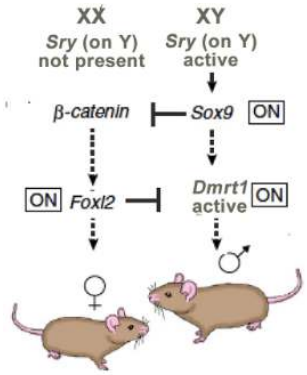
This page titled 12.2: Monitoring cellular processes- mitosis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

12.3: Sex-determination and its chromosomal basis

In eukaryotes, the generation of a new organism, often involves the process of sexual reproduction. Different types of organisms determine an individual's sex using different mechanisms, and in some cases, a single individual, known as a hermaphrodite, can display traits of both sexes at either the same time or sequentially.[427] There are basically two general mechanisms that determine the sex of an organism: genetic and environmental, although do not be confused, environmental processes are based on molecular and cellular switches encoded in DNA. In environmental sex determination various external signals influence the sex of the organism. For example in a number of reptiles, the sex of the adult is determined by temperature during key developmental periods, with different temperatures associated with male and female outcomes.[428] Climate change (global warming) has been implicated in altering sea turtle sex ratios.[429] In other organisms, all individuals originally develop into one or the other sex and, as they mature (often growing larger) they can transform into the other sex.[430] In some cases the presence of a mature animal of one sex can inhibit sex change in smaller individuals. As an example, the largest clownfish in a group is typically female; if that female is removed, one of the smaller males will develop into a female (think Nemo). In other species, the situation is reversed, the largest animal is a male, and if this male is removed, one of the (smaller) females develops into a male.[431]

In humans, and most mammals, birds, and reptiles the phenotypic sex of an individual is determined chromosomally, that is by the presence or absence of a "sex chromosomes". The other, non-sex determining chromosomes are known as autosomes.[432] In humans the sex (23rd) chromosome comes in two forms, known as X and Y (←).[433] An XX individual typically develops as a female, while an XY individual typically develops as a male. Most of the X and Y chromosomes are non-syntenic, as you might have suspected given that the Y chromosome is small and contains only ~50 genes, while the X-chromosome contains between 800 and 900 genes. The X and Y chromosomes contain small syntenic regions known as their pseudo-autosomal regions (←). Their organization effects how these chromosomes behave during meiosis.

A difference between X and Y chromosomes in therian mammals (marsupials and placental mammals, including humans), is the presence of the *SRY* gene on the Y chromosome. There is no copy of *SRY* on the X chromosome. The *SRY* gene is not found in monotremes (egg-laying mammals) and other vertebrates.[434] The *SRY*gene appears to have originated in the therian mammal lineage ~150 million years ago and to have been derived from a duplication of a gene encoding a SOX-type transcription factor. It contains a high-mobility group (HMG) DNA binding domain. The presence of a Y chromosome, and so (presumably) an active *SRY* gene, leads to male sexual development. If absent or inactivated by mutation, the individual undergoes female sexual development, even if the Y chromosome is present (→).[435]



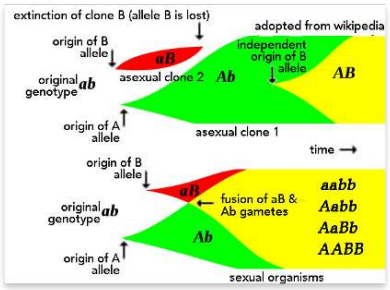
Expression of the SRY transcription factor initiates a down-stream gene regulatory cascade, activating some genes and inhibiting others; the end result is the generation of the various developmental differences associated with male and female anatomy and behavior.[436] In females other genes are expressed and act to inhibit the male differentiation system, just as Sry and its "downstream" targets act to inhibit female differentiation. In molecular studies, it is possible to show the importance of *SRY.* If the *SRY* gene is transferred to an autosome (one of the other chromosomes) it leads to male sexual development. The details of these processes are complex, so we defer further details to more advanced classes.[437] That said, as you can imagine, defects in any of the "downstream" genes and molecular networks can influence phenotypic outcomes. There are other sex determination strategies used by other organisms but we ignore them here.[438]

Essentially all macroscopic organisms, with the possible exception of bdelloid rotifers,[439] reproduce or can reproduce sexually.[440] In contrast to asexual reproduction, sexual reproduction generates genetically distinct organisms, different from either parent. In most cases, sexual reproduction is much more complicated and involves collaboration between male and female organisms.[441] So why is sexual reproduction so common? A simple reason is the generation of genetic variation. Why is variation important? One reason involves the presence of even more rapidly reproducing pathogens. Viruses, bacterial and microbial (eukaryotic) organisms typically reproduce over periods of minutes to hours to days, whereas larger, multicellular organisms reproduce over periods of months, years, and decades. Similarly, but on somewhat longer time scales, the level of genetic variation within a population enables a population to adapt to changing environmental conditions, of which pathogens and predators are a part. Susceptibility to infection by pathogens is itself a phenotype with a genetic component. The genetic variability within a population can serve as insurance against pathogens; even the most lethal pathogens known, viruses like

smallpox and bacteria such as those that cause plague, generally do not kill all of the organisms they infect. Those organisms that survive infection are often immune to subsequent infections, a phenomena that is the basis of vaccination and various other processes, including the CRISPR-Cas9 system of prokaryotes. This view of the selective advantage of sex is often referred to as the Red Queen Hypothesis, since organisms have to "run" constantly, in terms of generating genetic variation, to keep up with their predators, parasites and pathogens.[442]

> *"It takes all the running you can do, to keep in the same place." says the Red Queen to Alice*

Sexual reproduction, specifically the processes of meiosis and fertilization offers mechanisms that generate huge amounts of genetic variation within a population. Sexual reproduction speeds up the appearance of beneficial combinations of alleles, combinations that would take significantly longer to appear if they had to accumulate independently within a lineage (top panel, independent formation of organisms with beneficial AB individuals; bottom panel, sex-mediated formation of AB individuals →). The larger the population size, the more likely that beneficial, in terms of reproductive success, allelic combinations will appear and so facilitate adaptations to a changing environment. The reduction in genetic variation is one of the reasons that reductions in population size have been linked to an increased probability of extinction.[443]



In addition to the generation of variation, the process of sexual reproduction offers mechanisms by which populations can become reproductively isolated from one another, leading to the creation of two species from one. Generally males and females have to cooperate to reproduce; sexual reproduction is a social process. The participants have to be producing functional gametes at the same time, these gametes have to be able to meet each other, recognize each other, and fuse together. The diploid cell that forms has to develop normally producing an organism that can itself form functional gametes, and so on. Incompatibilities in any of these processes can produce a reproductive barrier between the individuals within different populations - that is, speciation. Reproductive barriers can be selected for if individual subpopulations have become well adapted to their ecological niches, while hybrids are not.

Questions to answer:

176. If you were to design a temperature sensitive form of sex determination, how might you go about it?
177. What might happen during meiosis if the regions of homology in X and Y chromosome were removed?
178. What are possible benefits of reproductive barriers and isolation between related populations?

Question to ponder:

- How might variations in sexual behavior come about, molecularly?

References

[427] We will not go into any great detail about hermaphroditic models of reproduction, but this is an interesting paper related to the subject: Sexual selection: lessons from hermaphrodite mating systems.

[428] Environmental sex determination mechanisms in reptiles

[429] Climate change is turning 99 percent of these baby sea turtles female

[430] Phylogenetic Perspectives on the Evolution of Functional Hermaphroditism

[431] Functional hermaphroditism in teleosts

[432] In other species (e.g. birds, some reptiles, and some insects) the system is based on Z and W sex chromosomes. In contrast to the XY system, males are ZZ while females are ZW.

[433] X chromosome regulation: diverse patterns in development, tissues and disease and Y-chromosome

[434] "Environmental sex determination is widely employed in fish, where a range of stimuli from social cues to temperature establishes sex. Temperature sex determination is also extensively utilized in reptiles."

[435] see Molecular Mechanisms of Male Sex Determination: The Enigma of SRY for more details.

[436] In a recent study, the primary sex determination event in humans has been found to be associated with changes in ~6500 genes: see 6,500 Genes That Are Expressed Differently in Men and Women

[437] Sex determination: a primer

[438] The evolutionary dynamics of haplodiploidy

[439] Uptake and Genomic Incorporation of Environmental DNA in the “Ancient Asexual” Bdelloid Rotifer Philodina roseola

[440] C. Zimmer. 2009. On the Origin of Sexual Reproduction

[441] Origins of Eukaryotic Sexual Reproduction: http://cshperspectives.cshlp.org/content/6/3/a016154.ful

[442] see Sexual reproduction as an adaptation to resist parasites

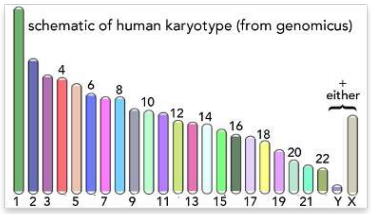
[443] Timing and causes of mid-Holocene mammoth extinction

This page titled 12.3: Sex-determination and its chromosomal basis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

12.4: Steps in meiosis- from diploid to haploid

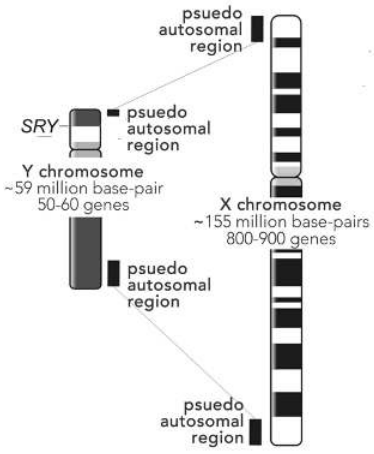
In animals sexual reproduction results in a diploid cell, a zygote, that goes on to generate an new organism. As development proceeds some cells differentiate into what is known as the germ line; the remainder of the organism is known as the soma or body. The germ line goes on to produce haploid cells known as gametes. In some organisms, the haploid (gametic) stage can persist and live independently,[444] but generally the haploid stage of the life cycle is short. After their formation haploid gametes, from two distinct "parents", fuse or die. In some, primarily unicellular, species there are multiple "mating types", and only gametes of different types can fuse. One aspect of the haploid state is that it can reveal the presence, and lead to the elimination, of deleterious recessive alleles. Haploid cells that contain, and are dependent upon the expression of such alleles will be eliminated, removing the alleles from the population, which can have a strong evolutionary effect.[445]

In organisms with multiple mating types, rather than two sexes, the gametes of different mating types are morphologically similar. The energetic investment to produce a gamete is the same for all mating types. The situation is different in multicellular organisms. The gametes of the male and female sex are morphologically different and differ is size. The sex that produces the larger, generally immobile gamete, an oocyte, is known as female (♀) while the sex that produces the smaller, often motile gamete, a sperm or spermatozoa, is male (♂). The difference in the size of the gametes is an example of a sexual dimorphism with evolutionary implications. The two sexes have, from the start discordant investments in the production of gametes (large versus small). This difference can become more pronounced when the two parents differ in their investment in the growth and nuturing of offspring, a fact that underlies sexual selection, a feature of modern (Darwinian) evolutionary theory.[446]



In females meiosis typically generates a single gamete, known as an oocyte or egg, and three non-viable mini-cells, known as polar bodies. In males, meiosis produces four gametes - sperm. Each gamete contains one and only one copy of each autosomal chromosome present in the original diploid cell. Historically, chromosomes were numbered based on their apparent size in histologically stained specimens. In humans, the largest chromosome, chromosome 1, contains ~250 million base pairs of DNA and over 2000 polypeptide-encoding genes, while the smallest, chromosome 22 contains ~52 million based pairs of DNA and ~500 polypeptide encoding genes (→). Homologous chromosomes are also defined by the order of genes found along their length. Human chromosome #5 contains different genes than are found on chromosome #6. Moreover, the maternal (from the mother) version of each chromosome can contain different alleles of these genes compared to those found in the paternal version.

In mammals (humans) males have both an X and a Y chromosome; meiosis generates four gametes that each contain one copy of each autosomes and either an X or a Y chromosome. Females have two X chromosomes, so all gametes they produce contain an X chromosome. A male gamete fuses with a female gamete to form a new diploid cell, a new organism. If the male gamete contains a Y chromosome, the new (diploid) organism is chromosomally male, if the male gamete contains an X chromosome, the new organism is chromosomally female.[447] The fusion event, known as fertilization, is the most discontinuous event in the process of sexual reproduction. Even so, fertilization does not represent a true discontinuity, at least with respect to life – both sperm and egg are alive, as is the fertilized egg.[448] In a critical sense life (in the post-LUCA world) never begins – it continues and is transformed. That said, fertilization is the start of a new, genetically distinct organism. The fused cell that results from fertilization is known as a zygote. Through somatic (asexual) cell division (mitosis and cytokinesis) the zygote will develop into an adult, composed of diploid cells. To repeat, the cells that produce gametes are known as germ cells, and are produced by the organism's germ line. The rest of the adult is composed of somatic cells, cells that divide (if they divide) by mitosis. Meiosis is restricted to the germ line cells that produce gametes.



References

[444] see wikipedia – gametophyte: https://en.wikipedia.org/wiki/Gametophyte

[445] see: Evolution of haploid selection in predominantly diploid organisms and Haploid selection in animals

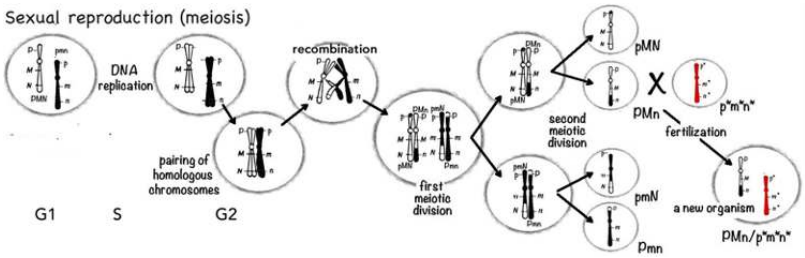
[446] How Darwin arrived at his theory of sexual selection and Mate choice and sexual selection since Darwin?

[447] While we not deal in detail with this topic, aspects of gender are complex traits: see Beyond XX and XY: The Extraordinary Complexity of Sex Determination

[448] In fact, there are examples of cell fusion within organisms - as an example, during the development of skeletal muscle, muscle precursor cells fused to generate large multi-nuclear cells, known as myotubes.

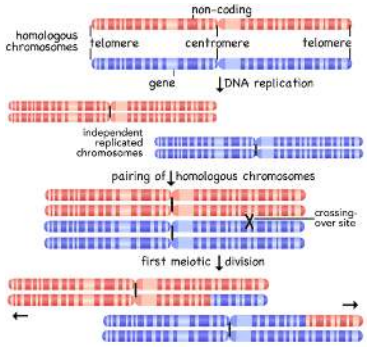
This page titled 12.4: Steps in meiosis- from diploid to haploid is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

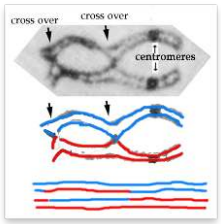
12.5: Recombination and independent segregation

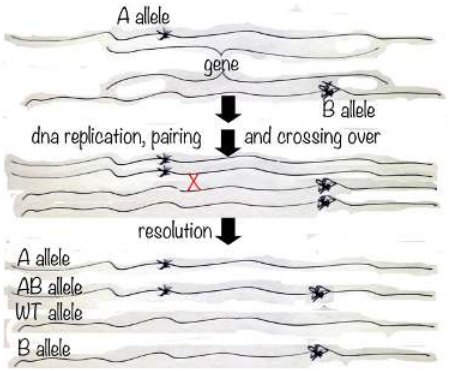


We begin our description of meiosis (↓) with a diploid germ line cell that contains two copies of each autosome and, in mammals, either two X chromosomes or an X and a Y chromosome. The chromosomes delivered to the fertilized egg and subsequent zygote by a female gamete are known as maternal chromosomes, while those delivered by the male gamete are known as as paternal chromosomes. The maternal and paternal chromosomes are known as homologs. To generate gametes, a diploidgerm-line cell enters meiosis (see video link). Meiosis consists of a single round of DNA replication followed by two rounds of cell division.

As the diploid germ line cell enters meiosis it moves from G1 into S, just as in mitosis. Each of its individual chromosomes is duplicated. The resulting replicated (double-stranded) DNA molecules remain attached to one another through a structural complex known as the centromere. Here is where meiosis diverges from mitosis. In an asexual (mitotic) cell division each replicated chromosome remains independent of its homolog and each replicated chromosome interacts independently with the mitotic spindle through its centromere and associated kinetochore complex. In meiosis, during G2 the now duplicated homologs (the maternal and paternal chromosomes) align with one another (→). These four DNA molecules are known historically as a “tetrad”; each consists of four double-stranded DNA molecules. The pairing of the homologous chromosomes is based on the association of syntenic chromosomal regions.[449] The DNA sequences along the homologous chromosomes, while not identical, are extremely similar, with the same genes located in the same order on each. When they are not, due to chromosomal rearrangements, things can get messy - as we will see. After chromosome pairing, and at essentially random positions along the length of the chromosomes, "crossing-over" or recombination events can occur. A DNA endonuclease, produces double-strand breaks in two of the four (double-stranded) DNA molecules (for example, at the site marked by “X” above ↑).[450] The DNA molecules are then rejoined, either back to themselves (maternal to maternal, paternal to paternal) or to the other DNA molecule. Maternal to maternal or paternal to paternal crossing over events are generally invisible and have little impact. Crossing-over between homologs (paternal to maternal or maternal to paternal) lead to visible (←) crossing-over events. Typically, multiple “cross-over” events occur along the length of each set of paired (replicated) homologous chromosomes. Whenever maternal-paternal crossing over occurs the resulting "recombinant chromosome" contains a different set of alleles than either the original chromosomes. You can convince yourself by following any one DNA molecule from beginning to end.



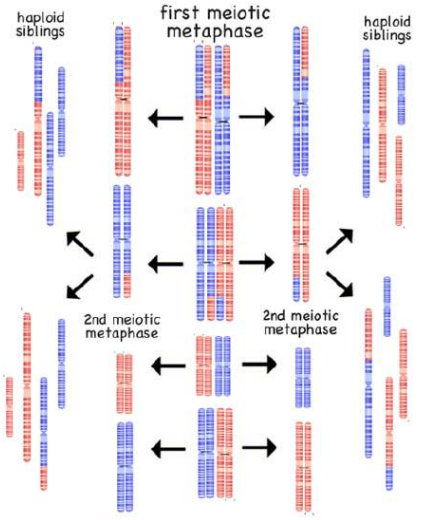




In addition to shuffling alleles, crossing over can create new alleles. Consider the situation in which two alleles of a particular gene are different from one another (→). Assume that each allele contains a distinct sequence difference (as marked). If, during meiosis, a crossing over event takes place between these sites, it results in one allele that contains both molecular sequences (AB), and another allele with neither (indicated as wild type "WT"). A new allele (AB) has been created, without a new mutation!

In the case of the X and Y chromosomes, the chromosomes pair with one another through their common pseudo-autosomal regions (see above), which are syntenic. Outside of these regions there is no significant synteny between the X and Y chromosomes, so crossing-over does not occur. In contrast, crossing occurs normally (that is, just like for autosomes) between the two X chromosomes in females.

Meiosis leads to yet another source of variation. At the first meiotic division, the duplicated (and recombined) homologous chromosomes remain attached at their centromeres; each of the tworesulting daughter cells receives either the duplicated maternal or paternal chromosome centromere region. However, what set of chromosomes (defined by their centromeres, maternal or paternal) they inherit is determined by chance. The process is known as the independent assortment of homologous chromosomes during the first meiotic division, or independent assortment for short (←). For an organism with 23 different chromosomes (such as humans), the first meiotic division can produce 2^{23} (8,388,608) different daughter cells.



There is no DNA replication between the first (M1) and the second (M2) meiotic divisions. During the second meiotic division each replicated chromosomes, held together at their centromere, attach to the spindle, very much as in mitosis. But because of earlier recombination events, the two chromosomes are no longer necessarily identical. This further increases to rather astronomical levels the number of different chromosome sets a particular haploid gamete can inherit. When the chromosomes separate, each of the two resulting sibling cells receives one and only one copy of each chromosome. Which particular chromosomes they inherit is stochastic. In males, all four haploid cells form gametes; they differentiate into sperm, that can fuse with an egg cell. In females only one of the four haploid cells forms a gamete, differentiating to form an oocyte that becomes a haploid egg that can fuse with a sperm cell. The other three haploid cells produced are known as polar bodies. Polar bodies do not fuse with sperm, they donate their cytoplasm to the oocyte - supporting the development of a large, fertilizable gamete.

The result, and basically the point, of meiosis is to generate a population of gametes in which the alleles present in the maternal and paternal chromosomes have been shuffled in various ways, so that any resultant offspring have a unique genomes related to, but distinct from that of either of their parents.[451] Fertilization (the fusion of gametes) combines two such genomes, one maternal and one paternal, to form a new organism, with a novel combination of alleles generated through independent assortment of homologous maternal and paternal chromosome and recombination events. Most phenotypes are influenced, to a greater or lesser degree, by the set of alleles within a genotype, and new combinations of alleles will lead to new phenotypes and phenotypic variations that can impact reproductive success, and so lead to evolutionary effects.

Questions to answer:

179. Consider the odds of an organism obtaining the three new mutations necessary for the appearance of a new trait. Predict which would be faster (in terms of the number of generations required) in achieving this goal, sexual or asexual reproduction and why.
180. You are working with an organism with five autosomes and one sex chromosome. Considering only the effects of independent assortment during meiosis, how many different types of gametes could be generated? A drawing of the process could help.
181. Indicate (in a drawing and associated explanation) how a deleterious mutation within a gene could be generated by or eliminated through recombination.
182. Would genetic diversity be altered if meiotic recombination occurred during meiosis II, rather than meiosis I?

Questions to ponder

- Under what conditions might you expect the evolution of sexual reproduction to be selected against.
- Why are parents and their siblings not necessarily good donors for organ transplantation?

References

[449] Synaptonemal complex formation: where does it start?

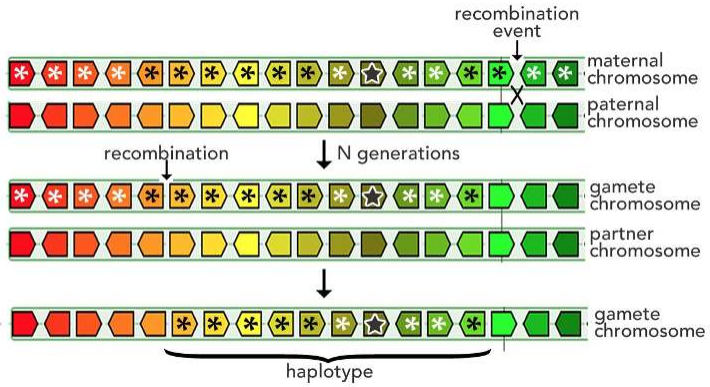
[450] adapted from The Centenary of Janssens's Chiasmatype Theory Koszul et al., 2012. Genetics **191:** 309-317.

[451] This even applies to hermaphrodites, in which one organism acts as both mother and father!

This page titled 12.5: Recombination and independent segregation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

12.6: Linkage and haplotypes

An important feature of meiotic recombination is that it can “disconnect” the alleles of genes located near one another along a chromosome. Consider the situation when a mutation occurs that creates a new allele in gene X; let us call it X^{select}. Now let us assume that this allele is subject to strong positive or negative selection. That means that the presence of the X^{select} allele in an organism has a strong effect on its reproductive success. Because it is either strongly (positively) selected for or (negatively) selected against. The frequency of the allele will tend to increase or decrease in subsequent generations. The change in the frequency of the X^{select} allele will also influence the



frequency of alleles of genes located near to the X gene on the chromosome. Selection of X^{select} will impact the the alleles neighboring "linked" genes. This can increase the frequency of linked alleles that are (mildly) deleterious or decrease the frequency of alleles that are, in fact beneficial. The closer the genes are to each other along the chromosome, the longer (over more generations) such linkage effects can be expected to persist. Why? because the probability of recombination between two sites along a chromosome is a function of how close they are to one another along the chromosome. As the distance increases, the probability that a recombination event will occur between them increases. When the probability of a recombination event between two genes reaches 50% or greater per meiotic division; the genes behave as if they are on different chromosomes – they are “unlinked.”

Linkage distances are calculated in terms of centimorgans, named after the geneticist Thomas Hunt Morgan (1866-1945).[452] A centimorgan corresponds to a 1% chance of a crossing over event between two specific sites along a chromosome. In humans, a centimorgan corresponds to ~1 million base pairs of DNA, although this value varies in different regions of different chromosomes. Two genetic loci that are 50 or more centimorgans apart are separated by ~50 million or more base pairs. In meiosis two genetic loci on the same chromosome, but separated by >50 centimorgans, have the same probability of being inherited together as if they were on two different chromosomes. We will return to this again, when we consider the interpretation of genetic crosses.



Consider a particular allele of a particular gene, marked by the star (★)(→). Let's assume that this allele is associated with a visible trait. We mark the alleles found in neighboring genes on this chromosome with asterixes (*). For the sake of clarity assume that different alleles (un-marked) are found on the homologous chromosome. During meiosis, recombination events occur randomly across the chromosomes. Over time multiple recombination events occur and reduce the size of the region of the original chromosome containing the ★ allele. This region is known as a haplotype; it is a group of alleles that are inherited together from a single parent. From a formal point of view, it is not clear which variation within the haplotype region is responsible for the trait observed, since they are inherited together. In the era of genetic (pre-molecular biology) days, multiple rounds of crosses (breeding cycles) were required to identify where, exactly along a chromosome the allele responsible for a particular trait was located. With each generations, the size of the haplotype region becomes smaller.

Now consider how the alleles within a particular region can be maintained together. Let us assume that the original allelic variant has effects on the expression of neighboring genes (→); how might this occur? Two obvious mechanisms suggest themselves: the allele could influence the packaging of the chromosomal region, so that the genes’ accessibility to regulatory factors is modified or the allele can itself effect or be in an gene regulatory element (an enhancer) that plays an important role in the regulation of multiple genes in its molecular neighborhood. Both options could lead to selective effects based on the maintenance of the integrity of the chromosomal region, the haplotype - that is, recombination events within the region can occur, but if they have a negative effects on reproductive outcomes they could be selected against.

Questions to answer:

183. Graph, as a function of distance, the likelihood that recombination will disconnect a selected (whether positively or negatively) allele from alleles in surrounding genes.
184. Why might a crossing over event inhibit nearby crossing over events?
185. How can you use the size of a conserved genomic region to estimate time of isolation of a population?
186. What are the benefits of recombination in terms of environmental adaptation?

Questions to ponder:

- How does the size of haplotype regions reflect the reproductive history of a population?
- How does the presence of a deleterious allele influence the selective pressures on an organism? How might it open up time, new evolutionary possibilities?

References

[452] Thomas Hunt Morgan

This page titled 12.6: Linkage and haplotypes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

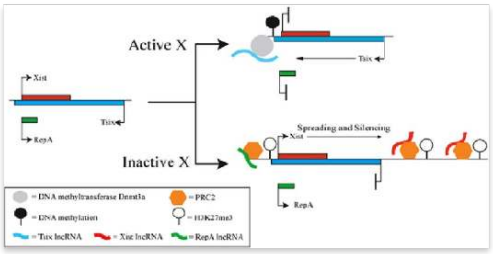
12.7: X-inactivation and sex-linked traits

One aspect of the XY chromosome-based system of sex determination is that the two sexes have different genotypes with respect to these chromosomes. As noted, the Y chromosome is short and encodes relatively few genes, while the X chromosome is longer and encodes many more genes. This creates a genetic imbalance between the two sexes in terms of gene copy numbers. A single gene can direct the synthesis of only so many RNA molecules per unit time, based on the rate of RNA polymerase binding, activation, and RNA synthesis. This is one reason for haplo-insufficiency, a phenomena associated with genes on autosomes, where a null allele leads to a dominant phenotype because a single functional copy of the gene does not produce enough gene product. Without



some "balancing" mechanism, we would predict that female cells would have about twice as many RNAs for genes on the X as do similar cells in a male (and most cells in males and females are, in fact, similar). There seems to be a need for a form of "dosage compensation"; either genes on the X in males have to be expressed more efficiently or genes on the X in females should be expressed less efficiently. The strategy used in humans and many other placental mammals is a process known as X-inactivation. Early in embryonic development, in a stochastic event one or the other of a female's X chromosomes becomes associated with specific RNAs and proteins; this results in that copy of the X being packed into a compact structure that can no longer support gene expression.[453] Once the choice of which X chromosome to inactivate is made, it is stable and inherited through subsequent mitotic cell divisions, generating clones of cells with one or the other X chromosome active (and the other inactive). A failure of X-inactivation generally leads to developmental arrest and embryonic death in female embryos. While gene expression from the inactivated X is inhibited, the replication of the inactivated chromosome continues with each cell cycle. We can see the effect of this X chromosome choice in female calico cats, in which the different coat colors reflect domains in which one or the other X chromosomes is actively expressed, while the other X chromosome is inactive (→). As you may have already deduced, a gene involved in the generation of coat color is located on the X chromosome.

The X-chromosome inactivation system involves of two genes, *XIST* and *TSIX*. *XIST* encodes a functional ~19.3 kilobase long non-coding (Lnc) RNA; it does not (as far as is currently known) encode any polypeptides (next↓page). XIST is expressed only in cells with two X chromosomes – so it is not expressed in males.[454] Which of the two X-chromosomes expresses *XIST* is determined stochastically during embryonic development. When expressed, the XIST RNA associates with regions adjacent to the *XIST* gene and eventually comes to localize along the entire length of the X-chromosome on which the active *XIST* gene is located. The XIST RNA associates with a number of protein complexes involved in inhibiting gene expression and producing the compact state of the inactivated X, known as a Barr body, named its their co-discoverer Murray Barr (1908–1995).



On the DNA strand anti-parallel to the *XIST* gene is an over-lapping gene known as *TSIX* (↓). The *TSIX* gene on the active X-chromosome is expressed. *TSIX's* promoter is distinct from that of *XIST*; expression of *TSIX* is expected to interfere with *XIST*expression. The *TSIX* gene encodes a ~40 kilobase lncRNA that is partially complementary to the XIST RNA. TSIX RNA acts to inhibit XIST activity; when present it inhibits the action of XIST on the active X chromosome, blocking inactivation of the chromosome actively expressing TSIX. Together the XIST/TSIX system ensures that one and only one of the two X chromosomes is active in a particular cell.

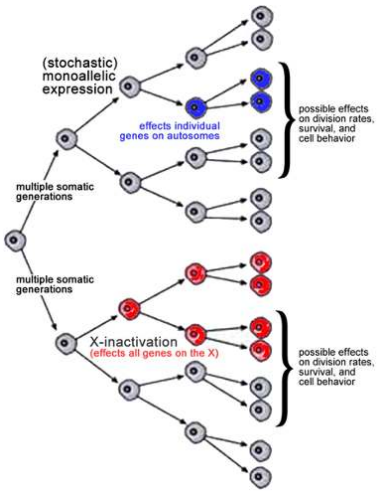
References

[453] X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos

[454] X-inactivation-specific transcript (OMIM)

This page titled 12.7: X-inactivation and sex-linked traits is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

12.8: X-linked diseases and mono-allelic gene expression



While calico spots occur only in female cats, there are a number of genetic susceptibilities that are more commonly seen in males. These arise because males have only a single X chromosome. The result is that, in contrast to the rest of the genome, genes on the X are effectively haploid in males. The result is that phenotypes associated with recessive alleles of genes located on the X chromosome are more often visible in males. In females, who are formally heterozygotic for such genes, some cells express one allele while others express the other. The female is mosaic, composed of clones expressing one or the other allele on the X chromosome.

In a typical diploid cell, it is sometimes the case that one gene is active while the other copy of the gene, on the homologous chromosome is inactive, due to stochastic "transcriptional silencing" events.[455] In some cases of such stable monoallelic expression there is what is known as somatic selection, which we will return to. Given that there are two alleles, when they are different which one is expressed may influence cell growth, division, and even survival, so that over time, cells expressing one allele may come to dominate (in numbers) those that express the other (→) in particular regions of the body. The extent to which random monoallelic expression influences human development and disease is just now being recognized and examined carefully.

Questions to answer:

187. What does it mean to be mosaic for an allele?
188. Why do males and females differ in the traits they display?
189. Why do males and females differ in the display of phenotypes associated with genes on the X chromosome?
190. Can you provide a plausible mechanism to explain why (autosomal) random monoallelic expression occurs?
191. How might monoallelic expression impact an organism?

Question to ponder:

- Under what conditions might monoallelic (autosomal) gene expression be beneficial?



Short Chapter Summary

- Cell cycle controls accurate genome duplication and segregation (mitosis).
- Meiosis halves ploidy and reshuffles alleles (independent assortment + recombination).
- Linkage, haplotypes, X-inactivation, genetic background and stochastic effects complicate trait inheritance.

References

[455] Monoallelic Gene Expression in Mammals

This page titled 12.8: X-linked diseases and mono-allelic gene expression is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

13: How Alleles Arise - Mutations

13.0: Introduction

13.1: Mutations into alleles

13.2: Luria and Delbrück- Discovering the origin of mutations

13.3: Forward and reverse genetics

13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies

13.5: Longer term mutation and evolution studies

This page titled 13: How Alleles Arise - Mutations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

13.0: Introduction



In which we consider how mutations (changes in genomic DNA) appear and become alleles within a population and distinguish between the effects of chance and selection on allele frequencies.

We are far enough along to recognize that beginning with a particular genome any change, such as those due to errors in DNA replication as well as un-repaired or mis-repaired environmentally induced damage involving chemical reactions or the effects of radiation results in a mutation. If the cell/organism survives the mutational event and if the mutation lies within a gene, it can produce what is known as an allele, a specific version of a gene. If it lies outside of a gene, it becomes known as a sequence polymorphism or sequence variation. With the advent of economical genome sequencing and related technologies, it is possible to estimate the rates of mutation in a particular organism or in a particular cell type.[456] Here we distinguish between mutations in the germ line (leading to eggs and sperm) that can be inherited by offspring and those that occur in somatic cells, the cells of the body. Mutations not inherited from one's parents are said to occur spontaneously. A spontaneous mutation may be passed on to offspring only if it occurs in a germ line cell.

As a first approximation, mutations occur randomly within genomes, although there are known mutational "hotspots". CpG dinucleotides are mutated ~10X more frequently than other dinucleotides. In addition to single nucleotide changes, mutations can involve small insertions and deletions of nucleotides known as indels. Indels are defined to be less than 20 base pairs (bps) in length to distinguish them from larger changes, known as structural variants; these can be thousands to millions of bases in lenght. It has been estimated that each generation sees the addition of ~3 indels and ~0.16 structural variants in a person's germ line. Another type of structural variant, known as a copy number variation (CNV), involves changes in the number of copies of a particular DNA region. It can lead to multiple copies of gene(s).[457] There are also mutations influencing cell division that can lead to the loss or gain of a chromosome or the duplication of the entire genome.

Mutations occur more frequently in the soma of an organism than in the germ line because there are more cells (trillions) and so more cycles of DNA replication and cell division involved. Similarly, there are fewer cell divisions involved in the generation of oocytes than in the generation of sperm and the number of mutations in the male germ line increases with age. Germ line mutations can be passed from generation to generation, while somatic mutations are lost on the death of the organism. A current estimate is that the chance of a *de novo* germ line mutation in humans is ~$1x10^{-8}$ per base pair per generation (the human genome contains ~6 x 10^9 base pairs of DNA). We will discuss both germ line and somatic mutations and their effects soon.

References

[456] see: The origins determinants, and consequences of human mutations

[457] Copy Number Variation & Indels

This page titled 13.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

13.1: Mutations into alleles

For a mutation to become an allele within a population the first criterion is that it does not produce an early (pre-reproductive age) lethal phenotype - that is a phenotype that results in the death of the organism before it can produce offspring. Why? In a diploid organism a new mutation will involve only one of the two genes present. For it to have a phenotype, it needs to be "dominant" over the other allele present. Of course this is not the case in prokaryotes, which are effectively haploid. If the mutation is not dominant lethal and if it occurs in the germ line, a gamete can carry it into the next generation where it has a chance to persist within the population (assuming that it does not render the organism carry it sterile. Again, this assumes that the presence of a single (dominant) copy of the allele does not result in a lethal phenotype in gametes or the early embryo, since where and when a gene is expressed has a lot to do with the phenotypes it is associated with and influences.

A new non-lethal dominant or recessive mutation has to avoid elimination through the stochastic effects of meiosis and genetic drift. Remember that when it first appears in the germ line of a sexually reproducing organism there is only one copy of the mutated allele present in the population. It is possible that gametes carrying the new allele will fail to find or fuse with another gamete to form a new organism – if so, the mutant allele will be lost. Similarly, the mutant allele may make it into the next generation if it is not too deleterious, just by chance.

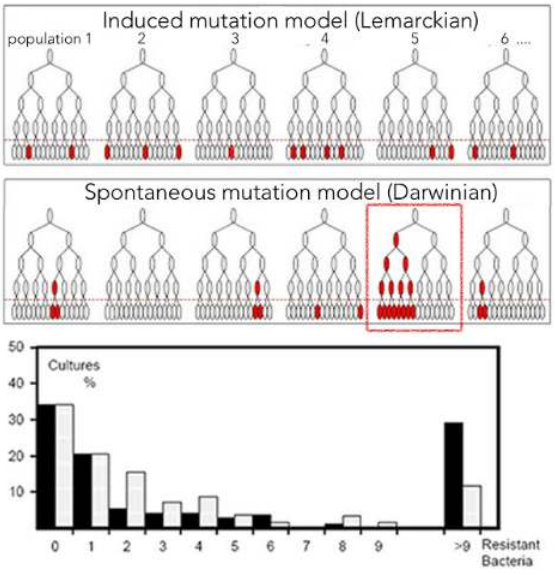
If a mutated gene survives these early events, it comes to be referred to as an allele, particularly when it is found in >1% of the population. Mutations outside of a recognized gene are known as polymorphisms. What is, and what is not, part of a gene, can be tricky to determine; not all sequences involved in the regulation of gene expression may be known. The total genetic variation within a population, the sum of alleles and polymorphisms reflects the population's past history; the combination of selective pressures and non-adaptive events, such as founder effects, bottlenecks, and genetic drift, and serves as the basis for subsequent evolutionary change.

This page titled 13.1: Mutations into alleles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

13.2: Luria and Delbrück- Discovering the origin of mutations

Darwin and Wallace lacked a clear understanding of where genetic variation came from, how it might be stored or passed on from one generation to the next. An important question that arose early in the history of evolutionary studies was whether mutations were due to chance (stochastic) events or whether they were somehow purposefully generated in response to the "needs of the organism". Darwin assumed that evolution involved random variations that arise in individuals. A Lamarckian mechanism involves induced (and purposeful) responses by individuals.[458] In the absence of a clear understanding of how genetic information is stored, replicated, and passed from generation to generation, there was really no way to distinguish between Darwinian and Lamarckian mechanisms, although Lamarckian mechanisms seemed more direct.[459]

To understand how this question was resolved, consider the classic experiment, known as the Luria-Delbrück experiment after Salvador Luria (1912-1991) and Max Delbrück (1906-1981), the scientists who carried it out.[460] Their study was published in 1943, before DNA was recognized as the genetic material and well before anyone understood how genetic information was stored.[461] Luria and Delbrück examined the resistance of bacteria to viral infection. They used bacteria that could be infected and killed by a specific bacteriophage. Mutations arose spontaneously. When they appeared they rendered the bacteria, and their off-spring, immune to viral infection. The question Luria and Delbrück asked was, do phage resistance mutations appear randomly all of the time or does the presence of the virus 'induce" their appearance. Is immunity learned or lucky?[462] If the generation of phage resistance mutations is adaptive, then we would expect that the frequency of resistance producing mutations will be more or less uniform from one population to the next – repeating experiments on different cultures should produce resistant bacteria at approximately the same rate in each (top panel next↓page). If, on the other hand, the mechanism occurs by chance, that is stochastically (middle panel ↓), we would expect that the number of mutational events will vary dramatically from one population (culture) to the next - the variation in the frequency of phage resistance (and the mutations that produce it) between independent populations will be large.



Luria and Delbrück started a number of bacterial cultures to which they then added enough virus (at the time of the horizontal red line in the top two panels) to kill every sensitive bacterium. They then plated out the cultures and counted the number of phage-resistant bacteria present, each of which grew up into a macroscopic clonal colony. The number of phage resistant cells in a culture reflects when, in the history of the culture, the resistance mutation appeared. If the resistance mutation appeared early in the history of the culture, as in the red-boxed culture (→) it would be common, whereas if it appeared late, it would be rare. The two models (induced/Lemarckian versus random/Darwinian) make dramatically different predictions. In the induced/Lemarckian model, the variation in the numbers of resistant bacteria between cultures is expected to be low, since resistance arises through a common “inductive”, physiological process, even though we do not know how that process works. In contrast, in the spontaneous/Darwinian model we expect large variations, with many cultures having no resistant bacteria and some having many. When the mutation occurs late, or not at all, as in lower panel, population 2, there will be few or no phage resistant cells. If the mutation occurs early there will be many resistant bacteria. Luria and Delbrück calculated what the two models predicted. The observed results (black bars) matched the prediction for the spontaneous/Darwinian mechanism, leading them to conclude that, at least in this system, mutations occurred independently of the presence of the virus.

To date there is no evidence that environmental factors can specifically induce the generation of beneficial or useful mutations. What can happen, however, is that the general (non-specific) mutation rate can increase in response to various stress conditions, arising from internal or environmental effects. Typically an increased mutation rate involves effects on the efficiency of DNA error repair systems, which leads to increased levels of genetic variation upon which selection can act.[463]

The ability to control mutation rates occurs within the vertebrate immune system, through a process known as somatic hypermutation.[464] This process is involved in the maturation of the immune response and the generation of increasingly specific

antibodies, a topic well beyond our scope here. That said, the mechanism is known; these cells activate a gene that encodes an "activation-induced deaminase" or AID (OMIM:605257). The AID protein acts on cytosine residues in DNA to generate uracils that, when repaired, replace the original C:G base pair with an A:T base pair. The other genes in these cells appear to be at least partially protected by "selective targeting of AID and gene-specific, high-fidelity repair of AID-generated uracils".[465]

References

[458] This is perhaps one reason that collectivist ideologies, such as the Soviet Union under Stalin, so disliked Darwinian evolution (and harshly prosecuted geneticists). see http://blogs.plos.org/scied/2017/04/...itics-marches/

[459] This led to what was known as the "Eclipse of Darwinism"; biology emerged from this "darkness" with the development of an understanding of genes and genetic mechanisms to produce what became know as the "Modern Synthesis".

[460] Luria–Delbrück experiment

[561] Mutations of bacteria from virus sensitivity to virus resistance: http://www.genetics.org/content/genetics/28/6/491.full.pdf

[462] As we will see later on, there are molecular mechanisms, such as the CRISPR-CAS9 system that can learn and lead to acquired immunity.

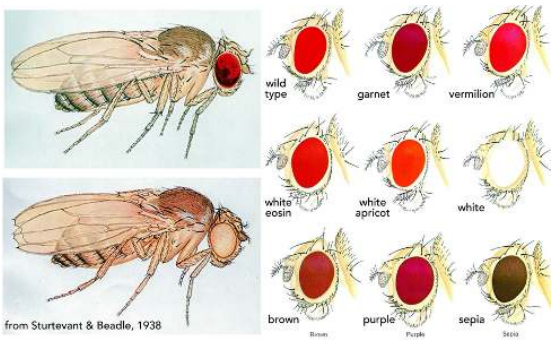
[463] A trade-off between oxidative stress resistance and DNA repair plays a role in the evolution of elevated mutation rates in bacteria

[464] Somatic hypermutation: wikipedia

[465] Two levels of protection for the B cell genome during somatic hypermutation

This page titled 13.2: Luria and Delbrück- Discovering the origin of mutations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

13.3: Forward and reverse genetics

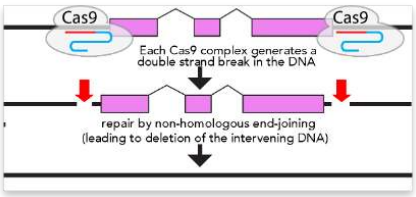


Originally, genetic analyses were carried out through what is now known as forward genetics. This involves the generation of mutations by chance and then identifying those individuals that carry mutations that disrupt a particular process or structure of interest. As an example, consider eye shape or color in the fruit fly *Drosophila melanogaster* (→); these are traits that are experimentally accessible because a *Drosophila* embryo can develop into a fertile adult without eyes. It is therefore possible to identify mutant alleles that alter the eye but allow other aspects of embryonic development to occur (more or less) normally, at least in the context of the laboratory. When we think about a particular trait or behavior, a specific phenotype, we want to know how many different genes are involve in producing or influencing that phenotype. On the other hand, if the product of the mutated gene plays multiple roles in the developing organism, perhaps in processes distinct from those involved in the formation of the eye, the embryo may die before eyes form and no mutations in that gene will be recovered, even though the gene's product plays a key role in eye development or pigmentation. It is for this reason that forward genetic screens for mutations that influence a particular process are never "complete", that is, they cannot identify every gene/gene product involved in a process.

This page titled 13.3: Forward and reverse genetics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies

While early geneticists worked with forward genetics, this approach generally fails to generate a complete map of all of the genes involved in a particular process. Alternative approaches are used to determine whether a specific gene is involved in a particular process. We leave descriptions of these methods for more advanced courses. We will, however, consider CRISPR-Cas9 mediated mutagenesis, which is based on one of a number of anti-viral infection systems found in prokaryotes.[469] In 2020, Emmanuelle Charpentier (b. 1968) and Jennifer Doudna (b. 1964) won the Nobel prize in Chemistry “for the development of a method for genome editing”. The Cas9 enzyme is an endonuclease that creates double-stranded breaks in DNA. What makes the system uniquely useful is that the site at which the endonuclease cuts the DNA is determined by a ~23 base pair RNA sequence, a guide RNA (gRNA) – this sequence is long enough to (often) occur once and only once within the genome of an organism, even an organism with a genome of more than a billion base pairs, such as humans. This gives the system an extremely high degree of target specificity. Versions of the system have been engineered to catalyze base changes at the target site, rather than cutting the DNA.[470] In the DNA cleavage system, the cell's DNA repair systems act to join the two ends of the cleaved DNA molecule back together again, but this joining is rarely accurate – base pairs can be lost or added, generating a mutated form of the original DNA sequence. If the gRNA sequence is present in both alleles of a gene, both alleles can be mutated at the same time. One variation, to ensure that a specific DNA region is removed, is to use pairs of gRNAs (→). If the CRISPR-Cas9 system is activated (or introduced) early in the development of an organism all or most cells can be mutated, which can lead to multiple phenotypes. Alternatively, it is possible to activate the system only in specific cell types, or at specific times of development, allowing for finer experimental control.



References

[469] Overview reference for the CRISPR-Cas9 system: wikipedia. The ADDGENE CRISPR website is useful - link.

[470] The next generation of CRISPR–Cas technologies and applications

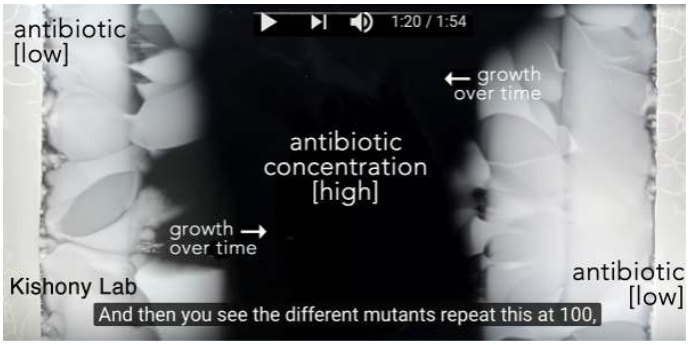
This page titled 13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

13.5: Longer term mutation and evolution studies

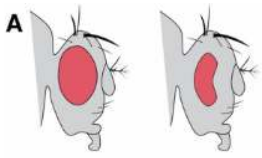
We can see the spontaneous mutation model applies throughout the biological world. Wherever we look mutations appear to arise by chance. If they persist within the population (see above), they become alleles. It is worth reiterating that because of non-adaptive processes such as genetic drift, new neutral or beneficial mutations may be lost because they are initially extremely rare within the population, while mildly deleterious mutations can become fixed by chance alone.

To study such evolutionary processes in a laboratory setting is not easy, but a now classic example of such a study was carried out by Richard Lenski (b. 1956) and his associates. They have been growing twelve originally identical populations of the bacteria *E. coli* for more than 25 years and 60,000 generations.[471] One, of many, characteristics of *E. coli* that distinguish it from other bacteria is that it is unable to metabolize citrate in the presence of O_2. In the course of their studies, Blount et al observed the appearance of *E. coli* variants that could metabolize citrate in the presence of O_2; a beneficial evolutionary adaptation, since it provided those cells with a previously un-utilized energy and carbon source.[472] By tracking backward, the investigators identified a "pre-disposing" mutation that occurred in this lineage around generation 20,000. The presence of this mutation made it more likely that subsequent mutations would enable cells to grow on citrate, producing a Cit^+ phenotype. Molecular analyses indicated that the initial Cit^+ phenotype, which appeared around generation ~31,500, was weak and involved a ~3000 bp genomic duplication that led to increased expression of the *citT* gene. *citT* encodes a protein involved in the import of citrate into the cell. Subsequent studies identified mutations in other genes in the Cit^+ strain that further improved the cells' ability to metabolize citrate.[473] One of these mutations led to increased expression of *DctA*, a gene that encodes a membrane transport protein that increases the cell's ability to import various nutrients normally released into the media, giving the cell a reproductive advantage when grown on citrate. An interesting aspect of these studies was the backlash from some creationists, who rejected the possibility of the evolution of new traits via mutation and selection.[474]

A second more recent study on bacterial evolution, involved looking at the evolution of antibiotic resistance. It involved a giant agar plate (a "megaplate") and an antibiotic gradient (↓). Bacterial cells were placed in the regions free of antibiotic and over time their ability to grow into regions of higher and higher antibiotic concentrations was visualized directly (video link). It is possible to watch the emergence of new variants at the boundary regions as new mutations arise.[475]



An important point to recall about such bacterial evolution studies is that these organisms are reproducing asexually, as clones. That means that they do not interbreed with other organisms in the population, but it also means that (in the absence of horizontal gene transfer) all mutations necessary for a phenotype need to occur independently in a single clonal population. If such mutations lead to a reproductive advantage they can, barring accidental death, take over the population – a process known as a reproductive sweep. This can lead to the loss of alleles present in other clones within the population. If these lost alleles were useful (that is enhanced reproductive success), they would need to appear again, independently, through random mutation and selection (or be transferred horizontally, something that is not occurring in this system). In sexually reproducing organisms, alleles from different individuals can be mixed to more rapidly produce beneficial phenotypes.



Questions to answer:

198. How can a "predisposing mutation" influence the direction(s) of future evolution?
199. In the antibiotic resistance video (watch!), why is there often (but not always) a delay before the bacteria grow into a region of higher antibiotic resistance? What is your model for this behavior and how might you test it?
200. How might the presence of horizontal gene transfer impact the megaplate experiment?
201. How might an evolutionary sweep effect a human population?

Question to ponder:

- How would evolution be altered if the mutations (alleles) were induced in a Lemarckian manner rather than selected?



One-Page Summary

- Mutation is ubiquitous and mostly random with respect to need; selection sorts outcomes.
- Forward/reverse genetics, long-term evolution, and CRISPR reveal mechanism → phenotype paths.

References

[471] *E. coli* long-term evolution experiment: wikipedia and the Lenski lab's *E. coli* Long-term Experimental Evolution Project site

[472] see Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli.

[473] see Genomic analysis of a key innovation in an experimental Escherichia coli population.

[474] The evolution of citrate metabolizing *E. coli*: the "Lenski affair"

[475] Baym et al., 2016 Spatiotemporal microbial evolution on antibiotic landscapes.

This page titled 13.5: Longer term mutation and evolution studies is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

14: Somatic Mutations and Genome Dynamics

14.0: Introduction

14.1: Rates and effects of somatic mutation

14.2: Non-disjunction- aberrant chromosome segregation

14.3: Genome dynamics

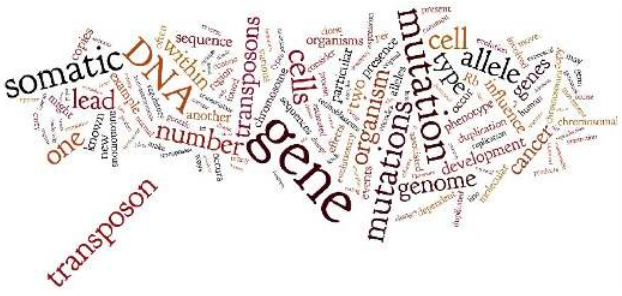
14.4: Gene duplications and deletions

14.5: Orthologs and paralogs

14.6: Transposons- moving DNA within a genome (and weird genetics)

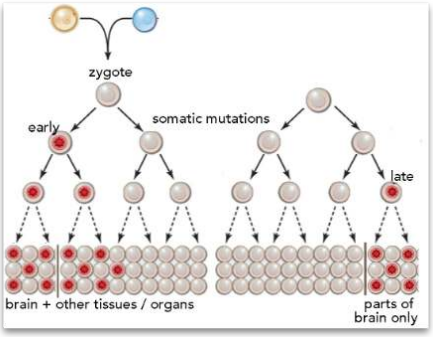
This page titled 14: Somatic Mutations and Genome Dynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

14.0: Introduction



In which we consider how mutations can influence cellular behaviors, lead to disease, and how genes can move around within the genome, often involving the actions of transposable elements.

Up to this point, we have considered mutations that become alleles inherited from one's parents. These parental alleles are shuffled during meiosis and upon the fusion of haploid gametes. Now we consider somatic mutations, mutations that occur during an organism's lifetime. While one's somatic cells contain the same parental alleles, the impact of these alleles will be influenced by the effects of molecular level events on their expression. A mutation that occurs within a somatic cell is passed on as part of a clone generated by asexual cell division (→). When and where a somatic mutation occurs will determine what the number and type of cells in the organism that carry the mutation. If the mutation leads to a cell lethal phenotype, cells that carry it will die, so no cells in the organism will carry the mutation. More often, somatic mutations are not lethal and can influence the rate and outcomes of cell divisions and a range of cellular behaviors.



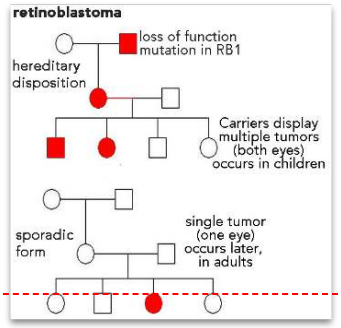
A multicellular organism is a social system. Cells cooperate in defined ways to keep the system functioning smoothly. In particular, when and where a cell divides is under strict regulatory control, involving both internal regulatory networks, as well as signals from other cells. Some somatic mutations disrupt one or more of these regulatory networks, leading to inappropriate cell division, a behavior that underlies the appearance of tumors and metastatic cancers. Carcinogenesis itself is a complex process, involving a number of steps, a number of distinct mutations within a particular clone. While a complete study of cancer is well beyond us here, certain common features are worth considering. In particular, somatic mutations can lead to cells ignoring signals meant to control their growth and behavior. A mutant somatic cell can generate a clone that will compete with wild type clones in various ways.

This page titled 14.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

14.1: Rates and effects of somatic mutation

The rates at which mutations are found within a particular cell type are based on the number of rounds of DNA replication leading to that cell type, the error rate associated with DNA synthesis, the rate of non-replication associated mutations, and the efficiency of DNA error repair. DNA error rates differ between species. In the mouse the error rate is estimated to be one mutation per ~5 x 10^{-9} per base pairs per generation; in humans the value is closer to ~1.2 x 10^{-8}per base pair per generation. Mutation rates in somatic cells appear to be higher than in germ line cells.[476] If we think about cumulative effects, that is from fertilized egg to the production of gametes, there are about 400 replication events in a human male, fewer in a female. It has been estimated that, compared with the chromosomes our parents supplied us, we each have ~100 new mutations in our germ line chromosomes.

Each new somatic mutation interacts with the pre-existing genome to produce a range of effects. These effects depend upon where and when the gene effected is normally expressed, the mutation's effect(s) on the gene's product(s), and its interactions with the rest of the expressed genome. If the mutation occurs early and in many cell types, many tissues may be affected; if late, its effects may be restricted to a specific cell type or region of the organism. As an example in the brain, a new pathologic phenotype can develop if as few as 10% of cells carry the mutation.[477] A somatic mutation can lead to the loss of growth control, and either cell death or over-proliferation - the formation of a tumor, either "benign" or malignant. In the case of cancer, a number of mutations occur that alter cell behaviors. These turn the well behaved somatic cell into a social cheater (chapter 4). Subsequent mutations can accumulate that enable the cancer clone to get better at competing with its normal neighbors and avoiding the host's defensive responses. The evolution of the cancer clone is, however, ultimately futile. The mutant clone will continue to divide and grow, but in the end such growth is incompatible with the survival of the host; both clone and host will die of the disease.[478] There are a number of ways that mutations can lead to cancer.[479]



We consider just one type of "predisposing" genetic interaction that leads to susceptibility to retinoblastoma, a cancer of the retina. Typically retinoblastoma is rare, but there is a form associated with an inherited dominant, loss of function allele (Rb^-) in the *RB1* gene (→).[480] Inheriting a single copy of the Rb^- allele is not, however, sufficient. A second, somatic, mutation is needed; this mutation inactivates the wild type *RB1* allele and leads to a dramatic increase in the probability of cancer. People who do not inherit the Rb^- allele can get retinoblastoma; the difference is that they have to accumulate two separate somatic mutations, a much more unlikely event. When sporadic forms of retinoblastoma appear, they are almost always restricted to one eye, and appear in older individuals. Such somatic mutations are unlikely to affect the germ line, and so will not be inherited. A similar pattern of inheritance is associated with breast cancer susceptibility gene 1 (BRCA1).[481]

References

[476] see Differences between germline and somatic mutation rates in humans and mice and from 2019 here

[477] see Somatic Mutation, Genomic Variation, and Neurological Disease; Discovery of autism/intellectual disability somatic mutations in Alzheimer's brains

[478] The exception is the occurrence of cellularly transmissible cancers, described in Tasmanian devils (*Sarcophilus harisii*) and a small number of other species- see Some Cancers Become Contagious

[479] Neomorphic mutations create therapeutic challenges in cancer

[480] Genetics of Retinoblastoma.

[481] BRCA1 and BRCA2: Cancer Risk and Genetic Testing

This page titled 14.1: Rates and effects of somatic mutation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

14.2: Non-disjunction- aberrant chromosome segregation

There is one more genetic disorder that we will consider briefly, non-disjunction. Non-disjunction refers to the situation where there is a failure of normal chromosome segregation. In the case of somatic (mitotic) cell division, one daughter cell may receive two copies of a chromosome, while the other daughter receives none. This can lead to lethality or differential reproduction (somatic evolution) within the two resulting clones. In the germ line, non-disjunction can lead to a gamete containing extra copies of one or more chromosomes, a situation known as chromosomal aneuploidy. Given that each chromosome, even the smallest ones, contains hundreds or more genes, the absence of the correct number of chromosomes leads to many changes in the patterns of gene expression. Generally, when a chromosomal aneuploidy occurs, the effect is embryonic lethality. Recent studies indicate that chromosomal abnormalities are surprisingly common in early human embryos.[482] For example, when a human embryo carries three copies of one of the smaller human chromosomes, chromosome 21 (the basis for Down Syndrome), ~80% of suchembryos perish *in utero* or in the neonatal period.[483] In cases where the early embryo is mosaic for chromosomal abnormalities, euploid blastomeres (embryonic cells) can replace aneuploid cells and lead to chromosomally normal embryos (and people!)[484]

Questions to answer:

202. A somatic mutation occurs early in development, what factors will influence the % of cells in the organism that carry the mutation over time?
203. How would a mutation in a cell cycle (mitosis or meiosis) checkpoint gene influence a somatic cell's clonal evolution?
204. What types of molecular defects would lead to chromosomal aneuploidy?
205. How might having three (or one) copy of a chromosome influence normal cell behavior (and gene expression)?
206. Propose a model that explains why inheriting a cancer Rb^- or BRCA1 allele leads to increase risk of cancer in some but not all tissues?

Question to ponder:

- Can you imagine a situation in which a somatic mutation became an inheritable allele in the next generation?

References

[482] Chaos in the embryo

[483] Morris et al. 1999.: Fetal loss in Down syndrome pregnancies. Prenat Diagn. **19**: 142-145.

[484] Mosaicism in preimplantation human embryos: when chromosomal abnormalities are the norm

This page titled 14.2: Non-disjunction- aberrant chromosome segregation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

14.3: Genome dynamics

Aside from the insertion of “external” DNA through horizontal gene transfer, something rare in eukaryotes or abnormal meiotic recombination events (see below), we might assume that the genome itself, is largely static. It is, however, clear that genomes are more dynamic than previously thought. In addition to the point mutations that arise from mistakes in DNA replication, different type of genomic variation were uncovered in the course of genome sequencing studies, these include the movements of transposable elements, known as “structural variants.” They include flipping of the orientation of a DNA region (an inversion) and sequence insertions or deletions, known as copy number variations.[485] It has been estimated that each person contains about 2000 “structural variants”.[486] Large chromosomal inversions or the movements of regions of DNA molecules between chromosomes can have effects on chromosome pairing during meiosis (described above), and can lead to hybrid sterility and inviability. The mechanisms that lead to these genomic changes are largely beyond our scope here.[487]

As before, if new genetic variants occur in the soma, rather than the germ line, they will be lost when the host organism dies. If a mutation disrupts an essential function, the affected cell will die and likely be replaced by surrounding cells. Multicellular organisms are social systems. It is often the case that organisms have both internal (cellular) and social (organismic) systems to guard against social cheaters. Mutant, "eccentric" or "misbehaving", cells can actively kill themselves through the process of apoptosis or they can be induced to die through interactions with their normal neighbors or, in organisms with an immune system, by cells that can identify them as abnormal and induce their death.[488]

References

[485] Copy number variation in humans:

[486] Child Development and Structural Variation in the Human Genome

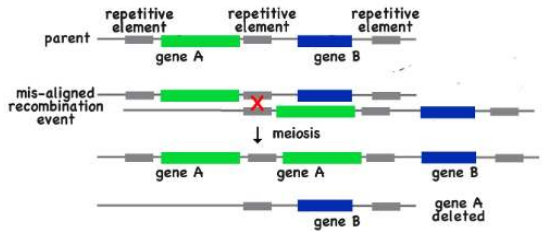
[487] Mechanisms of Gene Duplication and Amplification

[488] Conceptual simplicity and mechanistic complexity: the implications of un-intelligent design

This page titled 14.3: Genome dynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

14.4: Gene duplications and deletions

While meiotic alignment generally occurs accurately, there are times were mis-alignment happens. Consider the effects of repeated sequences within a chromosome. If the homologous chromosomes misalign, crossing over can lead to haploid cells that emerge from meiosis with either a gene duplication or a deletion (→). Such duplication events can have a liberating effect on subsequent evolutionary pathways.[489] Most obviously, having two copies of a previously single copy gene means that it is possible to make twice as many transcripts per unit time per cell. This extra activity can be useful. For example, imagine that the original gene product was involved in inactivating an environmental toxin; one copy of the gene might not make enough protein to allow the cell/organism to survive and grow, whereas two copies might. When one analyzes bacterial (or cancer) cells that can grow in the presence of a toxic compound, it is common to find that a gene(s) that encodes a protein(s) involved in degrading or exporting the toxin from the cell has been duplicated one or more times.[490]



Another adaptive mechanism depends upon the fact (noted above) that while a particular gene product typically has a clear "primary" activity, it can also have one or more weaker secondary activities.[491] Assuming that a gene product's primary function is essential for survival or reproductive success, changes that negatively influence survival or reproductive success will be selected against, even if they improve valuable secondary activities. The duplication of the gene allows the original activity to be preserved, while the duplicated gene can evolve freely, and may provide useful, off-target activities or alter when and where the gene is expressed.

References

[489] Ohno's dilemma: evolution of new genes under continuous selection: and Copy-number changes in evolution: rates, fitness effects and adaptive significance

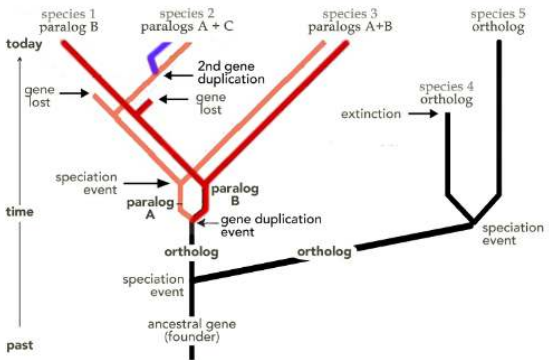
[490] Dihydrofolate reductase amplification and sensitization to methotrexate of methotrexate-resistant colon cancer cells:

[491] Enzyme promiscuity: a mechanistic and evolutionary perspective & Network Context and Selection in the Evolution to Enzyme Specificity

This page titled 14.4: Gene duplications and deletions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

14.5: Orthologs and paralogs

When a gene with similar sequences is found in distinct species of organisms, the general assumption is that the gene was present in the species' common ancestor. In this case, the genes are consider homologous and referred to as orthologs of one another. When a gene is duplicated, in an organism (and in a population) the two versions are termed paralogs of one another and can evolve independently (→). Even more dramatically, entire genomes appear to have been duplicated multiple times during the course of evolution.[492] In any gene duplication event, the duplicated genes can have a number of fates, they can act as a "back-up" for one another, they can be re-purposed, or one can be lost. Repeated gene duplication events can generate families of evolutionarily-related genes that are recognized by the presence of similar nucleotide and amino acid sequences and structural motifs in the encoded polypeptides. Orthologs are presumed to be derived from a single gene present in the last common ancestor of those species. Paralogous genes are present together in the ancestral organism. If one paralog of a pair is subsequently lost, it can be difficult to distinguish the remaining gene from the original ortholog (NIH LINK).



When paralogs are present in a species, detailed gene/polypeptide sequence comparisons are used to characterize the gene's family tree. The further in the past a gene duplication event occurred, the more mutational noise can obscure the relationship between the duplicated genes. For example, when looking at a DNA sequence there are only four possible bases at each position. A mutation can change a base from an A to a G; a subsequent mutation can change the G back to A. With time, this becomes more and more likely, making it impossible to accurately calculate the number of mutational events that separate the two genes. Many multigene families appear to have originated hundreds of millions or billions of years ago, the older the common ancestor, the more obscure the exact relationship. The exceptions involve genes encoding polypeptides/proteins that are very highly conserved because they are essential and do not tolerate changes. These gene/polypeptide/protein sequences evolve very slowly. In contrast, gene/gene products that are subject to less rigid constraints evolve more rapidly, since more different variants are compatible with life. Speedy evolutionary change complicates using sequence information for determining the relationships between genes found in distantly related organisms. While functional similarities are evidence for evolutionary homology, it is possible, particularly with highly divergent genes and gene products, that they are the result of convergent evolution. As with wings, there may be a different number of ways to carry out a particular molecular/cellular level function.

In an interesting recent set of observations, it appears that genomes contain multiple, previously uncharacterized genes that encode microproteins. Many of these sequences are found in DNA/RNA regions previously thought to be "non-coding". A surprising observation is that many of these microproteins arise rapidly. For example, there appear to be large numbers of microproteins encoding sequences specific to humans, and different from those of other organisms.[493] Their functional roles can complicate comparisons between model organisms and humans.

References

[492] Genome and gene duplications and gene expression divergence: a view from plants

[493] Sandmann et al.,(2023). Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. *Molecular cell*, *83*, 994-1011.

This page titled 14.5: Orthologs and paralogs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

14.6: Transposons- moving DNA within a genome (and weird genetics)

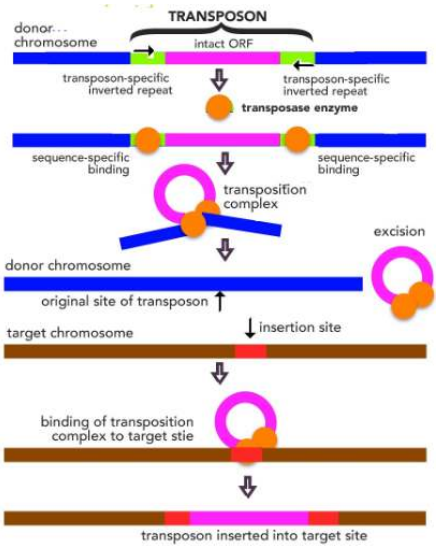
As we are thinking about DNA molecules moving into the genome through horizontal (lateral) gene transfer, and between genomes through conjugation, we can consider another widely important molecular system known as transposable elements or transposons. A transposon is a piece of DNA that can move (jump) from place to place in the genome.[494] The Nobel prize winner Barbara McClintock (1902–1992)(→) first identified transposons while studying maize (*Zea mays*).[495] In particular, she studied the phenomena known as variegation in the pigmentation of kernels (→). The variegation phenotype is due to what are known as unstable alleles; these are pairs of alleles in which one allele is associated with one phenotype (e.g. dark pigment) and the other allele is associated with a second phenotype (e.g. different or lighter color). During development of the kernel an allele can change from one state to another. Since tissues are built from (asexual) clones of somatic cells, the earlier in development an allele change occurs, the larger the region associated with its phenotype in the adult organism.[496]



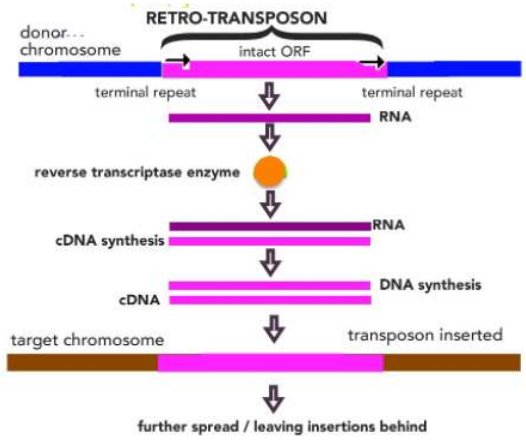


Transposons can have a number of different effects on the expression of the genes in which they are found.[497]For example, some transposons can disrupt a gene's coding region. When spliced out of the mRNA the gene can produce a normally functioning gene product.[498] Alternatively, the insertion of a transposon can inactivate the gene. Transposons come in two general types - those that move a DNA sequence from one place in the genome to another with no increase in total transposon copy number – these are known, for historical reasons, as type II transposons (next ↓ page). Type II transposons come in two forms, known as autonomous and non-autonomous. Autonomous transposons encode a protein known as transposase. The transposon contains repeated nucleotide sequences at each end. Transposase recognizes these repeat sequences and catalyzes the removal of the DNA region between the two repeat sequences and its subsequent insertion into another genomic site. The new site can be located anywhere in the genome where the chromatin is in an "open" (accessible) state. This property has been used to map the regions of the genome that are open, a method known as "Assay for Transposase Accessible Chromatin with high-throughput sequencing" (ATAC-seq).[499] In non-autonomous type II transposons, mutations have led to the loss of a functional transposase gene within the transposon. By itself, such a transposon cannot move, but if a autonomous transposon becomes active within the cell then the transposase it encodes can catalyze the excision and insertion of a dependent transposons. Why? because the transposase protein is synthesized in the cytoplasm and after it enters the nucleus, it can interact with multiple transposons.

The second type of transposon, known as a type I transposon, is also a DNA sequence but it uses a different mechanism to move. Type I transposons also come in autonomous and non-autonomous forms (↓). The autonomous form encodes a protein known as reverse transcriptase, a RNA-directed, DNA polymerase. When expressed, the type I transposon leads to the generation of an mRNA that encodes reverse transcriptase. Reverse transcriptase can recognize and make a complementary DNA (cDNA) copy of the transposon-encoded RNA. The cDNA can, in turn, be used as the template to generate a double-stranded DNA molecule that can be inserted, more or less randomly, into the genome. In contrast to a type II transposon, the original transposon's DNA sequence remains in place, and a new transposable element is created and inserted into the genome. If the transposon sequence is inserted into a gene, it may generate a mutation by disrupting the gene's regulatory or coding sequences. It can also act as a regulatory element, leading to changes in when and where the gene is expressed. In dependent (non-autonomous) type I transposons, mutations have render the reverse transcriptase non-functional; it can only make copies of itself if another, separate autonomous type I transposon is present and actively expressed in the cell.



Transposons do not normally encode essential functions. Random mutations can "kill" a transposon by modifying the molecular features involved in its recognition, excision, replication, and insertion within a genome. If you remember back to our initial discussion of DNA, human and many other types of genomes contain multiple copies of specific sequences - many of these are derived from once active transposons;



most are now "dead" – they are the remains of molecular parasites. About ~50% or more of the human genome consists of dead transposons. In particular the human genome contains ~1,000,000 copies of the Alu type transposon (~11% of the total genome); these are dependent, type I transposons that rely on the presence of autonomous transposons to move.[500] It is probably not surprising then that there can be movement within the genome of specific cells during an organism's life time, since some transposons are still active.[501] Moreover, since transposon movement is generally stochastic, as populations separate from one another, the patterns of transposons within the genome diverge from that of the ancestral population.[502] In addition, various stresses within an organism can enhance transposon movement, which may play a role in the generation of genetic variation - a primary driver of evolutionary diversity and adaptation.[503]

Questions to answer:

207. How many ways can you imagine that the movement of a transposon could influence gene expression?
208. What are the selective pressures on the maintenance or destruction of active transposons?
209. How could the movement of a transposable element NOT produce a mutation?

Question to ponder:

- Does the presence of molecular parasites represent an evolutionary design feature or an unintended consequence of molecular machines involved in"normal" DNA dynamics and mutational repair?

Short Chapter Summary

- Somatic change accumulates with age and environment; mosaicism is common.
- Copy-number change, transposons, and mis-segregation reshape genomes and disease risk.
- Gene duplication fuels novelty; paralogs specialize

References

[494] Transposons: The Jumping Genes: http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518

[495] Barbara McClintock: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1983/mcclintock-bio.html

[496] In you can't stop yourself, check out: Controlling elements in maize.

[497] Transposable Elements, Epigenetics, and Genome Evolution: http://science.sciencemag.org/content/338/6108/758

[498] The Maize Transposable Element Ds Is Spliced from RNA: https://www.ncbi.nlm.nih.gov/pubmed/3039661

[499] ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide

[500] Wikipedia: Alu element

[501] Active transposition in genomes

[502] The impact of retrotransposons on human genome evolution: https://www.ncbi.nlm.nih.gov/pubmed/19763152

[503] Stress and transposable elements: co-evolution or useful parasites? https://www.ncbi.nlm.nih.gov/pubmed/11012710

This page titled 14.6: Transposons- moving DNA within a genome (and weird genetics) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

15: Mendel and Weldon - Contexts and their Effects on Phenotypes

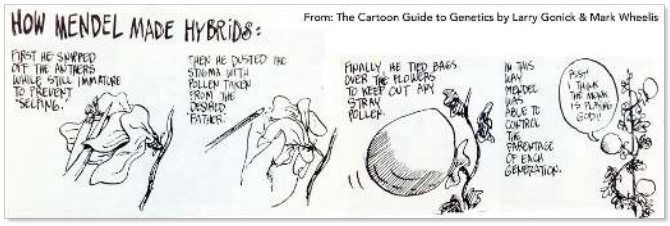
15.0: Introduction
15.1: How Mendel did what he did
15.2: Weldon's critique
15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
15.4: Dihybrid crosses- linkage and recombination
15.5: Genetic complementation
15.6: Interacting traits- synthetic lethality and co-dominance
15.7: Interacting traits- epistasis
15.8: Maternal and paternal effects
15.9: Mitochondrial inheritance
15.10: Imprinting- conflicts between mother, father, and fetus
15.11: Estimating the number of genes involved in a particular traits
15.12: On the nature of mutations (again)
15.13: Alleles, traits, and genetic diseases in humans.
15.14: Concordance between monozygotic twins and genetic influence on a trait
15.15: Measuring evolution’s impact on allele frequencies- Hardy-Weinberg
15.16: Genetic anticipation
15.17: The persistence of deleterious alleles

This page titled 15: Mendel and Weldon - Contexts and their Effects on Phenotypes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.0: Introduction



In which we consider Mendel's contributions together with their critique by William R.F. Weldon, including the realization that genetic elements behave in predictable but often complex ways and consider simplifications that can lead to inaccurate, genetically deterministic conclusions. We consider examples of how genes function within biological systems, including how molecular, developmental, and environmental variation combine to influence the uniqueness of individuals. [504]



As we think about the historical origins of genetics, it is worth considering the biases imposed by the way that Gregor Mendel (1822-1884) did his work. These reflect the realities of science – understanding does not appear fully formed, like religious revelation, it is built up through experimentally constrained insights, some of which are productive and others that turn out to be distractions. Subsequent observations and experiments lead us to recognize the implications and limitations of original ideas, better termed tentative hypotheses and working models rather than theories and drive their refinement or abandonment. The path from an idea to new discoveries and concrete conclusions is rarely as linear as they are often made to appear in a scientific paper. In Mendel's case, he began his work around 1854 and published it 11 years later in 1865; it took 35 years from the time he published his work until it was recognized (1900) as establishing something fundamental about genetic mechanisms–its significance was not immediately obvious. [505]

References

[504] Radick, G. (2023). Disputed inheritance: the battle over Mendel and the future of biology. University of Chicago Press.

[505] It is not as though people did not know of his work, "The methodical monk sent reprints of the article to 40 leading biologists around Europe, including Charles Darwin. Darwin's copy was found later, with its double pages still uncut: It had not been read." and "Mendel's work received little notice elsewhere and was cited a mere three times over the next 35 years."

This page titled 15.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.1: How Mendel did what he did

To make genetic behaviors intelligible, Mendel purposefully selected (and bred) plants whose mating partners he controlled (↑); plants that produced high numbers of offspring, and that displayed easily characterized and uniform (from plant to plant) traits. The traits he chose were independent of one another and were not dramatically influenced by environmental effects (growth conditions). His most famous work involved the garden pea *Pisum sativum*, which displays all of these features. Mating in peas involves male pollen (the plant equivalent of animal sperm). During fertilization a pollen cell fuses with an ovule cell, the plant equivalent of an animal egg. Pea plants can self fertilize, but this can be prevented and the experimenter can control the source of the pollen.[506]

Mendel identified or developed lines of peas that displayed one or the other of various pairs of traits (↓). In many cases, this involved "breeding out" natural variation. A case in point is pea color. The type of pea plants that Mendel worked with normally display a continuous range of seed colors, from green to yellow. Over a number of generations, Mendel selected the greenest and yellowest plants for in-breeding, leading to strains that produced seeds of uniform green or yellow color, with no intermediates. The plants "bred true" for seed color.

Results of all of Mendel's monohybrid crosses

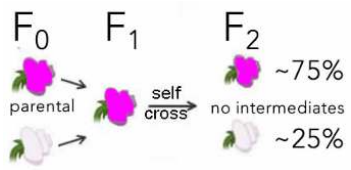
	Parental phenotype	F_1	F_2
1.	Round×wrinkled seeds	All round	5474 round; 1850 wrinkled
2.	Yellow×green seeds	All yellow	6022 yellow; 2001 green
3.	Purple×white petals	All purple	705 purple; 224 white
4.	Inflated×pinched pods	All inflated	882 inflated; 299 pinched
5.	Green×yellow pods	All green	428 green; 152 yellow
6.	Axial×terminal flowers	All axial	651 axial; 207 terminal
7.	Long×short stems	All long	787 long; 277 short

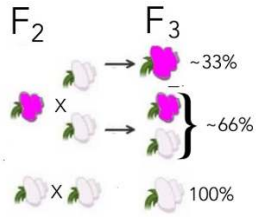
Griffiths et al., 2000

An individual plant is derived from a single pollen grain fertilizing an ovule. To say that a plant line breeds true means that when a plant is allowed to self-fertilize all of the offspring produced display the same form of the trait. These offspring will, if allowed to self-fertilize or to fertilize each other, again produce offspring that display the same form of the trait as the parent(s). The are in-bred, that is genetically homogeneous (similar), with little plant to plant variation - a situation quite different from most natural populations.

Next he crossed (fertilized) one plant with gametes from another. For example he fertilized a plant with white flowers with pollen from a plant with purple flowers, and examined the traits expressed in the offspring, known as the F_1 generation. On analyzing the traits of a large number of F_1 offspring he found that among this set of traits, only one of the pair of traits was displayed or expressed. When the parents (the F_0 generation) had purple or white petals, all of the offspring (F_1) individuals had purple flowers. It did not matter if the purple plant was the maternal or the paternal parent. In such a cross the parental trait displayed in the F_1 generation was said to be "dominant" to the "recessive" parental trait, that is the trait that was not displayed. The traits Mendel worked with all behaved in this way (that is one reason he studied them). Moreover, when two or three of these traits were displayed in the same individual, they did not influence each other - they behaved independently. The result was not surprising in that Mendel did not start with random traits, he selected traits that followed these rules, they were "well behaved", a fact we will consider further.

Mendel continued his experiments; he crossed true breeding F_0 individuals expressing one or the other trait, to produce F_1 individuals. He then crossed F_1 individuals to themselves or to F_1 individuals from other crosses (↓). Here came the surprise, from F_1 x F_1 crosses there emerged F_2 individuals that displayed the recessive form of the trait. As he collected large enough numbers of F_2 individuals, he discerned a pattern - approximately 25% of the F_2 individuals displayed the recessive form of the trait: a clear 1 to 3 ratio of individuals expressing the recessive traits compared to those that expressed the dominant trait. When the F_2 individuals that display the recessive trait were crossed to one another (or themselves), the resulting F_3 individuals all (100%) expressed the recessive trait (→). The F_2's that expressed the recessive trait were like the original in-bred, recessive F_0 parent. However, the F_2 individuals that displayed the dominant trait were not all the same. When Mendel crossed the F_2 individuals that expressed the dominant trait to recessive F_0 individuals and then examined large numbers of offspring, the results fell into two classes. In one third (~33%) of cases, all the offspring displayed the dominant trait, while in two thirds (~66%) of the cases approximately half of the offspring expressed the dominant trait and half expressed the recessive trait.





Mendel used his data to come up with a model for trait behavior. He assumed that each trait was controlled by two factors (alleles) of what we now refer to as a gene. In each of the parental lines these two factors were the same, they were "homozygous" for either

a dominant or a recessive allele. All of the gametes produced by an F_0 individual therefore carry the same allele of the gene associated with the trait. In a cross between parents homozygous for different alleles of a particular locus, the model predicts that all F_1 individuals are "heterozygous" and display the same "dominant" phenotype.

An F_1 individual heterozygous (for a particular gene) will (normally) produce equal numbers of two different types of gametes; 50% of the gametes will carry the dominant allele while 50% of the gametes will carry the recessive allele. When two F_1 individuals mate, there are four possibilities for the F_2 offspring. An F_1 gamete carrying the dominant allele can fuse with an F_1 gamete carrying either the dominant or the recessive allele. Similarly, an F_1 gamete carrying the recessive allele can fuse with a F_1 gamete carrying either the dominant or the recessive allele. If we assume that these events are equally probable. If the number of offspring is large enough we expect to find dominant to recessive phenotypes in the F_2 generation in a 3:1 ratio, All of the recessive phenotype individuals should be homozygous for the recessive allele. The individuals displaying the dominant phenotype can be either heterozygous or homozygous for the dominant allele. We can distinguish these two classes of individuals using what is known as a backcross; this involves crossing them to a homozygous recessive individual. If the individual to be tested is homozygous for the dominant allele, all of the offspring from the backcross will display a dominant phenotype. If the individual tested is heterozygous, if we collect a large enough number of offspring 50% should display the dominant phenotype and 50% should display the recessive phenotype. Given large enough numbers we will find that the F_2 generation consists of 25% homozygous dominant, 50% heterozygous, and 25% homozygous recessive, a 1 to 2 to 1 ratio. Mendel's observations were consistent with these ratios.

Key to Mendel's model was the assumption that the variants of a specific trait he examined were unambiguously distinguishable and determined by the alleles of a single gene. These are what are known as monoallelic traits. In addition, the alleles, and the traits they influenced displayed clear dominant-recessive behaviors with respect to one another; few alleles/traits behave this way. In many cases individuals heterozygous for a particular gene display a phenotype distinct from either homozygous form of the trait. Finally, while Mendel examined a number of traits, they all behaved independently of one another. That is because he selected them for their independence. As we know now, that means that these behaved (segregated) independently during meiosis. But remember, Mendel knew nothing about chromosomes and the molecular mechanisms involved, it was just that his choices of traits made his data intelligible and enabled him to build a relatively simple predictive model.

References

[506] A typical experiment begins with a hypothesis, a guess on how a particular perturbation influences the system. A study is more about observing and collecting data about a system. From observations, we make hypotheses about how the system will act under various conditions (an observational study) or how a perturbation (an experimental study) will alter the system’s behavior. Our prediction of the outcome is known as the null hypothesis - we examine the data collected to determine whether the prediction's null hypothesis is supported or not, or whether the data produced could have arisen by chance.

This page titled 15.1: How Mendel did what he did is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.2: Weldon's critique

Weldon argued that Mendel's focus on traits that displayed dichotomous dominant-recessive behavior in-bred pea strains differed significantly from the behavior of similar plants "in the wild", where traits often vary continuously with a population and unexpected outcomes were observed. He contended that Mendel's "rules" did not accurately reflect how inheritance works for most traits.[507]

Most traits are controlled by multiple genes and their alleles often do not act in a simple dominant or recessive manner. Many modern laboratory studies (including Mendel's) are carried out in in-bred genetic backgrounds, and different phenotypes can emerge in different backgrounds. In in-bred strains the organisms share a common combination of alleles at other genetic loci. Such genotypic homogeneity is an artifact of the way such strains were generated and experiments were conducted. Wild type populations display much more genotypic variation, generally many different alleles of many genes are present. Consider a dominant allele. In the wild, the phenotypic traits associated with that allele often vary - the extent of such variation is characterized through the terms expressivity and penetrance. Variable expressivity refers to the observation that even in the presence of the associated (dominant or homozygous recessive) allele, the phenotypes of individuals vary. As an example, consider a hypothetical pea; the exact degree to which each pea is wrinkled varies – some a little more, others a little less wrinkly – an example of variable expressivity. Similarly, it is possible that out of 100 individuals that carry a particular dominant or homozygous recessive allele, some will not display the trait associated with the allele. The percentage of individuals that display the trait is known as penetrance. Genetic background, together with molecular and cellular stochastic effects influence both the expressivity and penetrance of an allele. Various combinations of alleles of other genes can act to "suppress" and "enhance" the phenotype associated with an allele.[508] By restricting his work to fully expressive and penetrant dominant and recessive alleles, together with the availability of sufficient numbers of offspring of each class, Mendel was able to make sense of his observations. As noted previously, however, Weldon made explicit the impact of the complexity of genetic interactions, together with environmental and developmental processes that collaborate to shape an organisms' phenotype.

Questions to answer:

210. Why was it critical for Mendel's studies to be able to control crosses between individual plants?
211. What led Mendel to be able to discover recessive alleles?
212. Describe, in terms of meiotic behaviors, how the results of a monohybrid cross are produced.
213. Explain why, when small numbers of offspring are generated, the ratio of phenotypes in a F2 cross can differ from the expected 3:1 ratio.

Questions to ponder:

- Why are backcrosses to homozygous recessive individuals informative? Are backcrosses to homozygous dominant individuals useful?
- How does one determine, in practice, that a homozygous recessive individual is homozygous recessive?

References

[507] Wnat to know (much) more about Weldon check out this [link]

[508] here is a particularly relevant recent study: Genetic background limits generalizability of genotype-phenotype relationships

This page titled 15.2: Weldon's critique is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.3: Chi square (?2) analysis, hypothesis testing, and small numbers

Mendel's conclusions depended on the number of plants he examined. The various ratios he predicted are expected to be true and reproducibly observed only when the number of individuals examined is large. With smaller numbers, there can be serious divergences between what is observed and what the model predicts, a situation characteristic of stochastic processes. Which gametes contain which alleles and who they fused with are both stochastic events.[509] Consider the general question, how many rolls of a die would you need to convince yourself, with high confidence, that a particular die is fair? or perhaps better put, not unfair. While the stochastic nature of meiosis and fertilization does not effect the F_1 offspring of a cross between homozygous dominant and recessive plants, in which all offspring are expected to have the same (heterozygous) phenotype, it will influence the 3:1 ratio (in the F_2generation) of phenotypically dominant to recessive plants predicted to occur when F_1 individuals are crossed. How do we decide whether what we observe is consistent with our model or contradicts it? A model that does not produce the observed results will need to be abandoned or revised.

$$X^2 = \sum^{n} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

The answer is a statistical test known as a χ^2 (chi square) analysis.[510] It uses the equation (↓) together with two other concepts: degrees of freedom and the null hypothesis.[511] If we are testing a model that makes a mathematically precise prediction, such as the frequency of the phenotypic classes observed, our null hypothesis is that the data are unlikely to be generated by chance. Remember, we are not trying to prove that our specific hypothesis is correct; we are trying to estimate the probability that the values observed could have occurred by chance rather than by the specific mechanisms proposed.

To define the degrees of freedom of a study, we need to know how many independent variables there are. In our two phenotype system (wrinkled or round, purple or white, etc.), we assumed that all individuals have one or the other unambiguously distinguishable phenotype. If we know the number of individuals involved and the number of either phenotype, we automatically know the number of the other. In the case of two phenotypic classes, the degree of freedom is 1 (if there are four classes, the degree of freedom is 3, and so on). What is the degree of freedom for a six-sided die? By convention, currently under some discussion[512], we take an observation to be consistent with the null hypothesis if it can be expected to occur by chance at less that 1 time out of 20 (0.05) or one time out of one hundred (0.01); otherwise we have a good case to reject our hypothesis - that is, the data observed could well be due to chance alone.

For any particular experiment, we make observations to test our null hypothesis, are our predictions supported or rejected? Just for fun, let us consider Mendel's monohybrid crosses. The prediction of his model is that the ratio of round to wrinkled seeds in the F_2 will be 3:1. Mendel reported that he examined 7324 plants. Given his model, he would have predicted that 5492 of these plants would have round seeds, while 1849 plants would have wrinkled seeds. We can now do our χ^2 calculation. We have (5474 (observed) – 5492 (expected))^2 = (-18)^2 = 324/5492 (expected) equals 0.059 and (1850 (observed) – 1849 (expected))^2 = 1^2 = 1/1849 (expected) = 0.00054. The sum (∑) of these two numbers is 0.0595. To determine whether these observations are consistent with our null hypothesis, we consult a χ^2 probability table (↓). The higher the χ^2 value the more likely the difference between observed and expected data is due to chance, rather than because our assumption, our null hypothesis, is correct. Our value of 0.059 lies well below the 0.05 probability value of 3.841, suggesting that the observed numbers are consistent with our model and unlikely to be generated by chance. But keep in mind, consistency does not imply "truth." In fact, there have been suggestions that Mendel's observed numbers are too good, too close to what would be predicted from his model.[513] Be that as it may, Mendel's conclusions for the behavior of the types of traits he chose to study have been repeatedly verified - we can trust his general conclusions given his assumptions.

Selected percentile values of the χ^2 distribution

df*	.99	.95	.50	.10	.05	.01
1	.000157	.00393	.455	2.706	3.841	6.635

References

[509] It is similar to the question of which unstable isotope atom will decay next.

[510] Here is an alternative presentation from GENETICS AND GENE PROBLEMS

[511] chi square tutorial: http://www.radford.edu/rsheehy/Gen_flash/Tutorials/Chi-Square_tutorial/x2-tut.htm

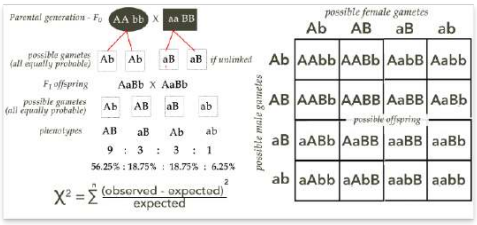
[512] Statistical errors and Colquhoun. 2014. An investigation of the false discovery rate and the misinterpretation of p-values

[513] see On Fisher's Criticism of Mendel's Results With the Garden Pea

This page titled 15.3: Chi square (?2) analysis, hypothesis testing, and small numbers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.4: Dihybrid crosses- linkage and recombination

Now we can move to more complex questions. As an example, consider two distinct traits (smooth/wrinkled and yellow/green seeds). We ask, do the alleles involved behave independently of one another or do they physically interact in some way? We begin, based on a monohybrid analysis, to identify alleles as recessive or dominant. We can assume a null hypothesis, that the two traits behave independently; that is they do not interact physically with one another. That means that they are either on different chromosomes or far away from one another on the same chromosome - they are "unlinked". Assum that we begin with two lines that breed true for these traits. As before, each parental F_0 organism can produce only one type of gamete, and all F_1 organisms will have the same AaBb genotype, independently of which parent was AA and which was bb. We can predict the outcome of crosses between F_1individuals. Assuming that the two genetic loci are independent, each F_1 individual will produce four different types of gametes in equal numbers and maternal and paternal gametes fuse randomly. We can visualize this behavior, and the outcomes of the cross, using what is known as a



Punnett square (→), which enables us to determine the possible phenotypically distinct outcomes and their relative frequencies given our assumptions.[514] There are 16 possible combinations of these alleles in the F_2 generation. Nine display a dominant:dominant phenotype: AABB (1), AABb (2), AaBb (4), AaBB (2); three display a dominant:recessive phenotype: AAbb (1), Aabb (2) or a recessive:dominant phenotype: aaBB (1), aaBb (2). One (aabb) displays a recessive:recessive phenotype. If we examine enough F_2 progeny we expect to find these phenotypic classes in a ratio of 9:3:3:1. Test crosses to recessive:recessive organisms can identify the genotypes of the various organisms. A χ^2analysis enable us to determine whether the outcome of a particular dihybrid (two trait) cross is consistent with our hypotheses that the alleles involved do not interact with one another and that they are unlinked.

What happens if we find that the cross produces the same phenotypic combinations but in numbers that do not match our predicted (expected) values? The simplest conclusion, and one not made or considered by Mendel because he excluded such traits, was that i) the genetic loci involved are go through meiosis together - they are "linked" and that ii) on occasion meiotic recombination can, separate these linked genes.[515] Consider one such example, we generate a dihybrid F_2 generation from AB phenotype F_1 offspring (→). We carry out a χ^2 analysis and obtain a value of 3492. A look at the probability table (↓) confirms our suspicion, namely that our null hypothesis, that the genes are unlinked, is rejected. An alternative hypothesis is that the genes are linked to one another and separated by a certain distance. We can generate an estimate of how close on the chromosome they are to one another.

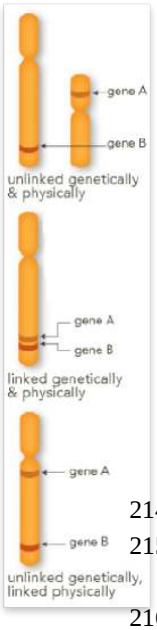
expected
AB : aB : Ab : aa
981 : 72 : 86 : 964
552 : 394 : 394 : 131
observed

Chi Square Values and Probability

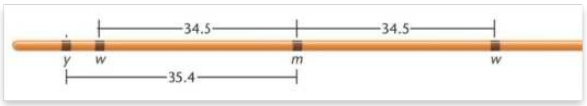
Degrees of Freedom	P = 0.99	0.95	0.80	0.50	0.20	0.05	0.01
3	0.115	0.352	1.005	2.366	4.642	7.815	11.345

We know from our cross that the parents (F_0) were AB and ab, and that the chromosomes were AB and ab respectively. If the A and B genes are located on the same chromosome, we can assume that, in the absence of recombination, only [AB] and [ab] gametes will be generated and that all F_1 organisms were [AB][ab], with the brackets indicating that the alleles are linked on the same chromosome. Again, in the absence of meiotic recombination, we can assume that F_1 organisms produce only [AB] and [ab] gametes. To produce aB or Ab gametes, there must have been a recombination event between the A and B loci. To calculate the frequency at which such recombinationevents occurred, we add the number of aB and Ab organisms and divide by the total number of organisms, in our case the result is 72 + 86 / 2103 = 0.0751. This indicates a recombination frequency of ~7.5%, significantly less than the 50% recombination frequency we would predict if the genes were unlinked. A 7.5% recombination frequency equals 7.5.

When the linkage distance exceeds 50 centimorgans (cM), the two genetic loci behave as if they are unlinked, that is, located on different chromosomes, even if they are actually located on the same chromosome. It is, of course, possible to walk along a chromosome using pairs of loci located near one another (←). In this way, we find that a typical chromosome is more than 50 cM in length. Because recombination (crossing-over) can be influenced by the physical state of the chromosome, for example crossing-over is inhibited within a chromosome's centromeric region, centimorgans do not directly or consistently convert into DNA lengths. That said, on average (in humans) a 1 centimorgan recombination distance corresponds to a physical distance of ~1 million base pairs (a megabase abbreviated Mb) of DNA. From an evolutionary standpoint gene linkage can influence the inheritance of alleles;



the closer two genetic loci are to one another the more generations, on average, it will take for recombination to separate them, so that they are inherited independently.



Using conventional genetic methods, we can extend our analysis of linkage from two to three or more genes, in order to identify the order of genes along a chromosome. If two different genes are linked to the same gene, for example, the *m* gene is linked to the *w* and the *y* genes (→), they can be in various orientations with respect to one another. Genetic crosses using organisms that are originally homozygous for all three alleles, assuming that there are multiple alleles at each locus and that homozygous organisms are viable, can be used to map genes with respect to one another. This enables us to determine if the *w* gene is located upstream or downstream, along the length of the chromosome, of the *m* gene. In an era (like today) of full genomic sequence data, it is easier to use web-based tools such as Genomicus (Chapter 16) [link].

Questions to answer:

214. What does it mean if the null hypothesis is not supported?
215. A dihybrid cross produces offspring that do not fall into the expected 9:3:3:1 distribution, what kinds of conclusions can we make?
216. In a dihybrid cross, the individuals that are homozygous for both recessive alleles are absent, what might you conclude and why?
217. Alleles in two different genes appear linked to an allele in a third gene, but they do not appear to be linked to each other. What can you conclude and why?

Question to ponder:

- Assuming that recombination is suppressed in the region of the centromere, do genes on opposite sides of the centromeric region of a chromosome appear closer or further away (genetically) than they are molecularly?

References

[514] Who was this Punnett fellow? see Reginald Punnett

[515] Why did he missing this type of genetic behavior, because i) he did not have linked traits in his analysis or ii) because he excluded traits that behaved in this way from his analysis - I have not checked with was the actual situation.

This page titled 15.4: Dihybrid crosses- linkage and recombination is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.5: Genetic complementation

When we make mutations in traditional ways, such as using X-rays or mutagenic chemicals, the organisms carrying these mutations are identified for further study based on their phenotypes. But conventional mutagenesis generates lots of mutations. We need to remove these unwanted mutation through “back-crosses”. Sometimes the phenotype of interest is lost when secondary mutations are removed. The strategies involved in “cleaning up” a mutation vary between different genetic systems, and we will not considered them in detail here.[516] In the modern world it is easier to generate mutations using CRISPR-Cas9 and related systems. The footnote describes a scheme for isolating recessive mutations in the fruit fly *Drosophila melanogaster*.[517]

A priori we do not know whether mutations that produce similar or related phenotypes are in the same or different genes. One way to answer this question is through genetic complementation tests. Let us assume that two (newly defined) mutant alleles influence molecular processes leading to clearly discernible traits. We can use dihybrid crosses to carry out a preliminary examination of the various types of interactions between these alleles. As an example, consider two independently derived alleles that produce the same apparent phenotype. Assuming that we can generate viable organisms homozygous for these alleles. If we cross these, let us call them a1/a1 and b1/b1, organisms, we expect that all of the F_1 generation will be genetically the same, at least at these loci. The F_1 organisms are expected to have a a1/+ b1/+ genotype. If they exhibit a wild type phenotype, we can tentatively conclude that these alleles are located in different genetic loci (genes) and are recessive. If they display a mutant phenotype, we can tentatively conclude that these are alleles of the same gene, with a a1/b1 genotype. We might seek to confirm these assumptions by asking whether the alleles are linked, although this can be difficult if a1/a1 and b1/b1 have similar phenotypes. We could avoid this problem if we have enough phenotypically distinct genetic markers; that enables us to determine whether the two genes are linked to the same or different genes. If they were found to be linked to the same markers (allelic versions of other genes), we might conclude that they are alleles of the same gene. If they are linked to different genetic markers, then it is likely that these are alleles of different genes.

Allelic interactions

Independent	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism displays both phenotypes.
synthetic	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism displays a new phenotype (such as lethality)
complementary	allele a in a gene is associated with a particular phenotype allele b in the same or a different gene is associated with the same or a different phenotype	a/b organism displays wild type phenotype
enhancement	allele a in a gene is associated with a particular phenotype allele b is in a different gene	phenotype of a/b organism is more severe than a/+
suppression	allele a in a gene is associated with a particular phenotype allele b is in a different gene	phenotype of a/b organism is less severe than a/+
epistasis	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism expresses only one of the two phenotypes.

Another formal possibility is that these two alleles are in the same gene, but display what is known as intragenic complementation, that is, while the a1 and b1 alleles are both recessive, leading to a mutant phenotype as homozygotes, the a1/b1 heterozygote displays a wild type phenotype. Intragenic complementation is relatively rare, since generally both allelic versions of the gene product are inactive (amorphic/null or hypomorphic), but there are cases, particularly involving polypeptides with multiple, distinct structural domains and proteins that contain multiple copies of the same gene product, in which the combination of allelic polypeptides retains sufficient activity to produce a wild type phenotype. Various other types of allele-specific interactions are possible.[518] This is one reason that researchers often examine the phenotypic effects of multiple alleles of a gene, as well as allelic phenotypes in a number of different genetic backgrounds. Genetic backgrounds can have substantial effects on phenotype, expressivity, and penetrance.[519] Given that different species (such as mice and humans) have dramatically different genetic backgrounds, arising from their evolutionary histories and ecological adaptations, it is not surprising that the same mutation defined in one organism can produce a different phenotype in another.[520]

References

[516] If interested, check out: The art and design of genetic screens

[517] Daniel The art and design of genetic screens: Drosophila melanogaster

[518] Genetic Background Limits Generalizability of Genotype-Phenotype Relationships (a paper cited above)

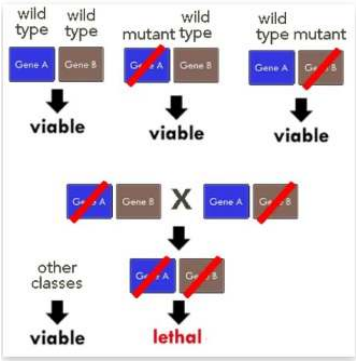
[519] Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases

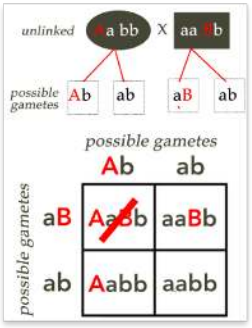
[520] Null mutations in human and mouse orthologs frequently result in different phenotypes

This page titled 15.5: Genetic complementation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.6: Interacting traits- synthetic lethality and co-dominance

The physical linkage of genetic loci is only one of the ways that genes interact (↑), others involve interactions between gene products and the biological processes they mediate. There are interactions between proteins encoded and other gene products within the concentrated confines of the cell (considered later). Perhaps the most dramatic type of interaction, from the perspective of phenotype is known as synthetic lethality.[521] In such a situation, often but not necessarily, carried out with dominant alleles of two distinct genes, both heterozygotes on their own are viable, while the double heterozygote is dead–the combination is lethal (→). Similarly, it can be the case that two recessive alleles, individually viable as homozygotes and non-viable or display a different phenotype in double homozygous individuals. We can detect the presence of synthetic lethality through various crosses in which individuals with specific expected combinations of alleles fail to appear in the progeny of a cross (←). As long as we can identify the phenotypes of expected progeny, and so count their presence in a population, such deviations can be detected using a χ^2 analysis similar to our approach to identifying linkage.





The presence of synthetic lethality suggests that the two gene products are involved in a common, essential process. Less extreme interaction outcomes are associated with other types of interactions between alleles of different genetic loci; these are recognized because the phenotype produced by the presence of both alleles is different from the phenotype of either allele on its own. This is distinctly different from the behavior of Mendel's genetic factors, as noted by Weldon and others).

Synthetic phenotypes can arise in a number of ways. A process may depend upon multiple gene products interacting to form a functional complex, necessary to produce a trait. Two, often paralogous, genes may produce functionally similar gene products. If one is mutated so as to produce little or no functional gene product, the product of the second gene may be sufficient, but if both are mutant, not enough of the functional complex is formed. In some cases, alleles of both genes may be recessive, but when present together, they may appear dominant. Such a situation can be generated using various methods, producing what is known as a "sensitized background" that reveals the roles of gene products in molecular and cellular contexts. There are many complexities associate with mutant behaviors.

Questions to answer:

218. What types of plausible scenarios can you imagine by which the products of two distinct genetic loci interact to produce a synthetic lethal phenotype.
219. If a functional gene is missing from a syntenic region, what might have happened to it?
220. How might the level of expression of one gene influence the phenotype associated with another?

Question to ponder:

Why (and how) did Mendel exclude interacting alleles from his analysis?

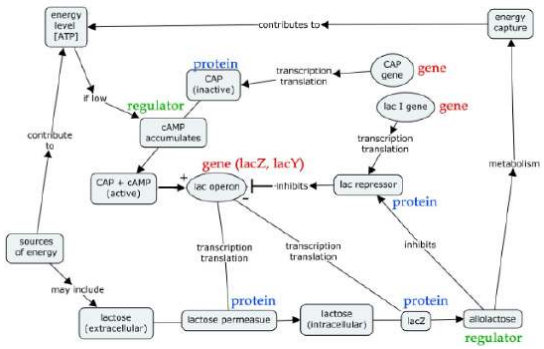
References

[521] Synthetic lethality and cancer

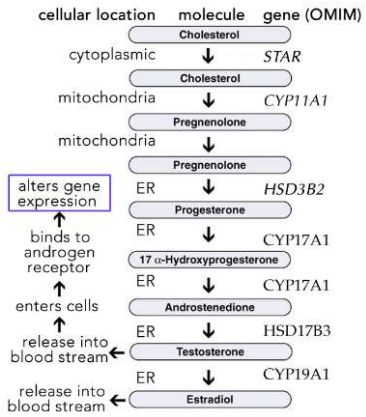
This page titled 15.6: Interacting traits- synthetic lethality and co-dominance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.7: Interacting traits- epistasis

The molecular systems that produce biological traits and behaviors (phenotypes) involve multiple gene products that influence macromolecular complexes and cellular behaviors and occur within pre-existing, adaptive, and homeostatic systems characterized by multiple feed-forward and feed-back interactions. A particular mutation will perturb the system in particular ways, and often unpredictable ways. As an example, let us return to the lac operon. We can generate a schematic of the interactions between genes, gene products, and regulatory molecules - in this case lactose, allolactone, and cyclic AMP (↓). Based on such a scheme, we could generate a mathematical model to serve as the basis for making predictions about the effects of mutations in the various genes involved in the process. Such models would include descriptions of the concentrations of the various components (gene products and other molecules), binding affinities between molecules, and such. If the model's predictions were confirmed experimentally, we would have increased faith that our understanding of the system is accurate and possibly complete. If the predictions are not confirmed, it is possible (likely) that we have missed important components of the system. At the same time, while DNA-dependent, RNA polymerase is a necessary component of the system, required to expressed the genes involved, it is not explicitly included in our model because mutations that alter polymerase function would disrupt essentially all cellular systems. These are known as pleiotropic effects arising from a mutation (allele).[522] Similarly, if components of the system are involved in other processes, the model may be influenced by indirect effects on those processes. Increasingly researchers are relying on generative AI models raising interesting questions about what understanding a biological system means.[523]



In a number of systems, there are parts of the network that act in a sequential manner, with one gene product acting on the next "down-stream" part of the system. An example is the testosterone/estradiol system. Both testosterone and estradiol are derived from cholesterol and both play key roles in the generation of male and female sexual characteristics in mammals. Beginning with cholesterol (and ignoring the reactions involved in cholesterol synthesis), we find a number of gene products, identified by their On-line Mendelian Inheritance in Man (OMIM) designations, that catalyze the steps in the pathway (→). These reactions occur in both the cytoplasmic and mitochondrial compartments of the cell. The entry of cytoplasmic cholesterol into mitochondria is facilitated by the STAR gene product. Within mitochondria an enzyme catalyzes the reaction that transforms cholesterol into pregnenolone that then can leave the mitochondria and accumulates in the endoplasmic reticulum (ER). A series of reactions then leads to the formation of testosterone, the "male" hormone, which can be transformed into estradiol, a "female" hormone. Estradiol is also involved in male reproductive function.[524] Both testosterone and estradiol are released into the blood stream, allowing them to interact with cytoplasmic proteins (androgen/estrogen receptors) in various tissues and cell types. Testosterone and estradiol act as allosteric effectors of transcription factor proteins, activating them to enter the nucleus and regulate the expression of specific target genes.



In the context of a pathway analysis, we find that the effects of mutations/alleles of genes can be ordered. For example, assume that there is a mutation in the CYP17A1 gene which leads to a non-functional version of the encoded protein. In an individual homozygous for this CYP17A1 mutation, we expect to find the accumulation of progesterone in the ER. Now consider a second null mutation in the CYP11A1 gene; an individual homozygous for this mutation would be expected to accumulate cholesterol in their mitochondria. Can you predict the phenotype, in molecular terms, of an organism homozygous for null alleles in both CYP17A1 and CYP11A1 genes? (assuming of course that both are viable and fertile). The result of such a genetic analysis can establish what is known as the epistatic relationship between genes (or more accurately gene products) in a particular process.[525]

A complicating aspect of most actual interaction pathways is that there are various forms of feed-back and feed-forward interactions that can influence the behavior of a pathway when its normal functioning is inhibited or perturbed. As an example, the

accumulation of one compound might influence the expression of other genes, or the activity of other enzymes. In some cases, this can result in a by-pass of the block, so that phenotypic effects are minimized. Given their roles as regulators of transcription factor activity, their presence (and concentration) or absence can influence the expression of multiple genes. At this point, what is important is to consider what the phenotypes of various genetic crosses might tell you about underlying molecular and cellular systems, while recognizing the limitations of such predictions.

Questions to answer:

221. What factors limit the usefulness of genetic crosses to establish epigenetic relationships?
222. How are genetic pathway maps useful, and what are their limitations?
223. Why is a forward genetic screen unlikely to identify all components of a particular process?
224. Consider a dominant allele in which the associated phenotype is lost on a particular genetic background. How might you reveal the presence of such an allele through a genetic analysis?

References

[522] Pleiotropy: One Gene Can Affect Multiple Traits

[523] see substack essay: What (exactly) does it mean to "understand" life or the brain? and De Lorenzo. The principle of uncertainty in biology: Will machine learning/artificial intelligence lead to the end of mechanistic studies?

[524] see The role of estradiol in male reproductive function

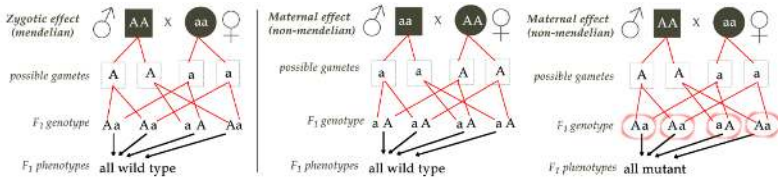
[525] Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems

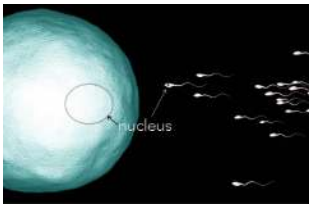
---

This page titled 15.7: Interacting traits- epistasis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.8: Maternal and paternal effects

Like any other process or trait, embryonic development can be studied and underlying mechanisms identified through the generation and analysis of mutations in the genes that influence the processes involved. From a genetic perspective, there are two general types of mutations - those that effect the formation of gametes, particularly the egg and those that effect development of the embryo directly or indirectly. Mutations (alleles) that influence oocyte formation and the maternally-constructed developmental environment are known as “maternal effect mutations”. Take for example a recessive allele “a” - it may be a typical zygotic effect allele or a “maternal effect” allele. How to distinguish them? Begin with a standard cross between homozygous individuals, the outcome will be the same whether the male or the female is homozygous for the "a" allele (↑). The traits Mendel use all behave in this way. The outcome of the cross will be dramatically different for a maternal effect allele if the female is homozygous for the wild type "A" or mutant "a" allele (↑). The genotype of the female parent (aa) rather than the genotype of the offspring (Aa) determines (influences) the phenotype, a decidedly non-Mendelian behavior. A similar situation arises if the maternal effect allele is dominant, assuming it specifically effects female reproductive success (fertility).





Gamete dimorphism (the differences in gamete size, morphology, and behavior)(↓) implies that some genes preferentially influence oocyte/egg or sperm behaviors and functions. In a number of organisms, particularly those that develop rapidly and outside the mother, most of the gene products and nutrients needed to support the early development of the new organism are supplied by the egg. Defects in the oocyte, due for example to recessive alleles in a homozygous mother, may lead to defects in the behavior of the fertilized egg and subsequent embryo that cannot be rescued by a sperm cell carrying a wild type (dominant) allele - they are dependent upon the maternal genotype and independent of the offspring's genotype. Paternal effects have also been identified.[526]

References

[526] What is a paternal effect?

This page titled 15.8: Maternal and paternal effects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.9: Mitochondrial inheritance

A obvious example of a maternal effect involves the inheritance of mitochondria. As noted previously, mitochondria have their own genomes, circular DNA molecules known as mtDNAs. mtDNA encodes a number of genes: 37 known in human. mtDNAs can, like any DNA molecule, accumulate mutations, whether during replication or in response to free radicals generated during the course of aerobic respiration. Mitochondria are typically supplied to the zygote by the oocyte/egg and not the sperm. The mitochondria present in the sperm cell either do not enter the egg or if they do, they and their DNA are degraded. Mutations in mtDNA can lead to dysfunctional mitochondria that can lead to a number of phenotypes. Defects in the mitochondrial genomes present in the egg cannot be rescued by sperm, and so produce a maternal effect on the zygote.[527]

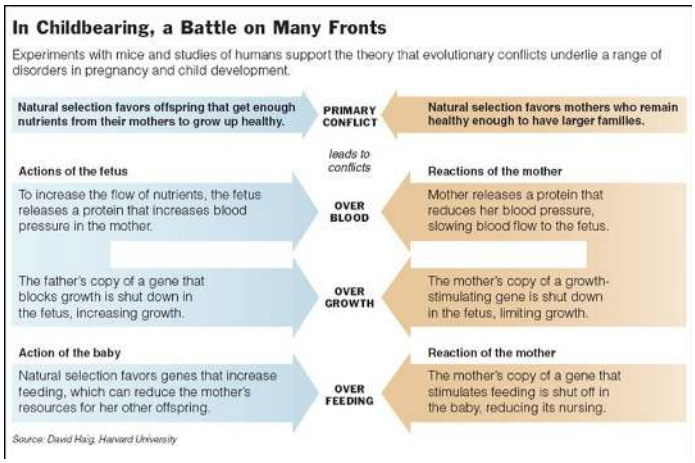
A complexity in the study of mtDNA mutations is that each mitochondrion contains a DNA molecule. A cell contains hundreds to thousands of mitochondria. Different cell types within the same organism can contain different numbers of mitochondria and differ in their dependence on mitochondrial function. The result is that we are looking at populations of mitochondria with the possibility of a number of different mitochondrial genotypes. The numbers of mitochondria in a cell or cell lineage can change, raising the possibility of population bottlenecks and associated genotypic changes. There can also be somatic selection - the differential replication of somatic cells based on their mitochondrial genotypes and activity. In any one cell or tissue, mitochondrial-dependent phenotypes will be influenced by the mtDNA genotypes present. A detailed consideration of mitochondrial influences on disease phenotypes in humans and other organisms is beyond us here, but the interested can find a database of mitochondrial DNA mutations at the MitoMap web site [link].

References

[527] Mitochondrial DNA mutations and human disease

This page titled 15.9: Mitochondrial inheritance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.10: Imprinting- conflicts between mother, father, and fetus



While we have considered sexual selection and the various conflicts between the reproductive interests of the two sexes particularly in sexually dimorphic species, other conflicts can occur between mother and fetus, particularly in placental mammals, such as humans (↓). In these organisms the risks to, and costs on, the mother in raising an embryo can be substantial. Under such a condition, carrying a pregnancy to term has the potential to harm the mother, and there may be situations in which it is to the mother's benefit if the pregnancy ends. In contrast, the embryo's (and in many cases the father's) overriding interest is to be born. Under these conditions, the embryo can benefit from suppressing or modulating the mother's "self-defense" responses. In turn embryonic defense strategies can be countered by maternal effects on zygotic gene expression. Both strategies involve a process known as imprinting, in which the DNA of sperm and egg are modified differently.[528] Imprinting involves sequence specific post-replication DNA modifications; these changes are epigenetic, they do not alter nucleotide sequence but influence when and where a gene is expressed. Because patterns of imprinting are different in males and females, the maternal and paternal alleles present in a new diploid organism may be expressed differently. In some cells only the maternal allele of an imprinted gene will be expressed, whereas in other cells only the paternal allele will be expressed.[529]

In a typical scenario the paternal (sperm-supplied) copy of a gene that promotes embryo growth (which if excessive can threaten the survival of the mother) is over-expressed. In response, the maternal (egg-supplied) copy of the gene is turned off. This balances the effect of the paternal copy, leading to normal development. A similar situation can occur if a maternal gene is expressed, leading to the suppression of expression of the paternal copy. Developmental problems can arise, however, if (for example) the paternal, expressed copy of the gene is defective or visa versa.[530] Imprinting often involves the epigenetic modification a gene's promoter region. Imprinting complicates things. As a result a deeper consideration of the general topic of human fertility is best addressed elsewhere.

Questions to answer:

225. How many mechanisms can you imagine that would lead to the expression of different genes in different regions of an embryo?.
226. Describe how imprinting can impact Mendelian allele behavior(s)?
227. Most of the genes involved in mitochondrial function are nuclear; how might that influence the phenotypes of mutations in mitochondrial DNA?
228. If you were to predict which tissues would be more severely effected by mutations in mitochondrial DNA, what would you base your predictions on?

Questions to ponder:

- Explain the evolutionary pressures egg and sperm behavior and the speed of early development.

References

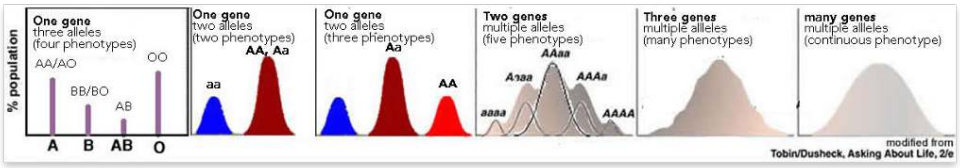
[528] Genomic Imprinting: http://learn.genetics.utah.edu/conte...cs/imprinting/

[529] The origin and evolution of genomic imprinting and viviparity in mammals.

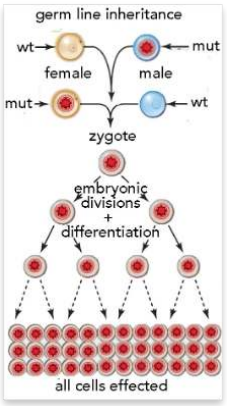
[530] genomic imprinting

This page titled 15.10: Imprinting- conflicts between mother, father, and fetus is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.11: Estimating the number of genes involved in a particular traits



Mutations that become alleles can be seen as lying along a continuum. At one end of this continuum are alleles that behave according to Mendel; these are alleles of a gene that control what we might term discrete features of a particular trait, such as the human (ABO) blood type or a number of genetic diseases that you either have or you do not have (↓ left side). As the number of genes (and the alleles) that influence a particular trait increases, the distribution of versions of the trait, for example, height, approaches a smooth curve, a curve often termed a bell curve (right side ↓). Such a distribution is characterized by a mean, a median (which is the same as the mean when the curve is symmetrical), and a standard deviation, which reflects the width of the distribution. The alleles in the various genes involved in a trait can display dominant, recessive, or synergistic (interactive) behaviors.



An important feature of germ line alleles is that all cells of the resulting organism (with the exception of the gametes produced by that organism and any new somatic mutations) will have the same genotype (next↓page). That said, for heterozygous loci, single cell RNA sequence has revealed what is known as monoallelic gene expression, where one or the other allele is expressed. The result can be differences (and selection) between genetically identical cells.[531]

Phenotypes associated with a particular allele can vary between cell types. Genes that encode common, often termed house-keeping functions, generally have global effects, while those expressed in only one or a few cell types may have effects in only these cells. The fact that many genes have been duplicated during evolution, to form paralogous genes that often have similar although rarely identical functions can also influence the phenotypes associated with various alleles. A gene may be expressed in a particular cell type, but the behavior of the gene product may be more or less critical in those cells because of the presence of functionally complementary gene products due to expression of genes in various compensatory or parallel molecular processes and pathways. We saw this effect in our discussion of somatic mutations; a germ line mutation can be inherited but not have a discernible phenotypic effect until a subsequent somatic mutation occurs that disables or alters the functioning copy of the gene, or compromises the function of a complementary gene.

References

[531] Monoallelic Gene Expression on Mammals.

This page titled 15.11: Estimating the number of genes involved in a particular traits is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.12: On the nature of mutations (again)

A point mutation changes a single nucleotide within a gene. To produce a phenotypic effect, a point mutation needs to alter a regulatory region, a coding region, or sequences involved in splicing. A point mutation that alters a codon without changing the encoded amino acid is referred to as a neutral or synonymous mutation; such a mutation can have effects if it changes a codon that is recognized by a highly expressed tRNA to one recognized by a infrequently expressed tRNA, an effect associated with codon bias. tRNAs with different codon-anti-codon interactions can bind with different affinities. The result is that some codons are misread more frequently than others, leading to an increase probability of a frame-shift or even translation termination.[532] When a single nucleotide change alters the amino acid encoded it is referred to as a missense mutation; such a mutation can influence the behavior of the encoded polypeptide. If, for example, the altered amino acid forms part of the active site of an enzyme, its three-dimensional structure, sites of post-translational modification or processing, or influences interactions with water or other polypeptides; it can alter the polypeptide's assembly, activity, stability, and cellular localization. For example, a single amino acid change can alter the energetics of polypeptide folding. The polypeptide may misfold and be unstable at lower (cold-sensitive) or increased (heat-sensitive) temperatures. This underscores the fact that organisms typically have an optimal growth temperature range. Its polypeptides/proteins are optimally functional in that temperature range and less functional outside it, where they may unfold or adopt non-functional configurations. Abnormal protein folding can lead to function-disrupting interactions with other molecules in the crowded cytoplasm.

Another type of "point" mutation, a non-sense mutation, introduces a non-coding or stop codon upstream of the normal translation termination site. A non-sense mutation leads to a truncated polypeptide that can fail to fold or function correctly, and may interact inappropriately with and disrupt the function(s) of other proteins. Because such mutations are generally disruptive there are mechanisms in eukaryotic cells when such a mutation occurs early a coding region that can trigger the nuclease mediated degradation of the mRNA, a process known as non-sense mediated decay (discussed above). Degradation of the mRNA suppresses the synthesis of the mutant polypeptide and so mitigates the effects of an aberrant (truncated) gene product.

References

[532] Different organisms vary in their use of different codons, which form the basis of what is known as "codon bias". Optimal expression of gene from organisms in another (e.g. a bacterium) often involves optimizing the codons used.

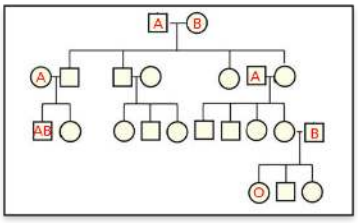
This page titled 15.12: On the nature of mutations (again) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.13: Alleles, traits, and genetic diseases in humans.

The range of alleles present within a population influences the various phenotypes observed - ranging from differences in body size and shape to disease susceptibility. Some (rare) alleles produce discrete traits that behave in a modified Mendelian manner, such as ABO blood type. There are three common alleles of the *ABO* gene that encode the protein ABO glycosyltransferase. A and B alleles behave in a dominant manner with respect to O, which acts in a recessive manner. A and B behave in a co-dominant manner with respect to one another; when both are present they generate a new phenotype, the AB phenotype. The distribution of these alleles in different human populations appears to be due, at least in part, to founder effects and selective advantages associated with specific alleles in specific environments.[533]

> *"Mourant suggested that the major differences in the geographical distribution of ABO blood groups may be the consequence of epidemics that occurred in the past. The concept of evolutionary selection based on pathogen-driven blood group changes is currently supported by studies on the genetic characterization of the ABO blood group in Neanderthals and ancient Egyptian mummies. These studies suggest a potential selective advantage of the O allele influencing the susceptibility to several different pathogens responsible for diseases such as severe malaria, H. pylori infections and severe forms of cholera".*
> *(see footnote 536)*

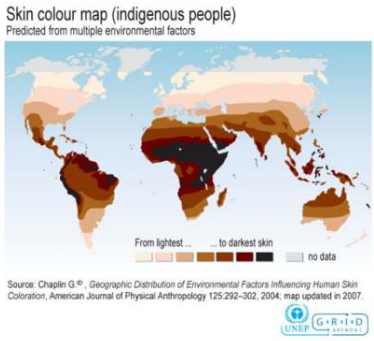
Because blood type can be determined unambiguously, the mode of interaction of these alleles is well defined, it is possible to trace their inheritance across multiple generations. If we know an individual's blood type, we have an initial (although incomplete) model of their genotype. As we examine the phenotypes of their progeny, we can further constrain their genotypes. In such studies, we assume that we know with certainty who mated with whom, something that may not be the case. In this family tree the presence of an AB individual in the second generation (←), indicates that the male (☐) parent must have had an AB or BO genotype. Other genotypes could not have produced an AB offspring. Similarly in the lineage giving rise to the O individual, we can conclude that its male parent (☐) had to be BO, while its female parent (◦) had to be OO. The more of the individual phenotypes we know in a pedigree, the more we can constrain the genotypes of members of their lineage.



In the (more) modern world we can use molecular markers to directly identify the alleles present in a specific individual. Pedigree analysis can lead to potentially embarrassing or disruptive conclusions; it can reveal that a father cannot be the biological father of a child. Generally, but not always, who the mother of a child is unambiguous.[534] Molecular details can influence these conclusions. The reactions catalyzed by the A and B enzymes are dependent upon an "upstream" enzyme - a fucosyltransferase. The product of another gene, it is necessary to create the substrate upon which the A and B enzymes act. If fucosyltransferase is not present due to a non-functional allele of that gene, a person with an A or B allele will have O type blood.

> *"When people think of skin color in Africa most would think of darker skin, but we show that within Africa there is a huge amount of variation, ranging from skin as light as some Asians to the darkest skin on a global level and everything in between. We identify genetic variants affecting these traits and show that mutations influencing light and dark skin have been around for a long time, since before the origin of modern humans."*
>
> *– Sarah Tishkoff*

A type of trait that differs between populations, as well as between individuals within a population, is skin color. Skin color has been linked to exposure to solar UV radiation and the role of UV light in the synthesis of vitamin D.[535] The extent of exposure of skin to sunlight depends on a number of factors. As genomic studies include more people from geographically diverse groups, DNA sequence analyses have revealed that a number of genes are involved in the determination of skin color. As one might predict,



given that humans appear to have originated in Africa, African populations display the most dramatic genetic diversity in skin color, a prediction confirmed by direct observation. Genomic studies indicate that four genomic regions (genes) are responsible for ~30% of the variation in skin pigmentation; the remainder is due to allelic variation in a number of other genes.[536]

Based on modern primates, it appears that our primate ancestor had largely unpigmented skin; they were protected from sun damage by their fur. Skin pigmentation is expected to have increased as fur was lost, an adaptation to a more active (heat-generating) life style and dependent on more effective cooling of the body. (←) As human populations migrated away from their site of origin different levels of UV exposure impacted their adaptation to the antagonistic pressures of skin damage and vitamin D production, leading to selection pressures based on skin pigmentation.[537] As populations migrated away from the equator, reduced levels of skin pigmentation were selected.

References

[533] Beyond immunohaematology: the role of the ABO blood group in human diseases

[534] That said there are strange situations, often involving embryological events, that can lead to unexpected results [link to add]

[536] Evolution, Prehistory and Vitamin D

[537] Genes responsible for diversity of human skin colors identified: (paper) Loci associated with skin pigmentation identified in African populations

This page titled 15.13: Alleles, traits, and genetic diseases in humans. is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.14: Concordance between monozygotic twins and genetic influence on a trait

An interesting phenomenon that can be used to characterize the genetic contributions to a trait involves twins. There are two types of twins. Fraternal or dizygotic twins involve two eggs and two sperm, leading to two distinct embryos developing together. Fraternal twins are generally born in rapid succession and are no more or less closely related than any two siblings born years apart. In animals that typically have multiple offspring, the individuals born generally arise from distinct zygotes. In contrast, identical twins are known as monozygotic twins. Identical twins occur when a single sperm fertilizes a single egg and generates a single zygote that begins development. Subsequently the embryo fragments to produce two embryos that develop independently of one another.[538] With the exception of somatic mutations, DNA modification, and stochastic effects during embryonic development, the two individuals are initially genetically identical. Their genetic identity enables us to measure the genetic concordance of a trait.[539] For example, if a trait is determined solely by the individual's genetics, then the concordance between identical twins should be 100% (blood type is one example). In other cases, while genotype plays a role it is not completely determinative. As an example, in the auto-immune muscle weakness disease myathenia gravis, the genetic concordance is ~35%, a level that implies other factors play important roles in the appearance and progression of the disease (variable penetrance and expressivity).[540]

As we are talking about twins, it is worth noting another type of outcome, which is known as a chimera.[541] In a chimeric embryo, two initially distinct embryos fuse. A single organism develops, but with two distinct "fraternal sibling" genotypes.[542] When dizygotic fusion is complete, a single normal, albeit mosaic, embryo and mature organism is generated, a situation that can lead to genotypic confusion. When fusion is incomplete, or occurs at a later developmental stage, incompletely fused embryos are formed - what are known as conjoined twins.

References

[538] https://www.genome.gov/genetics-glos...dentical-twins

[539] Does Higher Concordance in Monozygotic Twins Than in Dizygotic Twins Suggest a Genetic Component?

[540] Immunopathogenesis in myathenia gravis and neuromyelitic optica.

[541] It is even possible to generate chimeric embryos between different species: Humanized mice and porcinized people.

[542] Such human chimeras have been identified: see 3 Human Chimeras That Already Exist and One Person, Two Sets of DNA: The Strange Case of the Human Chimera – Natural human chimeras: A review

This page titled 15.14: Concordance between monozygotic twins and genetic influence on a trait is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.15: Measuring evolution's impact on allele frequencies- Hardy-Weinberg

In a population, each gene is represented by a set of alleles. Typically, different alleles are present in different frequencies in different populations. These differences reflect the history of the population and evolutionary pressures. To determine whether evolution is occurring within a population, we use the Hardy-Weinberg (H-W) equation, based on the work of G.H. Hardy (1877-1947) and Wilhelm Weinberg (1862-1937) – published independently in 1908. Their analysis was based on a set of five assumptions: 1) the population is infinite - processes such as genetic drift do not occur; 2) the population is isolated - no individuals leave or enter it; 3) no new mutations occur; 4) mating between individuals is random (no sexual selection); and 5) there are no differential reproductive effects, that is, natural selection is not occurring.[543] Under these (completely unreal) conditions, the allele frequencies found in the initial population do not change over time. If, on the other hand, allele frequencies are found to change, selection (or some other process) must be involved.

Before Hardy-Weinberg there was a belief that dominant alleles were somehow "stronger" than recessive alleles and must over time "swamp" recessive alleles out of existence. This incorrect assumption was called "genophagy", literally "gene eating".[544] It occurs only if various alleles influence reproductive success differently, that is evolution is occurring.

Consider the situation where there are only two alleles (A and a) of a particular gene. If the frequency of A in the population is p, the frequency of a is q. It is clear (hopefully) that p + q = 1. We can calculate the frequency of homozygotes and heterozygotes by expanding the term $(p+q)^2$; simple mathematical considerations indicate that within this population, the probability of an AA homozygote is p^2, the probability of an aa homozygote is q^2, and the probability of an Aa heterozygote is 2pq, such that:

$$p^2 + 2pq + q^2 = 1.$$

How is this possible? Remember, both p and q are less than 1. Our null hypothesis is that these alleles are NOT subject to selection; they have no effect on reproductive success. Now look at the frequency of recessive homozygotes in the population and calculate the χ^2 value and use it to estimate whether the population is at equilibrium, that is, no evolutionary changes are occurring, or whether there is active selection for or against certain alleles. For example, it might be that homozygous recessive individuals are either not viable, not fertile, or that their offspring die more often than the offspring of others. Alternatively, the heterozygote might have a reproductive advantage compared to the recessive homozygote; such a reproductive advantage can maintain significant levels of an allele that is deleterious as a homozygote within a population. The classic example of such a behavior involves alleles associated with the human hemoglobin B (*HBB*) gene. Alleles of this gene are associated with a dominant trait, resistance to malarial infection, as well as an often lethal recessive trait, sickle cell anemia. While the recessive trait is subject to strong negative selection, the dominant trait is subject to positive selection in environments where malaria is endemic. The same allele is responsible for both traits.

References

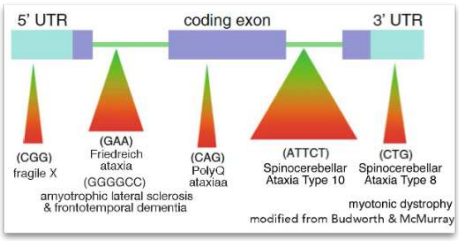
[543] Hardy-Weinberg Equilibrium: http://www.tiem.utk.edu/~gross/bioed/bealsmodules/hardy-weinberg.html

[544] genophagy

This page titled 15.15: Measuring evolution's impact on allele frequencies- Hardy-Weinberg is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.16: Genetic anticipation

There is a type of inherited allele that differs in interesting ways from conventional alleles, these are alleles that change from generation to generation, a behavior that has been termed genetic anticipation. Such alleles are associated with what are known as “trinucleotide repeat” expansion diseases. Some alleles involve repeated nucleotide sequences and generate what are known as microsatellite expansion mutations. Such repeated sequences (3 to 6 repeating units) account for ~30% of the human genome. Nucleotide repeat expansion diseases include several forms of mental retardation, Huntington’s disease, inherited ataxias, and muscular dystrophies.[545] There are regions of repeating nucleotides in the genes involved. Because of the "slippage" of DNA polymerase during DNA replication, the number of repeats can increase or decrease. The result? the allele delivered to an offspring can be more deleterious than the allele present in the parent - over generations, the symptoms of such an allele grow more and more severe. The length of the repeat correlates with the age of disease onset, but the age of onset is variable between individuals with the same repeat length, suggesting the impact of other factors. In addition to standard inheritance, many of these genes play roles in the function of nervous tissue, and it is possible that somatic (as opposed to germ line mutations) influence the allele's associated phenotypes. As an example, there is evidence that genetic anticipation is important in the context of schizophrenia and bipolar disorder, which together occur in ~1% of the population and have an estimated ~80% heritability risk, which means that on average, about 80% of the differences between individuals is due to genetic factors. Of course such estimates depend critically on how accurately various phenotypes can be quantified.



Mechanisms: Where nucleotide repeats are found and where their expansion leads to disease (→) suggests possible mechanisms behind the pathogenic state. Pathology-associated nucleotide expansion regions have been found within the transcribed regions of genes, including 5’ and 3’ untranslated regions and introns. When such a domain occurs in a coding region it can lead to stretches of repeating amino acids in the polypeptide. These may lead to toxic interactions with other cellular components.

To illustrate the potential complexity (a full exploration is beyond us here), consider recent work on the role of a nucleotide expansion domain in the gene *C9ORF72* (OMIM: 614620) that encodes a polypeptide implicated in intracelluar vesicle trafficking. The *C9ORF72's* expansion domain has been linked to both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Studies indicate that the expanded nucleotide region is targeted for inappropriate transcription; RNAs are synthesized bidirectionally. RAN (repeat-associated non-ATG translation) translation occurs from both sense and antisense expansion transcripts, resulting in the expression of six RAN proteins (antisense: Pro-Arg, Pro-Ala, Gly-Pro; and sense: Gly-Ala, Gly-Arg, Gly-Pro)".[546]These proteins accumulate in cytoplasmic aggregates in affected brain regions.[547] Another gene product, encoded by the *Supt4H1* gene (OMIM: 603555) appears to play a role in the inappropriate transcription of the C9ORF72 gene; reducing the levels of the *Supt4H1* gene product ameliorates the phenotypic effects of nucleotide expansion in C9ORF72.[548] The exact mechanisms of these types of alleles and associated phenotypes are complex, based likely on the effects of altered transcription on the functional roles of specific cell types.[549]

References

[545] A Brief History of Triplet Repeat Diseases

[546] Non-ATG–initiated translation directed by microsatellite expansions

[547] RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia.

[548] Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts

[549] C9orf72-mediated ALS and FTD: multiple pathways to disease

This page titled 15.16: Genetic anticipation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

15.17: The persistence of deleterious alleles

This page titled 15.17: The persistence of deleterious alleles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

16: Tools for Studying Genes and Genomes

16.0: Introduction

16.1: Synteny examined using Genomicus

16.2: Where is a gene expressed?

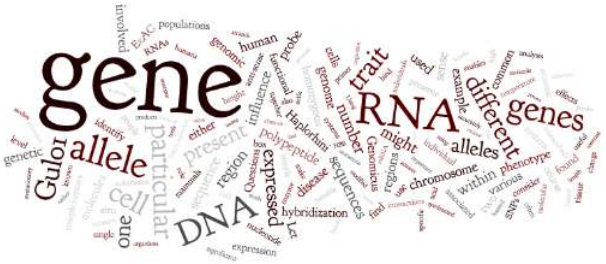
16.3: Using web-based bioinformatic tools- gnomAD

16.4: Using web-based bioinformatic tools- BLAST

16.5: Genome-wide Association Studies (GWAS)

This page titled 16: Tools for Studying Genes and Genomes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

16.0: Introduction



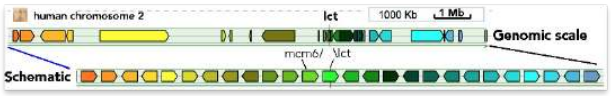
In which we consider a few of the tools available to examine the organization and evolution of genes and genomes and the role of genes and gene products in biological systems.

As we think about genes and their functional roles within biological systems it becomes increasingly useful to understand available methods, their power and limitations. We know a lot about biological system at the single cell (and single nucleotide) level, and can begin to piece together how they function. Here we introduce a few useful tools.

This page titled 16.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

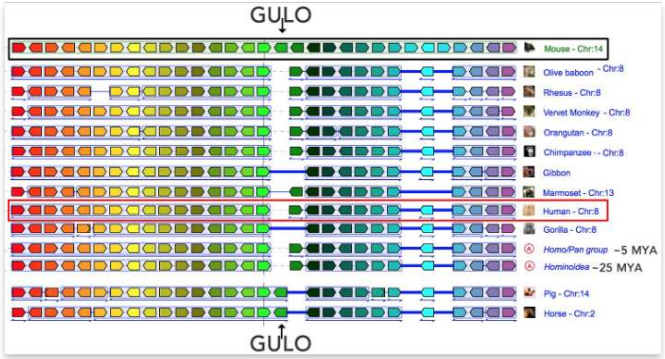
16.1: Synteny examined using Genomicus

In Genomicus [link] the user (you) inputs a gene name and the system displays the gene in its genomic context; where it is within a chromosome, as well as the genomic positions "of all its orthologous and paralogous copies in all the other sequenced metazoan genomes" together with "predicted ancestral genome structure".[550] We inputed the gene name *LCT* (OMIM: 603202)(↓). *LCT* encodes the enzyme lactase, normally expressed in infants and turned off as they mature into adults. The trait "adult lactose tolerance" is found in human populations associated with domesticated animals from which milk can be harvested. Adult lactose tolerance is associated with a ***failure*** to turn off expression of the *LCT* gene.[551] *LCT*expression in adults is negatively regulated by an enhancer element ~14 kbs upstream of the gene's promoter, located within an intron of the *MCM6* gene. Mutations within this enhancer element are associated with adult lactose tolerance, apparently the result of positive selection.[552]



Genomicus enables us to analyze the region around the *LCT* gene. Two views are possible. In the *genomic scale view*, the genes are displayed based on their actual size in base pairs), relative locations, and the direction of transcription, indicated by pointed boxes (↑). Different genes get different colors. Here indicates two genes transcribed in opposing directions (how is that possible?) Each pointed box indicates the size of the gene but not the positions of introns and exons. The intragenic regions between genes are indicated; their lengths are accurately displayed. In the *schematic view*, each gene is again indicated by a pointed box, but now all genes, no matter their actual length, are indicated by the same size box. It can be easier to recognize genes in the schematic view. On the web, holding your cursor on a gene (in either view) will display the gene name and more information. Note that the *MCM6* gene is located adjacent to the *LCT* gene (on chromosome 2). We could, if we wanted to, walk along the chromosome by inputing genes at each end of the region displayed. Genomicus also presents syntenic regions in other organisms, and provides predictions of the genomic organization of evolutionary ancestors.

To use Genomicus to study evolutionary change, let us consider the *GULO1* gene, another gene we introduced previously. In contrast to most vertebrates, the *Haplorhini* or dry nose primates are dependent on the presence of vitamin C (ascorbic acid) in their diet. It appears that a functional L-gulonolactone oxidase (*GULO1*) gene was lost in the last common ancestor of the *Haplorhini*. If we use the human genome as a reference, Genomicus fails to find the non-functional *GULO1* gene. If, however, we enter *GULO1* using the mouse or a *Strepsirrhini* (wet nose primate) genome, Genomicus finds the gene (↓), together with orthologs of the gene in the mouse and in a wide range of eukaryotes, including single-celled eukaryotes such as baker's yeast, which appears to have diverged from other eukaryotes about ~1,500,000,000 years ago. Each horizontal line in the diagram represents a segment of a chromosome from a particular species selected, together with predicted phylogenic (evolutionary) relationships based on synteny between species. We find that the genes surrounding the *GULO1* locus in mammals are largely the same; all mammals are estimated to have shared a common ancestor ~184 Mya. The syntenic region around the *GULO1* gene, and the presence of a *GULO1*gene in yeast and other distantly related organisms, suggests that the ability to synthesize vitamin C is a trait present in the ancestor of all eukaryotes.



An examination of the resulting map reveals the absence of humans and other Haplorhini primates – Whoa!!! what gives? The explanation, it turns out, is rather simple.[553] There is (apparently) no functional *GULO1* gene in any*Haplorhini* primate. But the *Haplorhini* are related to the rest of the mammals, aren't they? We can test this assumption, and circumvent the absence of a functional *GULO1* gene, by exploiting synteny – when we search for genes in the neighboring region we find that this region, with the exception of *GULO1*, is present and conserved in the *Haplorhini* (↑). The *GULO1* syntenic region lies on human chromosome 8 (highlighted by the red box) and similar syntenic regions are found in the homologous chromosomes of other *Haplorhini* primates. Genomicus analysis enables us to make a number of testable predictions. A newly discovered *Haplorhini* primate will be predicted to share the same syntenic region and will be missing a functional *GULO1* gene, whereas a newly discovered *Strepsirrhini* primate, or any mammal that does not require dietary ascorbic acid, should have a functional *GULO1* gene within this syntenic region. We

might also predict that adding a functional *GULO1* gene, for example from a mouse, would make a human cell (or a human) vitamin C independent, perhaps something a future genetic engineer with do.[554] Such an analysis also reveals that genes and chromosomal regions can and often do move around within the genome.

Questions to answer:

231. If you were to add a mouse Gulo1 gene to a human genome, where would you put it and why?
232. If a gene is missing from a syntenic region, what might have happened to it?

Questions to ponder:

- Would growers of citric fruits be right in working to ban the genetic engineering of vitamin C independent people?

References

[550] Genomicus in 2022: comparative tools for thousands of genomes and reconstructed ancestors

[551] Lactose digestion and the evolutionary genetics of lactose persistence

[552] World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection.

[553] see Visualizing and teaching evolution through synteny

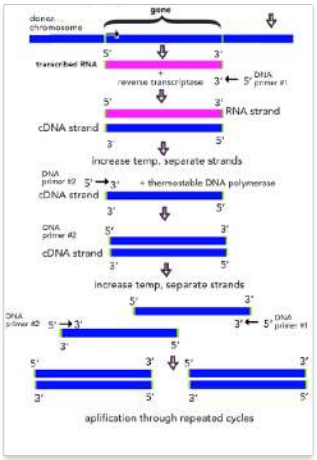
[554] Functional rescue of vitamin C synthesis deficiency in human cells by expression of murine l-gulono-γ-lactone oxidase

This page titled 16.1: Synteny examined using Genomicus is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

16.2: Where is a gene expressed?

When we consider the role of a particular gene in generating a particular phenotype, an important question is whether the effect is direct or indirect - is the gene expressed in the cells/tissues/organ that produces the phenotype or does it influence an earlier event? How, exactly do we know where and when a specific gene is expressed within an organism? This question can be answered by methods that reveal either where a particular gene is expressed or where the polypeptide encoded by the gene accumulates. We consider them briefly here.

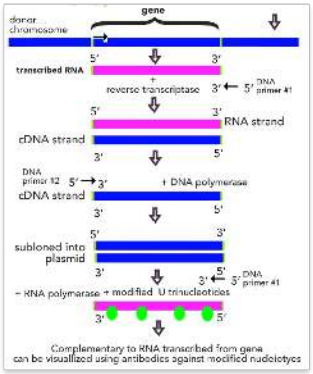
RT-PCR: Polymerase chain reaction (PCR) is a transformative technology, made possible by the discovery of heat stable DNA-dependent, DNA polymerases, isolated from archaea that live in high temperature environments (thermophiles and hyperthermophiles). PCR can be used to isolate and manipulate genes, as well as to visualize gene expression and in genome sequencing.



In the context of gene expression analysis, we use PCR to quantify the amount of a particular transcribed (expressed) RNA within a particular tissue, or together with single cell isolation technology (→). After cell isolation, we make a DNA copy of the transcribed RNA - this enables us to avoid the genomic DNA copies of genes that are present in every cell. We isolate RNA from a tissue and then use a "reverse transcriptase" (RT). RTs are enzymes derived from viruses and transposable elements.[555] In RT-PCR, we use the RT enzyme and a DNA primer to make a DNA copy complementary to the RNA strand, a cDNA. The RNA-DNA strands are then separated by increasing the temperature of the system. A second DNA primer acts together with a thermostable DNA-dependent, DNA polymerase to generate a doubled stranded copy of the cDNA with primer sequences at each end. Next comes the amplification stage of the reaction. The two strands are separated by increasing temperature. The original two DNA primers are present in excess, so that when the temperature is lowered, they bind back specifically to the DNA strands, and initiate a new round of DNA-dependent, DNA synthesis. With each cycle the number of DNA strands doubles, so that there is exponential growth in the number of specific DNA molecules with each cycle. Because the primer sequences, designed by the investigator and synthesized *in vitro*, are complementary to, and specific for, a particular gene sequence (the RNA of interest), one expects to amplify one and only one of the RNAs (gene products) present in the tissue. If the gene is not expressed, no amplified DNA will be synthesized. By using various tricks (beyond us here) the process can be made quantitative, so that it is possible to accurately compare the numbers of different types of RNA molecules (the products of particular genes) present in the original sample, a measure of the level of gene expression, at least at the RNA level. With different sets of primers, it is possible to quantify the expression of various splice forms of a gene. You can get hands-on training on PCR at the SkillsCenter [PCR1][PCR2] (if you are in Boulder!)s

It is now possible to isolate and sequence the RNAs (or rather cDNAs derived and amplified from mRNAs) from single cells to characterize the genes expressed in individual cells.[556] Because mRNA is used, only exon sequences are (generally) included - and the result is known as an exome sequence. This method can characterize the genes expressed in a particular normal or cancer cell types.[557]

In situ hybridization: A limitation of the RT-PCR approach is it uses homogenized tissue samples or individual cells isolated from a tissue. Spatial resolution is lost. To see which cells are expressing specific genes in a tissue's spatial context we use other methods. The most common is known as in situ hybridization. When a gene is expressed, an mRNA molecule complementary to one strand of the gene is synthesized, and these “sense” RNAs accumulate in the cells that express the gene (there is little evidence for significant transport of RNA from cell to cell, across the plasma membrane.)[558] To identify cells that express a gene, we generate modified “anti-sense” RNA molecules (→). Typically, we first isolate and subclone a DNA molecule that encodes the sense and antisense RNA of a gene’s expressed (exonic) region – this can be based on a cDNA generated from an mRNA or a genomic exon. Using specific primers, recognized by different bacteriophage-derived DNA-dependent, RNA polymerases, we can generate either sense or anti-sense RNA molecules. In these reactions fluorescein or digoxygenin modified forms of the RNA nucleotide UTP are used; these modified

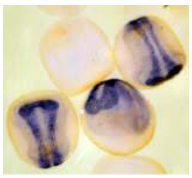


nucleotides can be used by the polymerase and are incorporated into the newly synthesized RNA. These are known as probes.

The overall process is relatively simple. The tissue is chemically stabilized and permeabilized so that molecules can diffuse into and out of it and then incubated with either a gene-specific sense or anti-sense probe. Because of the complementary nature of nucleic acids, the anti-sense probe RNA will bind to mRNAs, generated during gene expression. In contrast, the sense probe is the same sequence as the RNA transcript, and so does not bind. The sense probe is used as a null control, since it should not be complementary to any of the RNAs present in the sample. By controlling the hybridization temperature, we remove low affinity, non-specific interactions, leaving high affinity sense (mRNA)-anti-sense complexes. The probe is retained in regions that express the gene, and washed away from regions where the gene is not expressed. The level of binding to genomic sequence is too low to be visible. Antibodies, conjugated with various enzymes (typically alkaline phosphotase or horseradish peroxidase) can then be used to recognize the modified probe RNA:mRNA complex, and color-generating reactions, catalyzed by the enzymes, allows the distribution of probe to be visualized. The example here (←) is a neurula stage *Xenopus laevis* (clawed frog) embryo in which a gene (Snai2/Slug) expressed in the neural crest has been visualized by in situ hybridization.[559] In situ hybridization can also be carried out using fluorescent probes, allowing for higher resolution imaging. These methods can provide single cell resolution and can distinguish cells that do and those that do not express a particular gene. The specificity of the technique is influenced by the length of the probe and the hybridization temperatures used.



Single cell RNA Sequencing: The advent of more efficient DNA sequencing methods, together with PCR-based amplification, make it possible to isolate and sequence the RNA molecules within a single cell. Once sequenced, the number of molecules of each RNA (each gene product) can be counted to provide a catalogue of the genes and their level of expression within each cell. In heterozygotes single cell RNA sequencing can reveal whether one or both alleles are expressed. The result is that cells that were once considered identical have been shown to vary in terms of gene expression. These variations can give rise to variations that can influence cell behaviors and organismic phenotypes.



Immunocytochemistry: One limitation of RT-PCR, in situ hybridization, and single cell RNA sequencing is that they monitor RNA levels. In cases where the ultimate gene product is a polypeptide, it can be the case that RNA levels are not strictly correlated with the level of the accumulated polypeptide. One approach to avoid this disconnect is to use antibodies, proteins generated by the vertebrate immune system that bind specifically to particular molecular targets (epitopes). We will ignore how antibodies are generated since it involves an understanding of the immune system, a complex cellular system, but basically antibodies act very much like anti-sense RNA in situ probes, binding to specific molecular targets. A full characterization of the proteins present in a cell or tissue relies on physicochemical approaches, such as mass spectrometry, to define the proteome (another subject beyond us here).[560] The example provided here (→) is a neurula stage *Xenopus laevis* embryo stained for the transcription factor protein Sox3.

Questions to answer:

233. A gene can be spliced various ways - design primer sets to distinguish the splice variants of a gene.
234. Explain why a sense strand RNA probe serves as a useful control for in situ hybridization studies; what does it control for, and why does it work?

Questions to ponder:

- Why might the number of polypeptides in a cell differ from the number of RNAs that encode it?

References

[555] insert reference to reverse transcriptase.

[556] A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications

[557] see Defining murine organogenesis at single-cell resolution

[558] although things may actually be somewhat more complex: see Brain Cells Share Information With Virus-Like Capsules

[559] from: An NF-κB and Slug Regulatory Loop Active in Early Vertebrate Mesoderm

[560] Here is an example of proteomic analysis: Region and cell-type resolved quantitative proteomic map of the human heart

This page titled 16.2: Where is a gene expressed? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

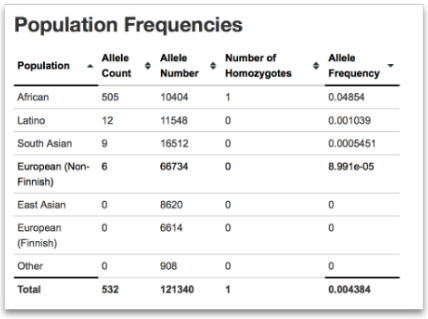
16.3: Using web-based bioinformatic tools- gnomAD

When studying a disease that appears to have a genetic component, it is common to identify the causative allele(s) of the gene involved. In the case of recessive alleles, such studies often involve pedigree analyses of families. Once a disease-associated allele is identified, it can be informative to determine whether that allele is found in individuals who do not display the disease trait. For dominant alleles, the presence of an allele without the disease phenotype indicates genetic background effects that influence the disease allele's penetrance and expressivity.

The accumulating library of exomic sequence data includes (July 2024) more than "730,947 exome sequences and 76,215 whole-genome sequences from unrelated individuals" from around the globe and continues to increase and become more diverse - more non-European people analyzed over time. This data library can be searched using gnomAD.[561] The user (you), inputs a gene's official name, as listed in OMIM or GenBank; gnomAD then displays sequence data from unrelated individuals. This allows for the identification of alleles and mutations present in a range of human populations. Let us try using the gene associated with sickle cell anemia, the *HBB* gene (hemoglobin, beta, OMIM: 141900). Mutations (disease-associated alleles) in HBB have been implicated in a number of human diseases. The allele associated with the sickle cell phenotype involves a missense mutation from GLU to VAL, known as GLU7VAL (↓). We discover that within the gnomAD database of "normal", that is disease-free individuals, this allele occurs with a frequency of ~0.0044 (with a single homozygous individual identified)(→). The heterozygotic individuals would not be expected to display any overt phenotype under most conditions, while the homozygous individual would be expected to have sickle cell disease. The vast majority of the people with the HBB Glu7Val allele are of African descent, as is the one homozygous individual. When this was originally written (June 2019) there was only one other homozygous individual within the library (Glu122Gln). 71 out of 85 of the people carrying this allele are of African descent, as is the homozygous individual.



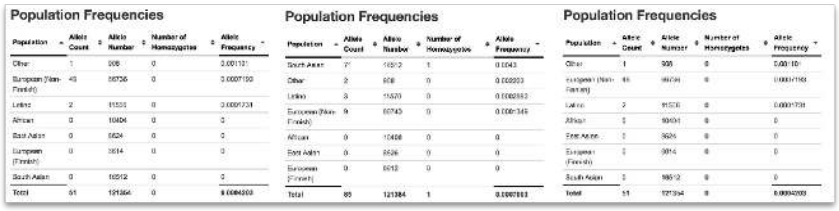
Variant	Chrom	Position	Consequence	Filter	Annotation	Flags	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
11-5248232-T-A (rs77121243)	11	5248232	p.Glu7Val	PASS	missense		532	121340	1	0.004384
11-5248233-C-T	11	5248233	p.Glu7Lys	PASS	missense		149	121340	0	0.001228



Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African	505	10404	1	0.04854
Latino	12	11548	0	0.001039
South Asian	9	16512	0	0.0005451
European (Non-Finnish)	6	66734	0	8.991e-05
East Asian	0	8620	0	0
European (Finnish)	0	6614	0	0
Other	0	908	0	0
Total	**532**	**121340**	**1**	**0.004384**

Data from gnomAD enables us to make informed guesses as to the impact of various genetic differences on the activity of a gene product.[562] If, for example, a dominant allele has been linked to a disease and yet that allele is detected in the gnomAD database, we might suggest either that that allele is not the cause of the disease, or that the effects of the allele are influenced by variations in other genes, leading to reduced penetrance and/or expressivity. If an allele is present in a heterozygous condition, but not a homozygous one, we can tentatively assume that negative selection is acting on the allele. If, on the other hand, alleles are present at different frequencies in different populations, that may be evidence for the action of positive selection dependent on environmental factors. The frequency of alleles in different populations can reflects the effects of founder effects, bottlenecks, and genetic drift. Take for example three other HBB alleles, p.Gly70Ser, p.Glu122Gln, and p.Gln40Ter (Ter=stop)(↓). We see that the Gly70Ser and Glu40Ter alleles are present primarily in non-Finnish Europeans, while the Glu122Gln allele is found in South Asians. It is not clear exactly what the effects of such missense mutations will be on the functions of the polypeptide – it could change folding, change interactions with other polypeptides and molecules, add or remove sites of post-translational modification, or change catalytic activities, if the polypeptide has such an activity. It is likely that the Glu40Ter mutation will produce a short, likely non-functional 39 amino acid polypeptide, compared to the 147 amino acid long wild type polypeptide. It is unlikely that the truncated protein is functional, but if it accumulates it could interfere with the function or molecular interactions of the full length wild type polypeptide.



Population Frequencies

Population Frequencies

Population Frequencies

References

[561] Genomics, Big Data, and Medicine Seminar Series – Daniel MacArthur

[562] The ExAC browser: displaying reference data information from over 60 000 exomes.

This page titled 16.3: Using web-based bioinformatic tools- gnomAD is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

16.4: Using web-based bioinformatic tools- BLAST

This page titled 16.4: Using web-based bioinformatic tools- BLAST is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

16.5: Genome-wide Association Studies (GWAS)

The majority of phenotypic traits are not associated with simple Mendelian inheritance, rather a number of different genetic genes and the combination of alleles determines the genetic aspect of the trait. In addition, there are often non-genetic, developmental and environmental factors involved. Nutrition factors during developing, exposure to toxins or pathogens and other stressors can influence the final phenotype. A classic example of a trait influenced by both genetics and environment is height. It is known as a quantitative trait, characterized by a simple number.[563] Estimates for the heritability of height differ between populations, ranging from between ~60 to ~80% of the variation attributed to genetic differences and ~20 to ~40% environmental (nutritional) factors. In addition, height (in humans) is a sexually dimorphic trait - on average males are taller than females.

If variation in many genes is involved how do we identify the genes involved in a particular trait?[564] Begin with a trait that can be accurately measured. Height is more objectively measured than friendliness. Then we need to identify the genetic differences found between different individuals. Typically between 500,000 to 1 million single nucleotide polymorphisms (SNPs) are used. A useful SNP occurs at high frequency (>10 to 30%) in the population. The SNP need not be located within a particular gene; with a high enough density of SNPs, a SNP will be linked to almost every gene. Of course meiotic recombination can influence who is linked to whom.

The different SNPs present in a particular genome are identified based on nucleotide sequence. Samples of a person's genome are taken, often from white blood cells that have nuclei. Since alleles and SNPs differ in their nucleotide sequences, two perfectly complementary (single-stranded) DNA molecules bind more strongly to one another than two mis-matched molecules. We can use this difference in binding stability to identify which SNP or allele is present at a particular position. Finally, we ask how the presence of particular SNPs/alleles relates to the trait, for example the height of the person or the levels of low (LDL) and high density (HDL) lipoproteins in their blood. Of course you see some of the issues right away. People are different heights at different times of their lives, and different levels of LDL and HDL depending on their diet, and when they last ate. So the trait we are trying to study has to be accurately and reproducibly measurable.

We then ask which markers (SNPs or alleles) are correlated to the trait phenotype under study. In a large enough population we can identify those markers (alleles and SNPs) that are in or near specific genes associated with the phenotype. Correlation however does not prove causation. It may be that the allele/SNP involves a functionally significant allele. To prove that a particular allele plays a role in producing or modifying a trait, further experimental studies are necessary.[565]

Questions to answer:

238. What is critical to know before one can even consider beginning a GWAS study?

Questions to ponder:

- You discover a gene linked to a particular trait through a GWAS study, how might you go about establishing a significant physiological role for the gene in influencing that trait?

A few conclusions before we move on ...

At this point, you will have completed what is meant to be a two semester introductory course on modern molecular and cell biology. It is limited in scope because what it aims to teach (help you learn) is important to master in order to think about biological systems mechanistically. As noted by Oscar Whitney (per. comm.), who served as a learning assistant for the course (awhile ago), the goal of any such course should be to help you build effective and productive intuitions regarding biological systems. That does not mean memorizing large numbers of facts, but rather developing reasonable feelings for how a system could (plausibly) work. What molecular level processes are likely to be involved.

So what comes next? Typically that might be courses in cell and more advanced molecular biology - looking at common mechanisms regulating the behaviors of biological systems. More and more details, but all anchored in the core concepts introduced here. In the next section we consider how these processes are applied in the context of developing systems. It emerged after MWK started teaching a developmental biology course.



Short Chapter Summary

- Synteny reveals evolutionary history; comparative genomics clarifies function.
- Expression maps (where/when) and variant databases (e.g., gnomAD) connect genotype ↔ phenotype.
- BLAST and friends: homology as a discovery engine.

References

[563] How much of human height is genetic and how much is due to nutrition?

[564] Chapter 11: Genome-Wide Association Studies

[565] The interplay of common, rare variation in autism

This page titled 16.5: Genome-wide Association Studies (GWAS) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

17: Supplemental Chapter - Fundamental concepts and developing systems

17.0: Introduction

17.1: How do systems change at the molecular level?

17.2: Steady state and changing molecular concentrations- synthesis and degradation

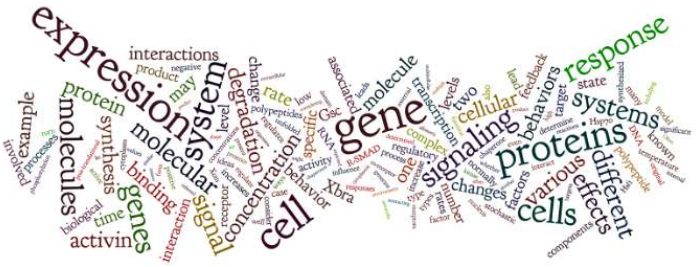
17.3: Direct and indirect cellular responses to signaling molecules

17.4: Modeling gene expression

17.5: Reversible, irreversible, and cascade effects

This page titled 17: Supplemental Chapter - Fundamental concepts and developing systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

17.0: Introduction



In which we consider the basic molecular & cellular processes involved in the behavior of groups of cells, including those involved in the transformation of a single cell, the fertilized egg, into a complex multicellular organism composed of multiple and integrated cell types.

By this point, you have been introduced to many and perhaps most of the core molecular and cellular ideas needed to understand the facts and mechanisms involved in the behavior of biological systems.[566] We will call upon these to build models of specific processes. With modification, these ideas serve as the basic toolkit that working biologists and biology students call upon to design models of a process, and design experiments to test the model's assumptions and either refine or abandon the model and development new ones. The ability to generate plausible (rather than correct) models, and understand the model's predictions is a high level skill that involves recognizing and applying relevant facts and concepts and ignoring irrelevant and distracting ideas.[567] It allows us to simplify and focus our thinking on what is important. Given the complexities of biological systems, which may be beyond what a human brain can readily comprehend, simplification can be necessary.[568] It starts by reflecting on what ideas apply to specific situations; a skill that takes practice and informed feedback - one reason that manuscripts written by the most experienced scientists often benefit from peer review leading to thoughtful response and revision, and sometimes new experiments. Here we consider how fundamental ideas are applied to understanding the behaviors of developing (animal) systems.

By developing systems we mean a cell or group of cells and how they change over time in response to various signals and perturbations, including their interactions with one another. Cells monitor their external and internal environments; their regulatory networks are critical to maintaining the living state (a process known as homeostasis), adapting to changing conditions, as well as decisions on whether to grow and divide, or to "differentiate". These decisions involve interconnected molecular and cellular networks that control all aspects of cellular behavior, including coordination with surrounding cells - together they produce the emergent behaviors we referred to as the living state. Emergent behaviors are those that "cannot be predicted through analysis at any level simpler than that of the system as a whole".[569]

When we think about emergent behaviors, an obvious example is the development of multicellular organisms. In these systems a fertilized egg, formed by the fusion of haploid gametes, goes on to form a multicellular embryo composed of multiple cell types in specific juxtapositions, and controlled by a combination of intracellular and extracellular asymmetries and cellular responses to those asymmetries. Similar processes are found among unicellular organisms. Microbes of various types sense their neighborhood, including the numbers (concentration) of related and unrelated organisms. They can alter their cooperative, competitive, and adaptive behaviors in various ways through quorum sensing, positive and negative feedback interactions, and molecular cascades that produce changes in gene expression, morphology, and behaviors (phenotypes).

References

[566] adapted from the blog post: on teaching developmental biology from a biofundamentalist perspective

[567] Making mechanistic sense: are we teaching students what they need to know?

[568] What (exactly) does it mean to understand the brain (and life in general)

[569] review of Dyson, 1997. Darwin among the Machines: The Evolution of Global Intelligence

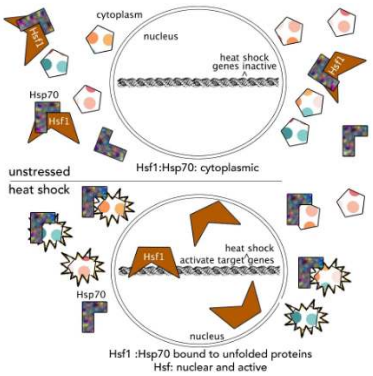
This page titled 17.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

17.1: How do systems change at the molecular level?

The fundamental biological system is the cell. Many essential processes are common and referred to as "housekeeping" functions. At the same time, cells have the ability to respond in different ways to external and internal factors. In multicellular organisms, various factors combine to produce the patterns of cell division and differentiation that underlie the formation of specific cell types, tissues, organs, and organismic behaviors. How a cell responds to external and internal signals is a product of the organism's evolutionary history, its genes, the genes it is expressing, together with the proteins (and other molecules) present, their activities and cellular locations. When cells change, when they take on different shapes, express different genes and gene products, and display different behaviors, these changes occur in response to interactions between internal systems and external factors. Again these are emergent behaviors.

How, exactly, do cell's change their behavior(s)? They can respond to changes in specific signaling or adhesion molecules. Considering the effects of physical changes that directly effect cellular components. Radiation, such as UV light, can provide the energy to initiate chemical reactions. Such processes lead to the tanning of skin, the synthesis of vitamin D, the capture of energy, and the generation of mutations. Starvation, the lack of necessary nutrients, can lead to stress responses associated with the interruption of on-going processes dependent upon coupling to thermodynamically favorable reactions. Without ATP and other molecules involved in coupled reactions, the thermodynamically unfavorable reactions associated with maintaining the living state as well as most metabolic reactions will cease. Reactions can stall, and aberrant molecules can accumulate.

A classic example, the heat shock response, involves changes in temperature. At the temperatures that a cell normally experiences, many of proteins are semi-stable, folding and partially unfolding. A class of evolutionary conserved proteins known as chaperones act to enable unfolded (denatured) proteins to refold. Other proteins act to degrade (recycle) unfolded or abnormally folded proteins for degradation, removing potentially toxic molecules.[570] Because the number of chaperone molecules in a cell is limited, there will be a competition - proteins normally associated with a chaperone may lose that interaction in the presence of increasing numbers of unfolded proteins generated by various stressors. There are evolutionarily conserved cellular responses to stress that increase the expression of genes that encode "heat shock proteins", chaperones and other "defense" factors. The transcription factor Hsf1 is constitutively expressed but normally sequestered in the cytoplasm in an inactive form through interactions with the heat shock protein Hsp70 (→). In response to temperature-induced protein unfolding/misfolding there is an increase in the concentration of denatured, Hsp70 binding proteins that leads to the movement of Hsp70 out of Hsp70:Hsf1 complexes and an increase in "free" Hsf1 that can enter the nucleus where it activates the expression of various genes. The expression of these genes further protects the cell from the potentially toxic effects of unfolded proteins. When the system temperature returns to normal, and unfolded proteins are refolded or degraded; the concentration of Hsp70 increases and sequesters Hsf1 in the cytoplasm. Genes dependent upon Hsf1 for their expression "turn off".



Considering this process, we recognize a number of common themes. First, binding interactions are based on molecular structure and the numbers of molecules present. There will always be a competition between possible "target" molecules for chaperone binding. Different proteins will differ in the stability of their functional state(s), so changing temperature will change the pattern of chaperone binding proteins and will influence the degree to which various chaperone:target complexes exist. This rule applies to transcription and associated factors and the genes and processes they regulate. The combination of binding site affinity and transcription factor concentration and modification will determine the extent to which specific DNA binding sites are occupied, and will influence the extent to which the genes they regulate are expressed. Changes in molecular shape, associated with unfolding, post-translational modifications, interactions with other proteins, or the binding of allosteric effectors can influence molecular behaviors and properties.

Question to consider: What can you expect to happen if cellular chaperones were to disappear?

References

[570] Rosenzweig et al., 2019. The Hsp70 Chaperone Network

This page titled 17.1: How do systems change at the molecular level? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

17.2: Steady state and changing molecular concentrations- synthesis and degradation

A key factor influencing molecular interactions are the free concentrations of the component involved, a function of synthesis and degradation rates, both regulated processes. Polypeptide and protein synthesis reflect rates of mRNA synthesis and processing, introns removal, 5' cap and 3' polyA tail addition, cytoplasmic transport (in eukaryotes), and interactions with ribosomes. The length of the Both transcription and translation are subject to stochastic effects leading to what is known as "bursting"– periods when multiple RNAs or polypeptides are synthesized and periods when few if any are.[571] When the time-averaged levels of a gene product are low, bursting expression can have functionally significant effects on the concentration of a gene product, influencing the behavior of biological systems at the single cell level. Given the cascade effects that we will discuss, a transient increase in a protein, particularly if it influences the pattern of gene expression, can lead to long lasting effects on cellular behaviors – such effects can generate phenotypic variations between cells within a homogenous environment.

A molecule's stability is characterized by its "half-life". As noted, unlike the half-life of a radioactive isotope, the degradation rate of a molecule is not intrinsic to the molecule but determined by active and regulated processes. Polypeptides can contain sequences or post-translational modification, such as covalent addition of ubiquitin (a small 76 amino acid long polypeptide) that can mark them for rapid degradation by proteolytic enzymes. Degradation is stochastic, so that the smaller the population size, the greater the statistical fluctuations - the more noise, the more variation. The effect is similar to that seen in genetic drift (allele behavior in populations) and the case of the bacterial lac operon (discussed previously). When the concentrations of gene regulatory factors are low, such variation will leads to noisy gene expression that can generate significant phenotypic variation between genetically identical cells and their progeny.

References

[571] What is a transcriptional burst? & Beyond initiation-limited translational bursting

This page titled 17.2: Steady state and changing molecular concentrations- synthesis and degradation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

17.3: Direct and indirect cellular responses to signaling molecules

A typical signaling system uses both fast acting responses (allosteric effectors and post-translational modifications, including proteolytic processing) and slower acting changes in gene expression (synthesis and degradation rates). Each signaling system can be characterized by common features that include i) the signal itself - generally molecules synthesized and released by other cells or the cell "talking to itself", a process known as autocrine signaling. ii) The receptor for the signal, generally receptors are proteins synthesized by the responding cell. Finally, iii) the effect(s) that occurs when the signaling molecule binds to the receptor. Many (most?) cellular signaling systems result in changes to molecular networks and patterns of gene expression.[572]

We can model the behaviors of biological systems. Molecular interactions are based on the thermodynamics of surface-surface and surface-solvent interactions. The binding energy determines the stability of the interaction, that is the average time an interaction, once formed, persists before it is knocked apart by collisions with other molecules. Low affinity interactions will likely be transient, persisting (on average) for shorter periods of time than higher affinity interactions. A long persistence time favors interactions with other molecules, and so the assembly of functional complexes. Multi-molecular components involving larger interacting surface areas are expected to be more stable than simpler ones involving smaller areas. Given their stochastic nature, while we can predict the average duration that two interacting molecules remain bound to one another, we cannot predict the behavior of any particular interaction. This matters at the cellular level, since there are only two copies of most genes and often limited numbers of regulatory molecules present. Noise in gene expression associated with low transcription factor levels is to be expected.

Often what were originally thought to be independent molecular interaction networks interact, producing systems of systems that lead to emergent behaviors. Various experimental and genetic manipulations can influence a system in multiple and complex ways. For example, the removal of a gene can be expected (naively) to lead to the "simple" absence of a gene product but the effects can be more complex.[573] If a gene product interacts with other gene products, then the behavior of these interacting gene products may be altered in unexpected ways. As an example, polypeptides "orphaned" by the absence of their normal interaction partner may interact with molecules they would not normally interact with leading to disruptive effects on normal function(s). Alternatively, they may fail to fold normally and so form toxic aggregates. These effects may be modulated by the levels of various chaperones, proteins that can (in some cases) reverse the effects of protein aggregation and misfolding.[574] But chaperone systems have a limited capacity; the effects of a new mutation may impact the number of targets they can "deal with".

Question to consider: How can the position of a mutation in a gene influence molecular interactions? What types of information would help you with your predictions?

References

[572] an example: Cytoskeletal control of gene expression: depolymerization of microtubules activates NF-kappa B

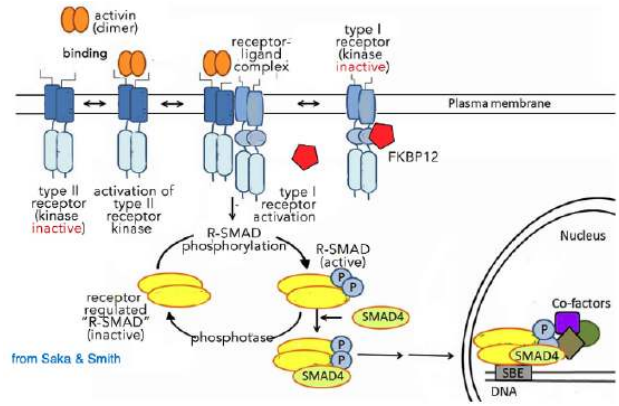
[573] Teng et al (2013). Genome-wide consequences of deleting any single gene

[574] Such behaviors are discussed here: *Filaments & phenotypes: cellular roles and orphan effects associated with mutations in cytoplasmic intermediate filament proteins*.

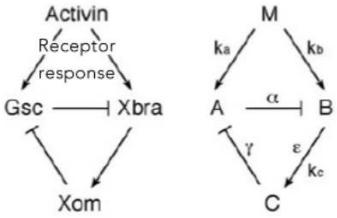
This page titled 17.3: Direct and indirect cellular responses to signaling molecules is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

17.4: Modeling gene expression

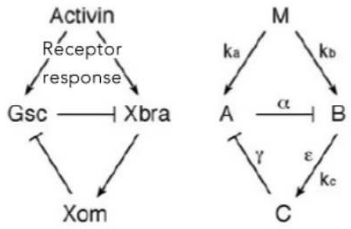
Let us use the model described by Saka & Smith[575] to illustrate a number of points. Their model aimed to understand how an extracellular signaling molecule regulates the mutually exclusive expression of two target genes. They considered cellular responses to the secreted signaling molecule activin, a member of the Transforming Growth Factor (TGF) family of proteins. Activin is synthesized and secreted by cells during embryonic development in *Xenopus laevis* and lots of other multicellular organisms.[576] Certain cells express the genes that encode polypeptides that assemble into cell surface receptors that bind activin.



When activin binds, there is a change in the receptor's three dimensional shape that influences its catalytic activity and/or its interactions with other molecules. The signaling molecule is an allosteric effector of the receptor. (←) The activin receptor has a protein kinase activity. The receptor's activin-binding site is extracellular while its kinase domain is intracellular. On activin binding, the activin:receptor complex binds to a co-receptor, another membrane protein leading to the activation of the receptor's kinase domain. The active kinase phosphorylates the co-receptor leading to changes in co-receptor's structure that i) favor the dissociation of a cytoplasmic inhibitor (FKBP12) from the co-receptor, ii) result in the activation of the co-receptor's protein kinase domain, and iii) the phosphorylation of cytoplasmic receptor-regulated SMAD (R-SMAD) proteins.[577] This phosphorylation results in shape changes so that phospho-R-SMADs dimerize and associate with a "co-SMAD" polypeptide, SMAD4. The SMAD4 polypeptide, normally cytoplasmic contains a transcription-activating domain. The phosho-R-SMAD:SMAD4 complex ("SMAD complex" for short) is transported into the nucleus through nuclear pores. In the nucleus the SMAD complex binds to specific DNA sequences and associated proteins, regulating the expression of target genes. There are a number of different R-SMAD proteins; different combinations of R-SMADs in a SMAD complex activate different sets of target genes. There are also proteins that can interact with a SMAD complex and inhibit its activity, turning it into a transcriptional repressor. At each point along the pathway there are inhibitors that can modulate the effects of extracellular activin: there are activin-binding proteins that block its binding to receptors, proteins that bind to the receptor and block its activation, and cytoplasmic proteins that block R-SMAD phosphorylation. The system is dynamic and, importantly, all of the events associated with activin signaling are reversible - including co-receptor and R-SMAD phosphorylation. R-SMAD dephosphorylation leads to the disassembly of the SMAD complex, the export of SMAD4 from the nucleus, and the inactivation of activin-regulated genes.



In Saka & Smith's model, the level of activin leads to SMAD-regulated expression of two genes, *Gsc* and *Xbra* (→).[578] At this point, what these gene names "mean" and where they come from is not important, what is important is that both genes encode sequence specific DNA binding proteins and act as regulators of transcription. Both *Gsc* and *Xbra*genes are directly regulated by the Activin signaling pathway; there are no intervening genes whose transcription and translation are necessary – the system is poised to respond to activin:activin receptor binding. There are, however, downstream effects based on the ability of the Gsc protein to inhibit *Xbra* expression and the ability of the Xbra protein to induce expression of the *Xom* gene. The product of the *Xom* gene is a transcriptional repressor that inhibits expression of *Gsc*. *Xom* is an indirect (downstream) target of Xbra. Generally, there are a limited number of direct regulatory targets of a signaling system; these act to control the regulatory cascade of downstream targets. In this case, while *Gsc*, *Xbra*, and *Xom* (and the polypeptides that they encode) are the focus of the analysis, it is reasonable to assume that the Gcs, Xbra, and Xom proteins directly regulate, perhaps tens to hundreds or more other genes and proteins - they might positively regulate some, and negatively regulate others, depending upon promoter binding affinities, proteins concentrations, modifications, and cellular contexts.



Analyzing, predicting, and understanding signaling effects is further complicated by multiple feedback interactions. In the activin-system there are three such feedback interactions. The Gsc and Xbra gene products negatively regulate each other's expression, so that at a high enough concentration of Gsc, *Xbra* expression is inhibited, and visa versa, even in the presence of activin-based activation. There is also a secondary, indirect negative feedback interactions mediated by the Xom gene product's effect on *Gsc* expression. Finally there can be negative feedback interactions that involve the degradation of receptors or other components of the signaling system; these act to turn down or turn off signaling after a period of activation, even if the signal is still present. There can be (but not here) positive feedback interactions, in which a gene product further activates the expression of the gene that encodes it. We will consider what limits such positive feedback loops shortly

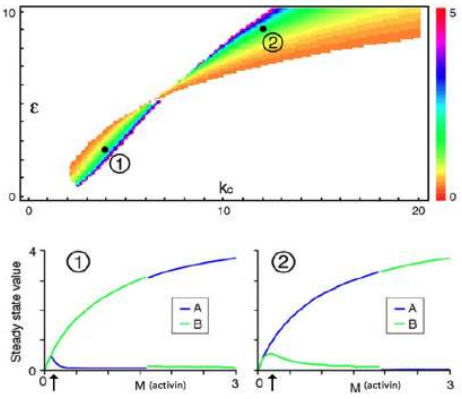
$$\frac{dA}{dt} = \frac{k_a}{1+C^{\gamma}} \cdot \frac{M^{\mu}}{1+M^{\mu}} - kd_a \cdot A$$

$$\frac{dB}{dt} = \frac{k_b}{1+A^{\alpha}} \cdot \frac{M^{\mu}}{1+M^{\mu}} - kd_b \cdot B$$

$$\frac{dC}{dt} = k_c \cdot \frac{B^{\varepsilon}}{1+B^{\varepsilon}} - kd_c \cdot C$$

k_a, k_b and k_c are the synthesis rates of A, B and C. α and γ reflect the cooperativities of repression by A and C, ε and μ are the cooperativities of induction by B and M. kd_a, kd_b and kd_c are degradation rates of A,B, and C proteins.

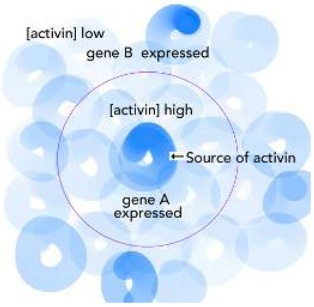
Predicting the behavior of the Activin-Gsc-Xbra system is not intuitively simple, we need to generate a quantitative model. We can abstract and generalize the system, replacing protein and gene names with symbols (←). In such a model, many of the molecular mechanisms involved are "collapsed" into more general variables and used to generate systems of (solvable) differential equations (→). These enable us to make quantitative predictions as to how the system will behave in response to various perturbations. In this case we characterize the relationship between the strength of the original signal (M), the relative effects on the direct (A and B) and indirect (C) genes, the concentrations of various proteins, and their affinities for their regulatory targets. The variables that apply to the system can take on a range of values. Variations reflect the situations in different cells. In addition, while the same activin (M) signaling system directly regulates transcription of the A and B genes, the rate of A and B synthesis can be quite different; for example, differences in the lengths of the RNA molecules and their coding region, as well as RNA and polypeptide degradation rates, folding and assembly rates (in the case of polypeptides that are part of a multimeric complex) will lead to different time delays for the appearance of functionally significant levels of the encoded proteins. The functionally significant level of a particular protein will depend on their binding affinities for (and the accessibility of) their various targets and their roles in generating a functional response.

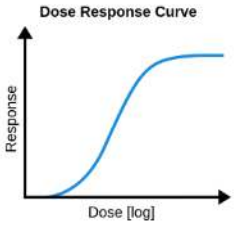


Modified from Saka & Smith: Simulations were performed assuming ka = kb = 5, α = γ = 3, μ = 1, kda = kdb = kdc = 1. Threshold values of M, which are color-coded, are plotted in the parameter plane (kc, ε). kc and ε determine how the value of C changes over time. Consider two sets of variables (top panel). At smaller values of kc and ε (area 1), B is on and A is off at low M, and vice versa with high M (once the system reaches) steady state. With larger values of kc and ε (area 2), A is on and B is off with low M, and vice versa with high M at steady state.

How the system behaves depends on these parameters, which may or may not be easily determined experimentally. Saka and Smith modeled the system's behavior at two parameter positions (marked 1 and 2 in the top graph (↓). In both, behavior is similar at low concentrations of activin (bottom graphs); both A and B genes (*Gsc* and *Xbra*) are expressed at low levels of activin signaling (the "↑s" in the lower panels). Expression behavior changes dramatically as activin concentration increases. In the two domains, expression of one or the other of the target genes increases, while the other drops to near zero. Expression of active genes continues to increase until activin concentration crosses a threshold, at which point the system flips, the expression of the previously expressed gene drops to near zero while the expression of the unexpressed gene jumps to high levels. If we were to think of a plane of cells, in which there is a localized source of activin that decreases with distance from that source, resulting in an activin concentration gradient, we might predict that, assuming that the cells are similar, that we will find a circular domain of cells expressing gene A surrounded by a domain of cells expressing gene B. The two domains would be separated by a distinct boundary (→). The expression of A or B would be expected to lead to different cellular behaviors, different "downstream" effects.

Another type of threshold effect: Often when the level of signal (or a transcription factor) increases, the effects on the target genes it directly regulates is not a linear one – the relationship between signal and response is not accurately described by a straight line. The dose-response relationship is best described by a sigmoidal curve, a smooth curve with a characteristic shape – it looks





like a flattened S (→). There is little or no response to low levels of signal, after which there begins a smooth increase until, at higher signal levels, the response flattens again. When response onset and saturation concentrations are close, the response curve looks more like a step function, basically an off-on (or on-off) switch. There are a number of reasons for why low levels of a signal fail to activate a response, these can involve the need to assemble a stable multicomponent complex before a response can occur. If the synthesis/activation rate of a necessary response component is a function of signal concentration, while the degradation/inactivation rate is constant, sufficient activity may only appear above a certain signal concentration, and then increase more or less linearly after that, essentially after the degradation machinery has been saturated and has reached its maximum rate. Given that there typically only two copies of a particular gene in a diploid cell, the number of RNA molecules that can be synthesized per unit time is limited by the rate at which RNA polymerase molecules can load onto one or the other of these genes. And, of course, as in the case with the gene regulatory system described above, there can be both positive and negative interactions between components, including positive and negative feedback loops, wherein one component effects its own synthesis, activity, or stability (degradation).

References

[575] Saka & Smith 2007. A mechanins for the shapr transition of morophogen gradient interpretation in Xenopus

[576] Activin is a member of the TGFb family of signaling molecules. see Chaikuad & Bullock 2016. Structural Basis of Intracellular TGF-b Signaling: Receptors and Smads

[577] SMAD stand for the a homologous of the "Suppressor of Mothers against Decapentaplegic" protein.

[578] As a reminder, gene names are in italics while the polypeptides encoded for by a gene is in standard font.

This page titled 17.4: Modeling gene expression is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

17.5: Reversible, irreversible, and cascade effects

A final consideration is whether, when a cell receives a signal, its response is transient - that is, does it return to its original state when the signal is removed or does it adapt to the presence of the signal, for example, through a negative feedback interaction that leads to decreased levels of receptor or critical response components, or does it move into a new cellular state, characterized by the expression of different genes and different cellular behaviors. As an example, if the signal up-regulates expression of a transcription factor that in turn regulates expression of other transcription factors that results in altered receptors and regulatory molecules, the cell can become physiologically and phenotypically different – becoming basically a new cell type. It may no longer respond to the original signal. In the first type of response, after a transient signal it "resets" back to its original state. The cell may even adapt (get used to) the presence of one level of signal, and require a higher level to continue to respond. The second type of response can be effectively irreversible, the cell has changed in terms of the genes it expresses, the proteins and molecules it contains. Chromatin organization may be altered, different genes become accessible or inaccessible. The same signaling molecule may produce no response or a different response. This second type of response is common in embryonic development, cells move from an originally totipotent state to an increasingly restricted one. An early embryonic cell may respond to signal(s) by forming into a range of cell types. At a later stage the same signals may not influence a differentiated cell. The differentiated state may be irreversible. A neuron, once formed, remains a neuron - a terminally differentiated cell type.

A technical breakthrough has been the development of protocols that can reverse terminal differentiation in some cell types. These enable us to reprogram a cell, producing what are known as induced pluripotent stem cells or iPS cells. These protocols do not work equally well with all differentiated cell types, which is one reason (among many) that the cells that go on to form gametes, the cells of the germ line, are maintained in a distinctive state compared to the cells that go on to form the body, the somatic cells, which are differentiated to various extents. The process of reprogramming a somatic cell is itself associated with stochastic effects, effects that can be best observed through single cell analyzes of gene expression. When a culture of supposedly identical cells are expose to the factors use to generate iPSCs, analysis of individual cells indicates that most cells fail to "reset", and that those that do can differ in significant ways from one another.[579]

Questions to answer

239. Make (and describe) a model for how a cell adapts to level of signaling molecule, and then required a higher level of signaling molecule to produce the same response.
240. Make (and describe) a model for an irreversible response to a pulse of signaling molecule; what factors will determine the behavior of the system.
241. Make (and describe) a model by would a point source of a signaling molecule could produce patterns (such as the "eye spots" in a butterflies wing.



Short Chapter Summary

- Development is change in state driven by changes in molecular concentrations, rates of synthesis/degradation, and network wiring; you can model much of it with simple production–decay and switch-like logics.
- Signaling reshapes gene expression directly (fast, local) and indirectly (through cascades and feedbacks), producing thresholds, memory, and spatial patterns.
- Reversible steps enable flexibility; irreversible commitments (e.g., bistable switches) lock in fates—cascades amplify small biases into robust outcomes.
- Practical skill: sketch minimal “box-and-arrow” models of a developmental decision (inputs → TFs/chromatin → target genes) and predict how tweaking a rate or affinity shifts the phenotype.

References

[579] see Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming

This page titled 17.5: Reversible, irreversible, and cascade effects is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

18: Supplemental Chapter - Social interactions between cells

18.0: Introduction

18.1: How do unicellular organisms generate phenotypic diversity?

18.2: Dying for others – social interactions between “unicellular” organisms

18.3: Quorum effects (somewhat redundant)

18.4: Transient and clonal ("true") metazoans

18.5: Unicellular to multicellularity to unicellularity

18.6: Evolutionary origins of clonal (permanent) multicellularity

---

This page titled 18: Supplemental Chapter - Social interactions between cells is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

18.0: Introduction



In which we consider the social behaviors displayed by cells within multicellular organisms.

Biology is often presented as a fragmented discipline. It is common to find multiple biology departments on a single college campus. Yet, underneath the diversity of organisms, systems of organisms (micro and macro ecologies), and idiosyncratic molecular mechanisms, there are evolutionary (family) resemblances that go deep. This is the reason we can use studies of dramatically different organisms to reveal the common mechanisms behind similar behaviors. As the result of evolutionary adaptations, different organisms can display behaviors in exaggerated forms, or can be more accessible (convenient and economical, or both) to scientific study. At the same time it is important to remember that a molecular/cellular mechanism characterized in one type of organism may be different, often in subtle, but important ways. Mice are not people, and there are mechanistically important differences between even the most closely related species, as well as between individuals of the same species due to genetic variation and life histories. Related (homologous) molecules may play different roles in different species.[580]

An important feature of many organisms are the social aspects of their behavior. How is it that unicellular organisms can cooperate with one another under specific circumstances to produce useful behaviors that they could not accomplish at the single cell level? Based on quorum sensing and the ability to produce multiple phenotypes from a single genotype, these behaviors range from self-sacrifice to the construction of complex molecular machines and communal feeding strategies. A particularly dramatic example occurs when normally unicellular organisms come together and coordinate their behaviors to form what we might term a temporary metazoan. In addition to self-sacrifice, we see examples of cellular differentiation in response to environmental and internal factors. Similar mechanisms are used in a wide range of responses, including those involved in producing a human from a fertilized egg. Network behavior and integration underlie the emergent behaviors of a range of systems, from the immune system to the brain. Now we consider what we can learn about general processes from studies of specific types of animals (we will largely ignore plants).

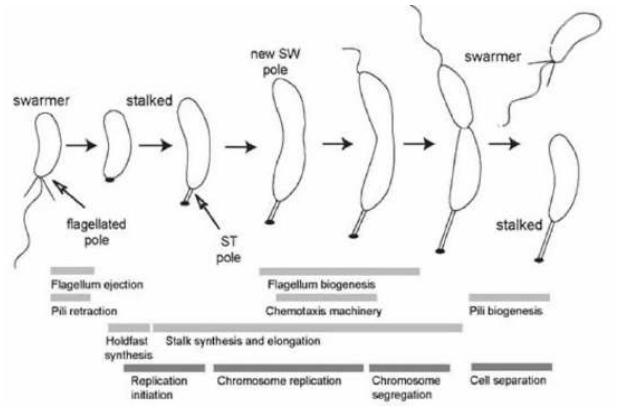
References

[580] an example: Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells

This page titled 18.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

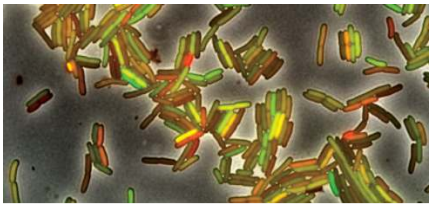
18.1: How do unicellular organisms generate phenotypic diversity?

In most unicellular organisms, the cell division process is reasonably uneventful, the cells produced are similar to the original cell – but not always! A well studied example is the bacterium *Caulobacter crescentus* [→].[581] The process of growth and cell division leads to the production of phenotypically distinct daughters. While it makes no sense to talk about a beginning (given the continuity of life after the appearance of LUCA), we can start with a "swarmer" cell, characterized by the presence of a motile flagellum, a molecular machine driving cellular motility.[582]



A swarmer cell will eventually settle down, lose its flagellum and replace it with a specialized structure, a holdfast, that anchors the cell to a solid substrate. The holdfast develops a stalk that lifts the cell away from the substrate. As growth continues, the end of the cell opposite the holdfast begins to differentiate – it begins the process leading to the assembly of a new flagellar apparatus. When reproduction (cell growth, DNA replication, and cell division) occurs, a swarmer cell is released. It can swim away and colonize another area or settle nearby. The holdfast-anchored cell continues to grow, producing new swarmers. This process is based on the inherent asymmetry of the system – the holdfast end of the cell is molecularly distinct from the flagellar end. This type of behavior is similar to that displayed by what are known as stem cells in multicellular organisms.

The process of swarmer cell formation in *Caulobacter* is an example of what we will term deterministic phenotypic switching. Cells can also exploit molecular level noise (stochastic processes) that influence gene expression to generate phenotypic heterogeneity, different behaviors expressed by genetically identical cells within the same environment. This process enables members of a population to sample phenotypic space.[583] Molecular noise arises from the stochastic nature of molecular movements and the rather small (compared to macroscopic systems) numbers of (most) molecules within a cell.[584] Which molecules are bound to which regulatory sequences, and for how long, is governed by inter-molecular surface interactions and thermally driven collisions, as well as their physical accessibility. How the chromatin is folded and what other proteins may be bound may influence expression. There are strategies that can suppress but not eliminate such noise.[585] As dramatically illustrated by Elowitz et al and others. Molecular level noise can produce cells (↓) with different phenotypes. Similar processes are active in eukaryotes (including humans), and can lead to the expression of one of the two copies of a gene. If the two alleles at a particular locus are not the same, monoallelic expression can lead to phenotypic differences between lineages.[586] Recent studies suggest the presence of competitive interactions between such clones.[587] Such stochastic phenotypic heterogeneity between what are genetically identical cells is rarely considered in most biology courses, but is becoming increasingly easy to identify using techniques such as single cell RNA sequencing and is found in essentially all cellular systems.[588] Control of such variation has been reported based on social / community responses.[589]



The ability to sample different phenotypes can be a valuable trait if an organism's environment is subject to significant changes. As an example, when the environment gets hostile, some bacterial cells transition from a rapidly dividing to a slow or non-dividing state, known as "persisters"; these non- or slow growers are found resistant to antibiotics, other drugs, and inhospitable environments. The result is that some cells in the population will survive until the environment becomes hospitable again.[590] In some cases, cells differentiate to form "spores" that are even more resistant to killing by dehydration, radiation, and other stressors. If environmental changes occur frequently and rapidly, a population can protect itself by continually having some cells (stochastically) differentiating into spores, while others continue to divide rapidly. Only a few individuals need to survive a catastrophic environmental change to re-establish the population once things get better. Some of these responses are transient, re-setting quickly as conditions change, while others may lead to the establishment of new phenotypes.

References

[581] explore more: Caulobacter microbewiki, C. crecentus & Hughes et al 20112. C. crescentus.

[582] from Jacobs-Wagner (2004). Regulatory proteins with a sense of direction: cell cycle signaling network in Caulobacter.

[583] Elowitz et al 2002. *Stochastic gene expression in a single cell.* Science **297** & Balázsi et al., 2011. Cellular decision making and biological noise: from microbes to mammals. Cell **144**

[584] Fedoroff, N. and W. Fontana 2002. *Small numbers of big molecules.* Science **297**:1129-1131.

[585] Lestas et al., 2010. Fundamental limits on the suppression of molecular fluctuations. Nature **467**:174-178.

[586] Zakharova et al., 2009. Monoallelic gene expression in mammals. Chromosoma, **118**:279-290 & Deng et al., 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science. **343**: 193-196.

[587] Ellis et al., 2019. "Distinct modes of cell competition shape mammalian tissue morphogenesis." Nature **569**: 497.

[588] Biology education in the light of single cell/molecule studies

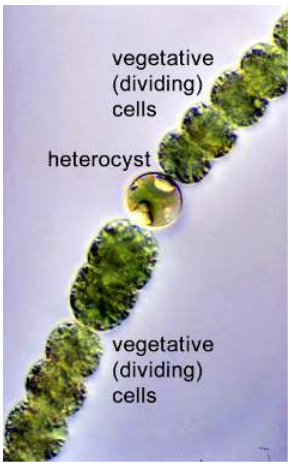
[589] see Cell competition corrects noisy Wnt morphogen gradients to achieve robust patterning in the zebrafish embryo (2019)

[590] Fisher et al., 2017. Persistent bacterial infections and persister cells. Nature Reviews Microbiology **15:**453.

This page titled 18.1: How do unicellular organisms generate phenotypic diversity? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

18.2: Dying for others – social interactions between “unicellular” organisms

Perhaps you wonder, aren't self-sacrificing behaviors are contrary to evolutionary mechanisms? Isn't it surprising to learn that one bacterial cell can sacrifice itself to benefit another. In fact, self-sacrificing behaviors are a common feature of biological systems.



An interesting example is provided by the cellular specialization decisions associated with photosynthesis and nitrogen fixation in cyanobacteria. These two processes require mutually exclusive cellular environments; molecular oxygen (O_2) released by photosynthesis inhibits nitrogen fixation. Nevertheless, both are required for optimal growth. The solution? Some cells differentiate into what are known as heterocysts (→), cells committed to nitrogen fixation, while most "vegetative” cells continue with photosynthesis. Heterocysts cannot divide, and eventually die – they sacrifice themselves for the benefit of their neighbors, the vegetative cells that can reproduce.

The process by which the death of an individual releases resources that can be used by its neighbors to ensure or enhance their survival is an inherently social process and subject to control by social mechanisms.[591] Social behaviors can be selected for because the organism’s neighbors, the beneficiaries of the self-sacrificial behavior, are likely to be closely (clonally) related to themselves. A result of such "self-sacrificial" social behaviors, mediated by “inclusive fitness”, can be an increase in the overall fitness of the population and an increase in the frequency of the genes, alleles, and regulatory networks involved.

Social behaviors can enable a subset of the population to survive environmental stresses. An obvious environmental stress involves the impact of viral infection. Recall that viruses are completely dependent upon the metabolic machinery of the infected cell to replicate. While there are a number of viral reproductive strategies, a common one is bacterial lysis – the virus replicates explosively, kills the infected cell leading to the release of virus into the environment. But, what if the infected cell or the larger organism's immune system kills the infected cell BEFORE the new virus particles are assembled? The result is the death of the virus and the end of the infection.[592] Often genetically programmed cell death responses are based on a simple two-part system, involving a long lived toxin and a short-lived anti-toxin (discussed previously).

References

[591] In an age of rampant narcissism and social cheating – the importance of teaching social evolutionary mechanisms

[592] One can imagine a similar process in the context of COVID-19. If an infected individual self-isolates themselves (a sacrificial behavior for most people) until their immune system eliminates the virus, they effective kill the virus and spare others from infection

This page titled 18.2: Dying for others – social interactions between “unicellular” organisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

18.3: Quorum effects (somewhat redundant)

Some types of behaviors make sense only when the density of organisms rises above a certain critical level. It makes no sense evolutionarily (or practically) for a single *Anabaena* cell to differentiate into a heterocyst (see above). The synthesis and secretion of a specific enzyme, a specific import or export machine, or the construction of a specific, complex and costly machine, such as a DNA uptake machine, may provide no advantage for an isolated organism. The secreted molecule will just diffuse away, and so be ineffective, the molecule to be imported (e.g. lactose) may not be present, there may be no free DNA to import.[593] As the concentration (organisms per volume) of bacteria increases, however, these behaviors become useful – there is DNA to eat or incorporate and the concentration of secreted enzyme can be high enough to degrade the target molecules (so they are inactivated or imported as food).

Of course, as in any social system, such "altruistic" (self-sacrificing and cooperative) behaviors are vulnerable to cheaters. A cheater might avoid programmed cell death (for example due to a mutation that inactivates the cell killing system) and could to take over the population over time. On the other hand, if such cheaters take over, the population will be less likely to survive the types of hostile environmental events that the social (altruistic) behavior evolved to address. In response to the realities of cheating, social organisms have evolved various strategies that enforce the commitment to social cooperation.

Questions to answer:

242. How might social cheaters be recognized by non-cheaters? What other ways might a cheater cheat?
243. Describe a situation in which the ability to produce multiple phenotypes from a single genotype is beneficial.

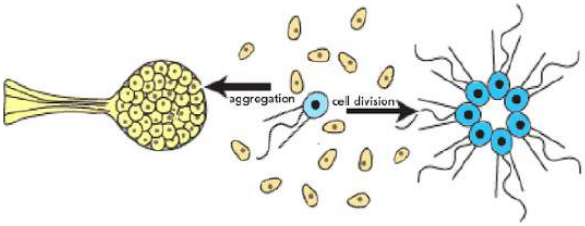
References

[593] page 208

This page titled 18.3: Quorum effects (somewhat redundant) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

18.4: Transient and clonal ("true") metazoans

Although we often think about developmental processes as restricted to multicellular organisms, there are versions that involve organisms that exist in both unicellular and transient multicellular forms. Because they are simpler, we can learn important and relevant lessons from studying transient metazoans.[594] Forming a transient multicellular organism requires that single celled organisms cooperate with one another, they get social.



The ability of individuals to cooperate, through processes such as quorum sensing and community effects, enables them to tune their responses so that they are appropriate and useful. Social interactions also make it possible for them to produce behaviors impossible for isolated individuals. Once individual organisms develop (evolutionarily) the ability to cooperate, new opportunities and challenges (cheaters) emerge. There are strategies that can enable an organism to adapt to a wider range of environments, or to become highly specialized to a specific environment, through the production of increasingly complex behaviors. Many cooperative strategies can be adopted by single celled organisms, but others require a level of multicellularity. Multicellularity can be transient – a pragmatic response to specific conditions, or it can be (if we ignore the short time that gametes exist as single cells) permanent, allowing the organism to develop the range of specialized cell types needed to build large, macroscopic organisms with complex and coordinated behaviors. We can divide multicellularity into two distinct types, aggregative and clonal. These appear to have arisen independently in a number of lineages .[595]

References

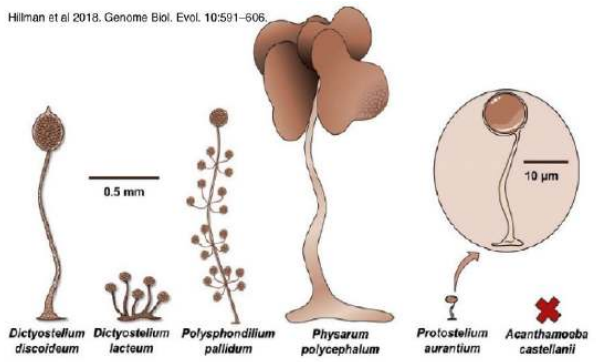
[594] We restrict our considerations to animals. Behavioral systems in multicellular plants (metaphyta) are beyond us.

[595] Bonner. 1998. The origins of multicellularity and Knoll. 2011. The multiple origins of complex multicellularity.

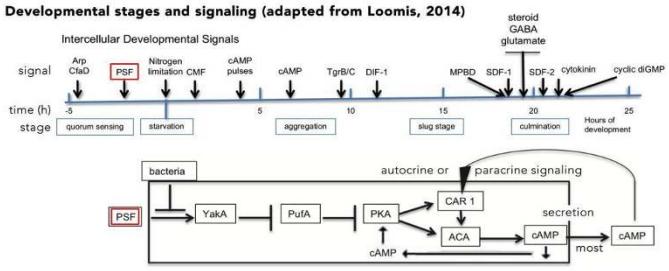
This page titled 18.4: Transient and clonal ("true") metazoans is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

18.5: Unicellular to multicellularity to unicellularity

Quorum and environmental/internal sensing systems enable single celled organisms to monitor the density of related organisms in their environment, as well as the supply of nutrients, and to turn on or off specific sets of genes necessary to produce specific and complex cooperative behaviors. A classic example occurs in cellular slime molds, with *Dicty*ostelium *discoideum* being the most extensively studied (→).[596] Under normal conditions, these unicellular amoeboid eukaryotes migrate, eating bacteria and such. In this state, the range of an individual's movement is restricted to short distances. When conditions turn hostile there is a compelling reason to abandon one environment and migrate to another, a journey impossible for a single-celled organism. Under such "hostile" conditions such signaling systems provoke the directional migration and aggregation of single celled amoeboid-forms to form multicellular aggregates (slugs). Slugs migrate and undergo a process of differentiation into multiple cell types. The formation of a multicellular slug depends upon the presence of sufficient number of cells enabling them to: 1) recognize one another's presence (through quorum sensing), 2) find each other through directed (chemotactic) migration, and 3) form a multicellular slug that goes on to differentiate. During differentiation about 20% form non-dividing stalk cells that organize themselves to the stalk that serves to lift the other ~80% of the cells into the air. These non-stalk cells (the survivors) differentiate to form resistant spores that can be released into the air where they can float away (escape) to new locations and establish new populations.



The process of cellular differentiation in *D. discoideum* has been worked out in molecular detail. It involves two distinct signaling systems: a secreted pre-starvation factor (PSF) protein and cyclic AMP (cAMP)(↓). PSF is a quorum signal leading to the inactivation of PufA and increased PKA activity. Active PKA induces the synthesis of two downstream proteins, adenylate cyclase (ACA) and the cAMP receptor (CAR1). ACA catalyzes cAMP synthesis, much of which is secreted as a signaling molecule. The membrane-bound CAR1 protein acts as a receptor for autocrine (feedback to the secreting cell) and paracrine (acting on neighboring cells) cAMP signaling. The binding of cAMP to CAR1 leads to the activation of PKA further increasing cAMP synthesis and secretion – a positive feed-back loop. As cAMP levels increase, down-stream genes are activated and inhibited leading cells to migrate toward, and to adhere to one another to form a migratory slug. The slug begins to migrate to an appropriate site; the processes of cellular differentiation, morphogenesis, and death lead to stalk and spore formation. The fates of the aggregated cells are determined stochastically. Social cheaters can arise. Mutations can lead to individuals that avoid becoming stalk cells. In the long run, if all individuals became cheaters, it would be impossible to form a stalk, so the survival purpose of social cooperation, forming a structure that disperses spores, fails. In the face of environmental variation, populations invaded by cheaters are more likely to become extinct. The various defenses against cheaters are best left to other, more advanced courses.[597]



Growing amoebae constituatively synthesize and secret the protein PSF; the extracellular concentration of PSF reflects cell density. Above a threshold [PSF] level, PSF activates the protein kinase YakA. YakA activation is inhibited by the presence of bacteria (that is , food). When active, YakA inhibits the inhibitor PufA, which inhibits protein kinase PKA. PKA activity is dependent on cylic AMP (cAMP). Activation of PKA leads to increase expression of adenyla cyclase (ACA) with catalyzes the synthesis of cAMP. Most cAMP is secreted and can bind to cell surface cAMP (CAR1) receptors on the secreting cell (autocrine signaling) or neighboring cells (paracrine signaling). Activated CAR1 activates ACA leading to increased cAMP levels that, in turn, lead to cell migration, slug formation, and differentiation.

References

[596] Loomis. 2014. Cell signaling during development of Dictyostelium; Hillmann et al., 2018. Multiple roots of fruiting body formation in Amoebozoa.

[597] Strassmann et al., 2000. Altruism and social cheating in the social amoeba Dictyostelium discoideum.

This page titled 18.5: Unicellular to multicellularity to unicellularity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

18.6: Evolutionary origins of clonal (permanent) multicellularity

An interesting aspect of the unicellular-multicellular-unicellular behaviors of social slime molds is that evolutionary selection acts on both uni- and multi-cellular stages. A major evolutionary transition, leading to the appearance of permanently multicellular plants, animals, and fungi is estimated to have occurred some time in the Cryogenian period (834–780 Ma).[598] Exactly how this transition occurred, on how many occasions, and exactly why remains unclear - presumably it involved selection for organisms that could exploit a new range of ecological niches.

Aggregative multicellularity involves an extension of quorum sensing and social cooperation between genetically distinct but related individuals. We can speculate on the drivers of true multicellularity, in which all cells of the organism are clonally related and one generation is related to the next through meiosis, gamete formation, and fusion.[599] Two plausible adaptive (evolutionary) drivers of permanent multicellularity spring to mind: as a way to avoid or discourage predators by getting bigger and as a way to produce varied structures needed to more efficiently exploit ecological niches and life styles, opportunities not available or as efficiently exploited by unicellular organisms.

An example of the first type of driver is offered by the studies of Boraas et al; they cultured the unicellular green alga *Chlorella vulgaris*, together with a unicellular predator, the phagotrophic flagellated protist *Ochromonas vallescia*.[600] After less than 100 generations they observed the appearance of multicellular, and presumably inedible (or at least less easily edible), forms of *Chlorella*. Once selected, this trait appeared stable, such that “colonies retained the eight-celled form indefinitely in continuous culture”. At the time of this writing the genetic basis for this multicellularity remains to be determined.

Another relevant evolution of multicellularity experiment are the studies by Ratcliff et al. They selected yeast that failed to separate after mitosis – the resulting clumps of cells fell to the bottom of culture tubes, forming multicellular aggregates known as "snowflakes". Within each snowflake cluster, the cells were linked mother to daughter in chains. A snowflake would "divide" into two when one of the cells in a chain died by programmed cell death (apoptosis).[601] This type of division is quite distinct from the process of sexual reproduction. The evolution of true multicellularity involves the appearance of specialized cells involved in reproduction, the germ line-somatic divide. [602]

Short Chapter Summary

- Even single celled organisms show social behavior: they diversify phenotypes, communicate (quorum effects), cooperate, and sometimes die to benefit kin—cheater control is essential.
- Transient collectives and clonal, permanent multicellularity emerge when coordination and conflict mediation outperform going solo; cycles between uni- and multicellularity are evolutionarily plausible.
- Big picture: social rules at the cellular level (signals, adhesion, policing) scale to tissues and organisms—understand these to understand development, immunity, and cancer.
- Practical skill: identify the payoff/cheater-control mechanisms in any cell collective you study and predict how a perturbation (e.g., reduced signal diffusion) alters group function.

References

[598] Snowball Earth climate dynamics and Cryogenian geology-geobiology and Uncertainty in the Timing of Origin of Animalsand the Limits of Precision in Molecular Timescales

[599] J.T. Bonner 1998. The Origins of Multicellularity

[600] Boraas et al., 1998. Phagotrophy by a flagellate selects for colonial prey: a possible origin of multicellularity see also Herron et al., 2019. De novo origins of multicellularity in response to predation

[601] Ratcliff et al 2015 Origins of multicellular evolvability in snowflake yeast & Pentz et al., 2016. Apoptosis in snowflake yeast: novel trait, or side effect of toxic waste?

[602] A nonadaptive explanation for macroevolutionary patterns in the evolution of complex multicellularity

This page titled 18.6: Evolutionary origins of clonal (permanent) multicellularity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

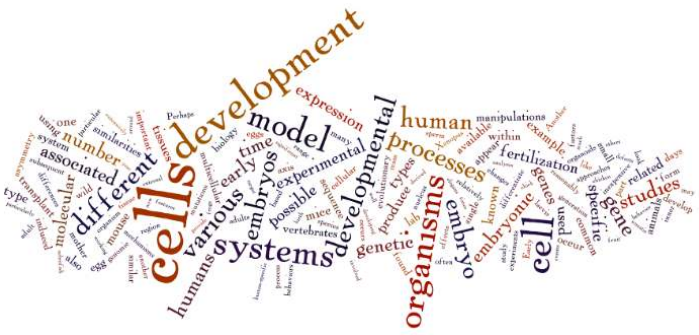
CHAPTER OVERVIEW

19: Supplemental Chapter - The role of model systems in understanding metazoan development

19.0: Introduction
19.1: Model Systems
19.2: Some model systems
19.3: Frogs and fish
19.4: Chick and Quail
19.5: The fruit fly Drosophila melanogaster
19.6: The nematode Caenorhabditus elegans
19.7: The Mouse
19.8: ESC and iPSC derived organoids

This page titled 19: Supplemental Chapter - The role of model systems in understanding metazoan development is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.0: Introduction



In which we consider some common model systems (organisms) use in the study of developmental processes, their strengths and limitations.

The study of developmental biology has its roots in the comparative study of embryos, but embryology as a discipline is beyond the scope of an introductory course in developmental biology, in part because there are 35 (assuming no more are discovered) different "types" (phyla) of animals – nicely illustrated at this website [BBC: 25 types of animals, most of whom are really weird]. What has transformed classical embryology into developmental biology has been the increased availability of genomic sequence data from more and more different types of organisms and a more detailed understanding of cellular and molecular processes. Together these have transformed embryology into evo-devo.[603] But for many biologists the principle driver of studying developing systems is to gain a better (practical and working) understanding of human birth defects and pathogenic processes.

While humans are connected to the rest of the tree of life, and specifically mammals, we are (now) a distinct species, derived from a population that separated from other mammals sometime around 6,000,000 years ago. In response to the various evolutionary pressures and events associated with this speciation event and subsequent human evolution there have been a number of functionally significant molecular changes specific to *Homo sapiens*.[604] Some lead to therapeutically significant differences in the response of humans to treatments that are effective in other organisms, such as the mouse.[605] At the same time, experimentation with humans is constrained by both ethical and practical considerations. These constraints are appropriate and necessary given the depressing history of medical atrocities and malpractice. To circumvent these experimental limitations, at least in the exploratory stages of biomedical research, it is common to turn to model systems. So what do we mean by a model system and what have we learned about development in general, and human development in particular from such studies?

References

[603] see Arthur, W. (2002) The emerging conceptual framework of evolutionary developmental biology and Wilson, E.B. (1940) The cell in development and heredity.

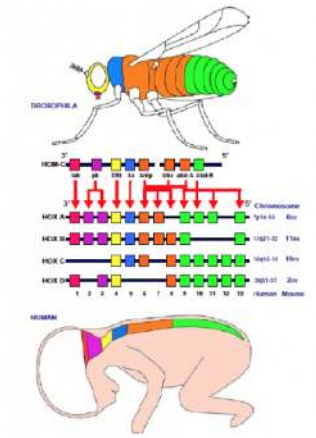
[604] Guo et al 2020. Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells

[605] NYTs 2013. Mice Fall Short as Test Subjects for Some of Humans' Deadly Ills

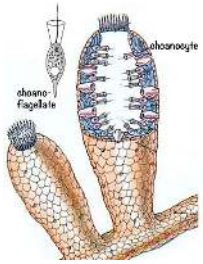
This page titled 19.0: Introduction is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.1: Model Systems

As our focus is on human development, we consider developmental processes in animals (and ignore plants). “All members of Animalia are multicellular, and all are heterotrophs, that is, they rely directly or indirectly on other organisms for their nourishment). Most ingest food and digest it in an internal cavity.”[606] From a macroscopic perspective, most animals have (or had at one time during their development) an axis of asymmetry. This asymmetry may pre-exist within the unfertilized egg or it may appear in response to external factors, such as sperm entry or early events during development. This axis of asymmetry underlies the development of the embryonic axes: anterior-head to posterior-tail (oral-aboral). Animals that can crawl, swim, walk, or fly typically also have a dorsal-ventral (back to belly) and a left-right axis. When seeking model organisms that can be studied profitably in terms of insights into human development, we look for common and practical features. First we need to be able to cultivate the organism in captivity (in the lab). We would prefer organisms that are small and can be fed non-esoteric foods; maintaining individuals and colonies should be reasonably inexpensive. A rapid replication time would be desirable, we would like to get experiments done in a timely manner. At the same time we would like the stages of early development to be experimentally accessible - external fertilization is associated with embryonic development occurs that outside the mother. Processes that occur within the mother are more technically challenging. At the same time, we might want to avoid organisms that display unique behaviors. An example would be the nematode *Ascaris suum*, in which ~13% of the genome is discarded in somatic cells. While this process may be of interest, since it occurs in a human parasite, it is unlikely to provide direct insights into processes associated with human development.[607]



On the other hand, there are deep molecular level similarities between organisms that appear to be completely different. Perhaps the most dramatic is the HOX cluster(s) of genes associated with anterior-posterior and proximal-distal (in limbs) axes specification. These genes encode DNA binding transcription factors and their genomic organization. Their patterns of expression are similar throughout the metazoans, from fruit flies to mice and humans (→).[608] It should be noted, however, that Hox gene organization is often presented in textbooks in a distorted manner (→).[609] The Hox clusters of vertebrates are compact, but they are split, disorganized, and even “atomized” in some types of organisms, an example of what seems to be a conserved feature can, through evolutionary processes, be altered.[610] Such molecular similarities extend to cell-cell and cell-matrix adhesion systems and the systems that release and respond to signaling molecules, controlling cell behavior and gene expression. Similarities reflect the evolutionary conservation and the common ancestry of all animals.[611] Differences reflect adaptions to specific niches and life styles, as well as unexpected events.



Where do these similarities come from? Presumably they were present in the common ancestor of all metazoans. Early in the history of comparative cellular anatomy, it was noted that there are striking structural similarities between the feeding system of choanoflagellate protozoans, a motile (microtubule-based) flagellum surrounded by a “collar” of microfilament-based microvilli, and a structurally similar organelle (←) found in choanocytes, cells present in multicellular organisms, such as sponges. The implication is that the Choanozoan ancestor was predisposed to exploit some of the evolutionary opportunities offered by clonal multicellularity. These pre-existing affordances, together with newly arising genes and proteins were exploited in multiple lineages in the generation of multicellular organisms.[612]

References

[606] Phil Myers. Animals

[607] The Occurrence, Role and Evolution of Chromatin Diminution in Nematodes and Silencing of Germline-Expressed Genes by DNA Elimination in Somatic Cells

[608] Figure from Lappin et al, 2006. HOX genes: seductive science, mysterious mechanisms.

[609] Duboule 2007. The rise and fall of Hox gene clusters.

[610] Similar to the limited repurposing of codons in some organisms (link?)

[611] Brunet & King. 2017. The origin of animal multicellularity and cell differentiation.

[612] Long et al., 2013. New gene evolution: little did we know.

This page titled 19.1: Model Systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.2: Some model systems

Here we briefly consider a number of the most commonly used model organisms, focussing in particular on the types of experimental analyses and developmental processes they are best suited for. While developmental processes have been studied in many organisms, over time scientists have narrowed their attention to just a few. These range throughout the animal kingdom, and generally have been chosen based on practical considerations.[613] Perhaps the most important is the availability of embryos throughout the year; experiments can be carried out as the are imagined and designed by researchers. Since one experiment is often inspired or necessitated by results and observations from the last, it is important not to have to wait until next year to do the follow on experiments. At the same time, the maintenance of organisms in the lab needs to be reasonably inexpensive; this tends to favor smaller organisms that can be housed in compact quarters. Other factors that influence choice of experimental organisms are the ease of their experimental manipulation; such manipulations are easier when eggs and embryos are large, and when fertilization and subsequent development occur outside of the mother. The ease with which organisms survive and heal from surgical manipulations can also be a factor. As we will see, different model systems offer benefits for answering questions about specific processes.

Early on the experimental manipulations available to researchers were limited. Regions of a developing embryo could be moved or removed. Alternatively, one could generate, select, and analyze mutations that influenced developmental processes. More recently a much wider range of molecular interventions have become available. Embryonic cells can be injected with various inert dyes to trace cellular lineages (what types of cells a particular cell in the embryo differentiates into). Molecular tools make it possible to construct plasmids that encode RNAs for wild type or mutant gene products; chimeric polypeptides that contain regions derived from fluorescent proteins can be used to visualize cell lineages and the intracellular localization of encoded polypeptides. DNA-based promotor reporters that reveal where different signaling systems are active. Monoclonal antibodies can be injected into cells, where they may bind to and disrupt intracellular protein function(s). The expression of gene products can be suppressed by reagents that block the translation of mRNAs (morpholinos) or act to destabilize or block the translation of target mRNAs (based on microRNAs). Most recently CRISPR Cas9-based approaches have been developed that can mutate target genes in various ways.

An equally important aspect of experimental studies involves the techniques available to analyze the effects of various manipulations on developmental processes. Early on, analyses were primarily based on microscopy-based examinations, often associated with the preparation of thin sections of the organism or tissue. Such sections can be stained with dyes to reveal various subcellular components, such as the nucleus, the nucleoli, or connective tissues. The tools available for analyzing experimental effects and mutant phenotypes have grown dramatically more sensitive and sophisticated. Microscopy, together with various fluorescent reagents has been extended to three dimensions and higher resolution using whole-mount confocal, light sheet, and two-photon microscopy. Single cells and subcellular organelles and their normal or abnormal morphologies can be characterized. It is possible to dissociate embryos or tissues into single cells and to sequence the mRNAs present providing a read-out of the genes expressed as well as the variation between superficially similar cells. Analogous methods exist (affinity-isolation and mass spectrometry-based proteomics) to examine the polypeptides present in a cell, as well as their interaction partners. The following is meant to be but a short introduction to a few key model systems.

References

[613] Hopwood 2019. Inclusion and exclusion in the history of developmental biology

This page titled 19.2: Some model systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.3: Frogs and fish



As a model system, the frog *Xenopus laevis* has a number of advantages, and some limitations.[614] Adults are remarkably disease resistant with a wholly aquatic lifestyle. Its lifecycle (from fertilization to sexual maturity) is relatively short, and that of the related species *X. tropicalis* is even shorter. Xenopus can be induced to lay eggs year round through the injection of commercially available hormones and produce functional sperm. Fertilization and subsequent development occur externally and rapidly, resulting in swimming tadpoles within a day or so. A single female produces a large number (hundreds) of eggs of a size that make injection of individual blastomeres (up the the 16-32 cell stage) and microsurgical manipulations possible with limited training.

Xenopus and other frogs have been particularly useful in identifying and in some cases resolving key questions about developmental mechanisms. Studies in frog embryos identified the "organizer", a region of the early embryo that acts to induce the formation of the embryonic anterior-posterior axis. Nuclear transplant experiments were used to reveal that genetic information is (generally) not lost during vertebrate development, an observation that laid the groundwork for somatic cell reprogramming, the generation of induced pluripotent stem (iPS) cells. Nuclear transplant experiments were facilitated by the identification of a dominant mutation in the gene encoding rRNA (0-nu); heterozygous 0-nu cells have a single nucleolus (the site of ribosomal gene expression), whereas wild type cells have two. Transplanting a one-nucleoli nucleus into a wild type cell enabled experimenters to confirm that the transplanted nucleus was driving development. Finally, because early development is supported by maternal components, isolated cells continue to grow and behave. The surrounding vitelline membrane / fertilization envelop of the early embryo can be removed, making microsurgical approaches, often using eyebrow hairs as scalpels, straightforward with a little practice and dexterity. Various types of embryonic explants have been used extensively to study cellular behaviors, morphogenic movements, and inductive interactions that drive developmental processes.

A type of analysis that is rare in *Xenopus* are genetic studies. While there are experimental approaches to the manipulation of gene expression, these are one off, involving the manipulation of a single embryo. In part this is because the generation time of *X. laevis* is much longer than that of most organisms used for genetic studies, and in part because *X. laevis* is effectively tetraploid. There has been some interest in genetic studies using the related species *X. tropicalis*, which reaches sexual maturity faster and is diploid.

Another vertebrate that has been used extensively for genetic studies is the zebrafish, *Danio rerio*. As with frogs, fertilization and embryonic development are external. Unlike frogs eggs and early embryos, which are pigmented and opaque, zebrafish embryos are nearly transparent, so high resolution optical microscopy is possible. Zebrafish are easy and (relatively) inexpensive to maintain in the lab, which facilitates classical mutagenesis and analysis. The advent of genome sequence data and directed (CRISPR-Cas9 mediated) mutagenesis makes genetic / molecular studies increasingly efficient. It is now reasonably straightforward to "knock-in" various alleles, for example alleles associated with diseases in humans, and examine the effects of related processes in the fish. There are companies that will edit fish genomes in various ways for you![615]

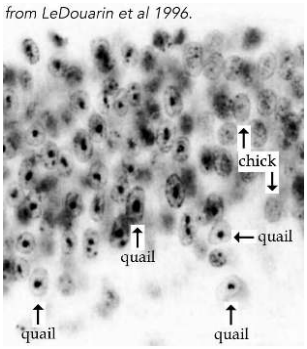
References

[614] Gurdon & Hopwood. 2000. The introduction of Xenopus laevis into developmental biology: of empire, pregnancy testing and ribosomal genes

[615] Zebrafish Genome-Editing Services

This page titled 19.3: Frogs and fish is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.4: Chick and Quail



Another classic system in which to vertebrate development has been studied extensively is the chick embryo. Fertilization occurs internally; the egg is laid after ~24 hours and hatches ~21 days later. It is possible to open the egg without disturbing embryonic development, which allows for tissue removal (extirpation) and transplantation studies. Fertilized chicken eggs are relatively inexpensive and the tools involved are fairly standard.[616] Another important factor is that it is possible to transplant tissues between quail (*Coturnix coturnix japonica*) and chicken (*Gallus gallus*) embryos. While these birds and their eggs are of different sizes, their developmental rates are similar. Importantly, cells from transplant and host can be distinguished based on chromatin organization: in both embryonic and adult quail cells heterochromatin is condensed into a small number (1 to 3) of aggregates in the central region of the nucleus. These can be visualized using histological staining (Feulgen-Rossenbeck) reaction (→). In more modern studies, it has been possible to transplant chick cells transgenic for GFP expression; such chimeric embryos often develop normally and may hatch.

References

[616] Le Douarin et al. 1996 Quail-Chick Transplantations

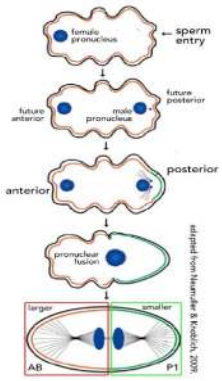
This page titled 19.4: Chick and Quail is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.5: The fruit fly Drosophila melanogaster

The fruit fly has many of the features we look for in a model organism, it is easy and economical to maintain in the lab. Mating (fertilization is internal) produces many offspring, resulting in the laying of embryos that develop quickly, produce motile larvae that undergo metamorphosis to produce sexually mature adults in 10-12 days. Adults can be anesthetized and easily sorted under a dissecting microscope while virgin flies can be distinguished, making controlled crosses of phenotypically and genotypically characterized males and females possible for genetic analyses. A number of chromosomal rearrangements are available to control for recombination effects, and recombination does not occur in males. The characterization of genetic mutations influencing early, and highly stereotyped events in embryonic development, as well as the identification of what are known as homeotic mutations, in which one body part or region is transformed into another, set the stage for the application of molecular techniques that revealed the distribution of gene products, their binding partners, and in the case of transcription factors, the genes they regulate, and defined many of the basic mechanisms underlying embryonic development, such as, the establishment of molecular gradients, and the responses of cells to such gradients.

This page titled 19.5: The fruit fly Drosophila melanogaster is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.6: The nematode Caenorhabditus elegans



Another primarily genetic organism, at least originally, is the soil nematode *C. elegans*, in partbecause most adults are self-fertilizing hermaphrodites.[617] Attractive aspects of *C. elegans* are that it is easy to grow in the lab, and embryos and adults can be frozen and revived![618] It is small (adults are ~1 mm in length). The embryo (and adult) are, like zebrafish, transparent. Its life cycle is about 3 days from fertilized egg to sexually mature adult. The embryo hatches to produce the first larval stage with 558 nuclei (some cells are multinucleate). The cell divisions that produce these cells occur in an invariant pattern, based on an early asymmetry within the egg and the site of sperm entry (→). The pattern of cell division and differentiation enables one to identify and study cells that undergo programmed cell death (apoptosis), and to look at how mutations change patterns of cell division and differentiation. Another aspect has been centered around the ability of double stranded. RNAs to silence target gene expression for multiple generations, a phenomena known as RNA interference (RNAi) and a form of transgenerational epigenetic gene regulation.[619] Studies of RNAi have elucidated the molecular mechanisms involved in related processes associated with small RNAs.

References

[617] C. elegans outside the Petri dish

[618] Corsi et al. 2015. A Transparent window into biology: A primer on Caenorhabditis elegans

[619] Spraklin et al., 2017. The RNAi Inheritance Machinery of Caenorhabditis elegans

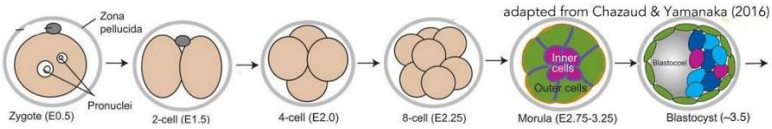
This page titled 19.6: The nematode Caenorhabditus elegans is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.7: The Mouse

For studies of development in mammals, in which both fertilization and subsequent embryonic development occur within the mother, the mouse has been the model system of choice.[620] While the costs associated with working with mice are significantly higher than for the other model systems considered so far, they remain reasonable (much lower, for example, compared to working with pigs or primates), and provide experimental access, particularly through the generation of various genetically manipulated mouse lines, to carry out quite sophisticated studies. Perhaps the technique that had the most dramatic impact has been the Cre-Lox (and related) systems in which genetic manipulations (gene deletions and such) can be activated in specific cell types and at specific times during embryonic development. These have now been extended using CRISPR-Cas9-based systems.

So why the mouse, *Mus musculus*? Mice and humans shared a common ancestor ~80 million years ago and while different share a number of physiological similarities. Pet mice have been kept for centuries, and most lab strains are derived from such mice; they are relatively docile and, to be sure, different from wild (as opposed to wild type, i.e. non-mutant) mice. Mice have a gestation period of 19–20 days (from fertilization to birth), reach sexual maturity in 6 to 8 weeks after birth, and produce litters of 5–8 offspring. At the same time, humans are, on average ~roughly 2500 times larger than mice.

In contrast to the other model systems introduced so far, the mouse (mammalian) egg appears grossly symmetric; sperm entry itself does not appear to impose any long lasting asymmetries. As the zygote divides, the first cells formed appear to be similar to one another. As cell division continues, however, some cells find themselves on the surface while others are located within the interior of the forming ball of cells, or morula (↓). These two cell populations are exposed to different environments, particularly when the embryo implants into the wall of the uterus. The surface cells differentiate to form the trophectoderm, which in turn differentiates into extra-embryonic placental tissues, the interface between mother and developing embryo. The internal cells become the inner cell mass, which goes on to form the embryo proper, the future mouse (or human). Early on inner cell mass cells appear similar to one another, but they experience different environments, leading to emerging asymmetries associated with the activation of different signaling systems, the expression of different sets of genes, and differences in behavior – they begin the process of differentiating into distinct cell lineages and cell types forming, as embryogenesis continues, different tissues and organs. It is possible to establish "embryonic stem cell" (ES) lines from inner cell mass cells. Such cells retain the totipotency displayed by inner cell mass cells, they can differentiate to form essentially any cell type found in the adult.



References

[620] Perlman 2016. Mouse models of human disease: An evolutionary perspective

This page titled 19.7: The Mouse is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

19.8: ESC and iPSC derived organoids

Model systems have provided a wide range of insights into the processes involved in development. That said, it is immediately obvious that there are important differences – after all people are instantly distinguishable from members of closely related species and certainly look and behave differently from mice. For example, the surface layer of our brains are extensively folded (they are known as gyrencephalic) while the brain of a mouse is smooth and referred to as lissencephalic. The failure of the human brain cortex to fold is known as lissencephaly, a disorder associated with several severe neurological defects.[621] With the advent of more and more genomic sequence data, we can identify human specific molecular (genomic) differences. Many of these sequence differences occur in regions of our DNA that regulate when and where specific genes are expressed. Sholtis & Noonan provide an example: the HACNS1 locus is an 81 basepair region that is highly conserved in various vertebrates from birds to chimpanzees; there are 13 human specific changes in this sequence that appear to alter its activity, leading to human-specific changes in the expression of nearby genes.[622] At this point ~1000 genetic elements that are different in humans compared to other vertebrates have been identified and more are likely to emerge.[623] Such human-specific changes can make modeling human-specific behaviors, at the cellular, tissue, organ, and organismic level, in non-human model systems problematic. It is for this reason that scientists have attempted to generate better human specific systems.

The Nobel prize winning work of Kazutoshi Takahashi and Shinya Yamanaka, who devised methods to take differentiated (somatic) human cells and reprogram them into ESC/PSC-like cells, cells known as induced pluripotent stem cells (iPSCs), represented a technical breakthrough that jump-started this field.[624] Since then progress as been rapid. In particular, Madeline Lancaster, Jurgen Knōblich, Yoshiki Sasai, and a growing community of others have devised approaches by which such cells can be induced to form tissue specific organoids. Cerebral organoids, which produce brain-like tissues, have been used to examine developmental defects associated with microencephaly and Zika-virus infection-induced microencephaly, lissencephaly, Down's syndrome and others. Both ES and iPS cells can be induced to differentiate into what are known as gastruloids. Gastruloids can develop anterior-posterior (head-tail), dorsal-ventral (back-belly), and left-right axes analogous to those found in human embryos.[625] Perhaps surprisingly (and perhaps not) human organoids develop along a time-line to that observed in intact human embryos, which means that these studies can take significant amounts of time.[626].

An ethical / moral / religious issue associated with this type of work is that, as gastruloids, assembloids, and embryoids become increasingly like natural human embryos, how they are treated becomes contentious. A similar issues involves the "excess" embryos generated in the course of in vitro fertilization protocols. These are discussion best considered in the context of non-science course.



Short Chapter Summary

- No model does everything; each highlights specific principles: external developers (frog/fish) for early patterning and imaging, chick/quail for graft/lineage experiments, flies and worms for genetic logic and circuit motifs, mouse for mammalian genetics/physiology, and organoids for human-relevant, tractable mini-tissues.
- Strategy beats memorization: pick the system whose life history, genetics, and tools best match your causal question, then map findings across species by homology.
- Practical skill: justify a model choice in one sentence ("We use X because Y lets us measure/perturb Z"), and sketch how results generalize (or don't) to humans.

References

[621] lissencephaly

[622] Sholtis & Noonan. 2010. Gene regulation and the origins of human biological uniqueness

[623] McLean et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits

[624] Takahashi & Yamanaka 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors and How iPS cells changed the world

[625] Turner et al 2017. Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids

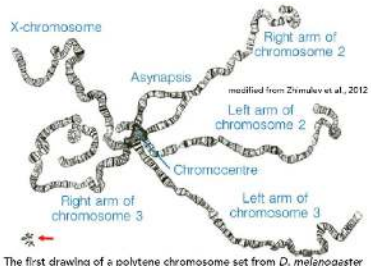
[626] Complete human day 14 post-implantation embryo models from naive ES cells and Gastruloids: Pluripotent stem cell models of mammalian gastrulation and embryo engineering

This page titled 19.8: ESC and iPSC derived organoids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

CHAPTER OVERVIEW

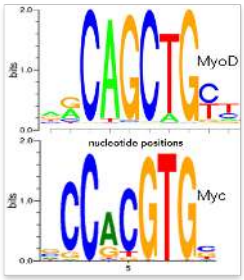
20: Appendix- Muller's Morphs

Another way to look at alleles is from a functional perspective. This was the approach taken by Herman J. Muller (1890-1967) in the 1920s and 30s. He exploited unique features of the fruit fly Drosophila. Geneticists had isolated a number of chromosomal duplications and deletions, something made possible by unique aspects of chromosome organization in the salivary glands of the fly (←). These cells are polyploid; each chromosome contains more than 1000 double-stranded DNA molecules lined up from end to end.[627] Based on the analysis of various mutations he could place mutations into distinct functional (with respect to a particular phenotype) groups: they were either amorphic, hypomorphic, hypermorphic, antimorphic, or neomorphic compared to the wild type ("normal") version of the gene. Note that a particular gene/gene product may have more than one functional role, and a particular mutation may influence its different functions differently. An allele could be hypomorphic for one trait and antimorphic for another. Here a note, all types of alleles can lead to complex, adaptive changes to the cellular system that can lead to phenotypic effects.



The first drawing of a polytene chromosome set from *D. melanogaster* made under light microscopy by T. Painter in 1934. The chromosomes have been spread by squashing them on a microscope slide. Each parental chromosome is tightly paired with its homolog, and there are regions where two homologous chromosomes are separated (asynapsis). All the chromosomes are linked together by the pericentromeric regionto create a single "chromosome center"/\. In the lower left corner, mitotic chromosomes (indicated by arrows) from ovarian tissue are shown at the same magnification, demonstrating the giant size and transveral banding in polyploid polytene chromosomes.

Neomorphic mutations (alleles) are particularly interesting. Such mutations (alleles) change the activity of the gene product, producing a new (neo-) function. As an example a mutation can change the specificity of an enzyme, something that can occur in the development of a cancer.[628] As an illustration, consider the transcription factor MyoD, a protein that regulates the formation (differentiation) of skeletal muscle cells. There are mutant alleles of *MyoD* associated with an aggressive form of embryonal rhabdomyosarcoma, a cancer of skeletal muscle. One missense mutant allele changes the leucine present at position 122 of the wild type MyoD protein to an arginine.[629] This change alters the DNA binding site preference of the MyoD protein. The wild type MyoD protein binds to a particular consensus sequence (top panel →); in contrast the consensus binding sequence for the mutant protein is altered (bottom panel →). The mutant MyoD protein's binding preference now closely resembles the sequences bound by the transcription factor Myc. Myc regulates genes associated with active cell division. The result is that a gene product that normally inhibits cell division and encourages the formation of non-dividing muscle cells (MyoD), now turns on a different set of genes inducing (aberrant) cell division – a key feature of cancer cells. The mutation is neomorphic because the mutated MyoD protein (known as MyoDAla$_{122}$→Arg) has a new function.[630]



The relationship between the type of mutation (in Muller's terminology) and recessivity or dominance is not simple. An amorphic allele can be dominant, a behavior known as halpoinsufficiency, if a single wild type copy of a gene may not produce a sufficient amount of gene product. If recessive, a single functional copy of the gene is sufficient to produce a wild type phenotype. There are mutations that do not change the amino acid sequence of the encoded polypeptide, but change the DNA sequence – these are known as synonymous mutations. Such mutations result in what are known as single nucleotide polymorphisms (SNPs), a feature in the DNA that can be detected by various molecular methods. SNPs are often used in the analysis of genomic similarities and differences, including human ancestry.

Finally, remember essentially all traits are dependent upon a number of gene products, and so are polygenic, while a particular gene product can influence a number of traits - it can be pleiotrophic.[631]

Questions to answer

244. Draw out the relationship between gene→RNA→polypeptide→protein, and describe the effects of mis-sense, non-sense, frame-shift, and intron-exon junction mutations on gene expression.
245. Can you produce some "rules of thumb" relating the position of a mutation within a gene to their effects on the gene product's function?
246. Why is the MyoD mutation neomorphic? What would you call it, if the mutated MyoD protein blocked the binding of wild type MyoD to its target DNA sequences but failed to activate transcription?

Questions to ponder

- A Drosophila polytene chromosome can have over 1000 DNA molecules (strands). How, do you imagine, does the banding pattern observed in these polytene chromosomes relate to the genes on the chromosome?
- How does the polyploid nature of these chromosomes make visualizing chromosomal duplications and deletions possible? What are its limits, do you think?

References

[627] Banding patterns in Drosophila melanogaster polytene chromosomes correlate with DNA-binding protein occupancy.

[628] Neomorphic mutations create therapeutic challenges in cancer

[629] from Myc and MyoD and Deep Sequencing of MYC DNA-Binding Sites in Burkitt Lymphoma

[630] We will return to this topic toward the end of book: see Neomorphic mutations create therapeutic challenges in cancer

[631] Pleiotropy: One Gene Can Affect Multiple Traits

This page titled 20: Appendix- Muller's Morphs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Michael W. Klymkowsky and Melanie M. Cooper.

Index

A

ABO blood type
- 15.13: Alleles, traits, and genetic diseases in humans.

accessible education
- InfoPage

Acetabularia acetabulum
- 7.2: Locating hereditary material within the cell

activation energy
- 5.3: Reaction rates
- 8.9: Regulating protein activity, concentrations, and stability (half-life)

activin
- 17.4: Modeling gene expression

adaptation
- 2.0: Introduction
- 3.3: Fossils and family relationships- introducing cladistics
- 3.6: Darwin and Wallace's conceptual leap
- 3.18: Speciation and extinction
- 3.19: Mechanisms of speciation
- 3.20: Mechanisms of reproductive isolation
- 4.11: Curbing runaway selection
- 13.2: Luria and Delbrück- Discovering the origin of mutations
- 15.13: Alleles, traits, and genetic diseases in humans.
- 18.3: Quorum effects (somewhat redundant)

adaptive behavior
- 9: Genomes, Genes and Regulatory networks

adaptive evolution
- 3.15: Genetic drift

adaptive traits
- 3.17: A brief reflection on the complexity of phenotypic traits

adenosine triphosphate
- 6.5: Generating gradients- using coupled reactions and pumps

adult lactose tolerance
- 16.1: Synteny examined using Genomicus

aerobic
- 9.4: E. coli as a model system

aerobic bacteria
- 6.0: Introduction

aggregation
- 18.6: Evolutionary origins of clonal (permanent) multicellularity

agonists
- 10.4: Signaling molecules and receptors

Albert Einstein
- 5.1: Approaching Thermodynamics

allele
- 3.8: The origins of polymorphisms
- 3.16: Gene linkage- one more complication
- 7.12: Genes and alleles
- 13.1: Mutations into alleles
- 15.16: Genetic anticipation

allele frequencies
- 3.13: Genetic diversity, population size, founder effects and population bottlenecks
- 13.0: Introduction
- 15.15: Measuring evolution's impact on allele frequencies- Hardy-Weinberg

allele frequency
- 12.6: Linkage and haplotypes
- 16.3: Using web-based bioinformatic tools- gnomAD

alleles
- 3.9: Genotype-phenotype relationships- discrete and continuous traits
- 3.10: Variation, selection, and speciation
- 4.8: Origins and implications of sexual reproduction
- 7.0: Introduction
- 7.14: DNA repeat diseases and genetic anticipation
- 13: How Alleles Arise - Mutations
- 13.0: Introduction
- 13.5: Longer term mutation and evolution studies
- 14.0: Introduction
- 15: Mendel and Weldon - Contexts and their Effects on Phenotypes
- 15.1: How Mendel did what he did
- 15.2: Weldon's critique
- 15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
- 15.4: Dihybrid crosses- linkage and recombination
- 15.5: Genetic complementation
- 15.8: Maternal and paternal effects
- 15.11: Estimating the number of genes involved in a particular traits
- 15.13: Alleles, traits, and genetic diseases in humans.
- 20: Appendix- Muller's Morphs

allopatric speciation
- 3.19: Mechanisms of speciation
- 3.21: Sympatric speciation

allopatric_speciation
- 3.20: Mechanisms of reproductive isolation

allosteric binding site
- 8.10: Allosteric and post-translational regulation

allosteric effector
- 8.10: Allosteric and post-translational regulation

allosteric effectors
- 17.3: Direct and indirect cellular responses to signaling molecules

allosteric factors
- 9.6: Types of regulatory interactions

allosteric regulation
- 8.10: Allosteric and post-translational regulation

altruism
- 4.1: Selecting social (cooperative) traits
- 4.4: Inclusive fitness, kin and group selection, and social evolution
- 4.5: Group selection
- 18.3: Quorum effects (somewhat redundant)

altruistic behavior
- 4: Social Evolution and Sexual Selection

amino acid
- 8.4: Protein synthesis- translation (RNA to polypeptide)
- 8.6: Effects of point mutations on polypeptides and proteins

amino acid sequences
- 14.5: Orthologs and paralogs

amino acids
- 2.7: Experimental studies on the origins of life
- 8.0: Introduction
- 8.1: Specifying a polypeptide's sequence
- 8.2: The origin of the genetic code
- 8.5: The translation (polypeptide synthesis) cycle
- 8.8: Turning polypeptides into proteins
- 18.5: Unicellular to multicellularity to unicellularity

amphipathic
- 5.10: Interacting with Water

anabolism
- 5.4: Coupling Reactions

anaerobic
- 9.4: E. coli as a model system

analogous traits
- 3.22: Signs of evolution- homology and convergence

ancestral relationships
- 3.2: Natural and un-natural groups

animal breeding
- 3.4: Evolution theory's core concepts

animal cells
- 6.12: An evolutionary scenario for the origin of eukaryotic cells

animal development
- 19.1: Model Systems

animals
- 12.4: Steps in meiosis- from diploid to haploid

antagonists
- 10.4: Signaling molecules and receptors

anterior posterior
- 19.1: Model Systems

antibiotic resistance
- 11.0: Introduction
- 11.1: What counts as sex in prokaryotes
- 11.2: Other naturally occurring horizontal gene transfer mechanisms
- 13.5: Longer term mutation and evolution studies

antibiotics
- 6.11: Osmosis and living with and without a cell wall

antiporter
- 6.10: Using the energy stored in membrane gradients

apoptosis
- 4.3: Active (altruistic) cell death
- 4.6: Defense against social cheaters
- 19.6: The nematode Caenorhabditus elegans

aquaporins
- 6.11: Osmosis and living with and without a cell wall

archaea
- 2.10: Life's impact on earth
- 6.12: An evolutionary scenario for the origin of eukaryotic cells
- 6.13: Making a complete eukaryote

artificial selection
- 3.4: Evolution theory's core concepts
- 3.10: Variation, selection, and speciation

asexual reproduction
- 11.0: Introduction
- 12.0: Introduction

astrophysics
- 2.6: Thinking about life's origins

atomic organization
- 5.5: Inter- and intra-molecular interactions

atoms
- 5.5: Inter- and intra-molecular interactions

ATP
- 6.10: Using the energy stored in membrane gradients

ATP synthase
- 6.5: Generating gradients- using coupled reactions and pumps
- 6.7: Chemo-osmosis (an overview)

ATP synthesis
- 6.7: Chemo-osmosis (an overview)

attractive forces
- 5.5: Inter- and intra-molecular interactions

autism
- 7.14: DNA repeat diseases and genetic anticipation

autocrine signaling
10.3: Intercellular signaling- signals, receptors and responses
17.3: Direct and indirect cellular responses to signaling molecules
autosomes
12.5: Recombination and independent segregation

B

bacteria
2.5: The Death of Vitalism
2.10: Life's impact on earth
6.12: An evolutionary scenario for the origin of eukaryotic cells
6.13: Making a complete eukaryote
11.1: What counts as sex in prokaryotes
13.5: Longer term mutation and evolution studies
bacterial behavior
18.2: Dying for others – social interactions between “unicellular” organisms
bacterial cells
18.1: How do unicellular organisms generate phenotypic diversity?
bacterial chromosome
11.0: Introduction
bacterial gene activity
4.2: Quorum Sensing
bacterial transformation
11.2: Other naturally occurring horizontal gene transfer mechanisms
bacteriorhodopsin
6.6: Simple Phototrophs
baker's yeast
9.6: Types of regulatory interactions
Barr body
12.7: X-inactivation and sex-linked traits
base pairing
7.5: Discovering the structure of DNA
7.9: Replication machines
behavior
3.0: Introduction
behavioral ecology
4.10: Sexual Selection
behaviors
3.24: Anti-evolution arguments
beneficial mutations
13.5: Longer term mutation and evolution studies
big bang
2.8: Mapping the history of life on earth
bilayer
6.1: Defining the Cell’s Boundary
bilayers
6.2: The origin of biological membranes
binding energy
17.3: Direct and indirect cellular responses to signaling molecules
biochemical pathways
8.5: The translation (polypeptide synthesis) cycle
biochemical reactions
5.4: Coupling Reactions
biochemistry
8.12: Molecular machines
biodiversity
3.0: Introduction
3.1: Organizing organisms (hierarchically)
3.3: Fossils and family relationships- introducing cladistics
3.19: Mechanisms of speciation
4.11: Curbing runaway selection
biofilm formation
4.2: Quorum Sensing
Biofundamentals
5: Molecular Interactions, Thermodynamics and Reaction Coupling
8: Peptide Bonds, Polypeptides and Proteins
bioinformatic tools
16: Tools for Studying Genes and Genomes
biological
6: Membrane Boundaries and Capturing Energy
biological behaviors
5.0: Introduction
biological communities
4.5: Group selection
10: Social Systems
biological diversity
3.21: Sympatric speciation
biological evolution
3.6: Darwin and Wallace's conceptual leap
biological experiments
7.2: Locating hereditary material within the cell
biological history
5.0: Introduction
biological processes
8.0: Introduction
biological reproduction
4.9: Sexual dimorphism
biological system
17.1: How do systems change at the molecular level?
biological systems
1.1: The interconnectedness (self-consistency) of science
1.3: Science is Social
1.5: Understanding Scientific Ideas
2.4: Spontaneous generation and the origin of life
5.0: Introduction
5.2: Coupling thermodynamically favorable and unfavorable reactions
5.3: Reaction rates
5.7: Bond Stability and Thermal Motion (a non-biological moment)
6.0: Introduction
6.7: Chemo-osmosis (an overview)
9.3: Interaction networks and model systems
15.0: Introduction
16.0: Introduction
17.0: Introduction
biological traits
15.7: Interacting traits- epistasis
biology
1.4: Teaching and Learning Science
2: Life’s Diversity and Origins
2.0: Introduction
2.1: What is life, exactly?
2.6: Thinking about life’s origins
4.0: Introduction
8.12: Molecular machines
18.0: Introduction
biomedical research
19.0: Introduction
biosynthesis
6.8: Oxygenic Photosynthesis
biotechnology
16.0: Introduction
birth rate
3.5: Limits on populations
bistable switches
17.5: Reversible, irreversible, and cascade effects
blast
16: Tools for Studying Genes and Genomes
blood type
3.9: Genotype-phenotype relationships- discrete and continuous traits
boiling points
5.9: The implications of bond polarity
bond energies
5.6: Covalent bonds
bond stability
5.6: Covalent bonds
BRCA1
14.1: Rates and effects of somatic mutation
Brownian motion
3.12: Considering stochastic processes
5.1: Approaching Thermodynamics

C

C. elegans
19.6: The nematode Caenorhabditus elegans
calico cats
12.8: X-linked diseases and mono-allelic gene expression
cancer
2.3: The organization of organisms
4.6: Defense against social cheaters
cancer genetics
14.1: Rates and effects of somatic mutation
cancer risk
14.2: Non-disjunction- aberrant chromosome segregation
carbohydrates
9.5: Adaptive behavior and gene networks- the lac response
carriers
6: Membrane Boundaries and Capturing Energy
6.5: Generating gradients- using coupled reactions and pumps
Cas9
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
catabolism
5.4: Coupling Reactions
catalysis
5.3: Reaction rates
catalysts
5.3: Reaction rates
5.4: Coupling Reactions
catalytic proteins
8.9: Regulating protein activity, concentrations, and stability (half-life)
cell
6: Membrane Boundaries and Capturing Energy
7.1: Discovering how nucleic acids store genetic information
17.1: How do systems change at the molecular level?
cell behavior
10: Social Systems

cell biology
3.12: Considering stochastic processes
6.3: Transport across membranes
6.13: Making a complete eukaryote
7.2: Locating hereditary material within the cell
10.2: Nuclear targeting and exclusion
cell cycle
12: Asexual and Sexual Reproduction in Eukaryotes
12.0: Introduction
12.1: Ploidy during the cell cycle
12.2: Monitoring cellular processes- mitosis
14.2: Non-disjunction- aberrant chromosome segregation
cell cycle regulation
12.2: Monitoring cellular processes- mitosis
cell death
4.3: Active (altruistic) cell death
cell differentiation
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
cell division
2.2: The cell theory and the continuity of life
4.6: Defense against social cheaters
4.7: The appearance of multicellular organisms
12.0: Introduction
12.2: Monitoring cellular processes- mitosis
12.5: Recombination and independent segregation
17.1: How do systems change at the molecular level?
18.1: How do unicellular organisms generate phenotypic diversity?
cell membrane
6.11: Osmosis and living with and without a cell wall
cell signaling
10.0: Introduction
10.3: Intercellular signaling- signals, receptors and responses
17.5: Reversible, irreversible, and cascade effects
cell survival
4.6: Defense against social cheaters
cell theory
2: Life’s Diversity and Origins
2.0: Introduction
2.2: The cell theory and the continuity of life
cells
2: Life’s Diversity and Origins
2.1: What is life, exactly?
2.2: The cell theory and the continuity of life
2.3: The organization of organisms
cellular adaptation
17.5: Reversible, irreversible, and cascade effects
cellular behavior
10.3: Intercellular signaling- signals, receptors and responses
17.0: Introduction
cellular behaviors
14.0: Introduction
15.7: Interacting traits- epistasis
17.2: Steady state and changing molecular concentrations- synthesis and degradation
cellular biology
15.9: Mitochondrial inheritance
16.0: Introduction
cellular defense
4.3: Active (altruistic) cell death
cellular differentiation
4.0: Introduction
18.0: Introduction
18.5: Unicellular to multicellularity to unicellularity
cellular integrity
5.0: Introduction
cellular interactions
10.0: Introduction
cellular mechanisms
4.3: Active (altruistic) cell death
cellular membranes
6.4: Channels and carriers
6.10: Using the energy stored in membrane gradients
cellular noise
9.7: Final thoughts on (molecular) noise
cellular organization
2.10: Life's impact on earth
cellular phenotype
10.4: Signaling molecules and receptors
cellular processes
17.0: Introduction
cellular reproduction
11.0: Introduction
cellular response
4.2: Quorum Sensing
cellular slime molds
18.5: Unicellular to multicellularity to unicellularity
cellular specialization
4.7: The appearance of multicellular organisms
centimorgans
12.6: Linkage and haplotypes
channels
6: Membrane Boundaries and Capturing Energy
6.5: Generating gradients- using coupled reactions and pumps
chaperones
8.11: Diseases of folding and misfolding
charge polarization
5.8: Bond polarity, inter- and intramolecular interactions
cheaters
4.6: Defense against social cheaters
18.3: Quorum effects (somewhat redundant)
checkpoints
12.2: Monitoring cellular processes- mitosis
chemical reactions
1.5: Understanding Scientific Ideas
2.4: Spontaneous generation and the origin of life
2.7: Experimental studies on the origins of life
5.0: Introduction
6.3: Transport across membranes
chemical stability
7.3: Identifying DNA as the genetic material
chemiosmotic hypothesis
6.7: Chemo-osmosis (an overview)
chemistry
1.0: Introduction
1.1: The interconnectedness (self-consistency) of science
2.6: Thinking about life’s origins
chemotactic migration
18.5: Unicellular to multicellularity to unicellularity
chemotaxis
4.2: Quorum Sensing
chemotrophs
6.7: Chemo-osmosis (an overview)
6.9: Chemotrophs
chi square analysis
15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
chick embryo
19.4: Chick and Quail
chlorophyll
6.6: Simple Phototrophs
6.8: Oxygenic Photosynthesis
chloroplasts
6.13: Making a complete eukaryote
chromatin
9.0: Introduction
chromatin organization
17.5: Reversible, irreversible, and cascade effects
19.4: Chick and Quail
chromosomal aneuploidy
14.2: Non-disjunction- aberrant chromosome segregation
Chromosomal Basis
12: Asexual and Sexual Reproduction in Eukaryotes
chromosomal duplications
20: Appendix- Muller’s Morphs
chromosomal rearrangements
19.5: The fruit fly Drosophila melanogaster
chromosome
3.16: Gene linkage- one more complication
12.1: Ploidy during the cell cycle
12.2: Monitoring cellular processes- mitosis
12.6: Linkage and haplotypes
chromosome mutations
11.0: Introduction
chromosome segregation
12.0: Introduction
14.2: Non-disjunction- aberrant chromosome segregation
chromosomes
7.1: Discovering how nucleic acids store genetic information
7.10: Further replication complexities
12.5: Recombination and independent segregation
cladistics
3: Evolutionary Mechanisms and the Diversity of Life
classification
3.1: Organizing organisms (hierarchically)
3.2: Natural and un-natural groups
3.3: Fossils and family relationships- introducing cladistics
climate change
3.5: Limits on populations
climate changes
3.14: Population bottlenecks
clonal metazoans
18: Supplemental Chapter - Social interactions between cells
cloning
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
closed systems
5.1: Approaching Thermodynamics
CO2
6.9: Chemotrophs
coding region
15.12: On the nature of mutations (again)
coding regions
7.13: The origin of new (de novo) genes
codon
8.4: Protein synthesis- translation (RNA to polypeptide)

codons
8.1: Specifying a polypeptide's sequence
8.2: The origin of the genetic code
8.5: The translation (polypeptide synthesis) cycle
collaborative learning
InfoPage
colloid
5.10: Interacting with Water
combination of alleles
16.5: Genome-wide Association Studies (GWAS)
common ancestor
3.23: Homologies provide evidence for a common ancestor
3.24: Anti-evolution arguments
8.2: The origin of the genetic code
community behaviors
4.2: Quorum Sensing
comparative genomics
16.1: Synteny examined using Genomicus
comparative studies
3.2: Natural and un-natural groups
competition
3.5: Limits on populations
4.5: Group selection
competitive inhibitor
8.10: Allosteric and post-translational regulation
competitive interactions
18.1: How do unicellular organisms generate phenotypic diversity?
complexity
1.1: The interconnectedness (self-consistency) of science
concentration gradient
6.5: Generating gradients- using coupled reactions and pumps
6.10: Using the energy stored in membrane gradients
concentration gradients
6.4: Channels and carriers
confocal microscopy
19.2: Some model systems
conjugation
11.1: What counts as sex in prokaryotes
11.2: Other naturally occurring horizontal gene transfer mechanisms
consensus in science
1.5: Understanding Scientific Ideas
continuity of life
2: Life's Diversity and Origins
convergent evolution
3.22: Signs of evolution- homology and convergence
cooperation
4: Social Evolution and Sexual Selection
4.0: Introduction
4.1: Selecting social (cooperative) traits
4.7: The appearance of multicellular organisms
4.10: Sexual Selection
18.3: Quorum effects (somewhat redundant)
18.4: Transient and clonal ("true") metazoans
cooperative behavior
4.4: Inclusive fitness, kin and group selection, and social evolution
4.5: Group selection
cooperative behaviors
4.2: Quorum Sensing
4.11: Curbing runaway selection
copy number variations
14.3: Genome dynamics
cosmology
2.6: Thinking about life's origins
2.8: Mapping the history of life on earth
coupled reactions
6.0: Introduction
coupled transport
6.10: Using the energy stored in membrane gradients
covalent bond
5.8: Bond polarity, inter- and intramolecular interactions
covalent bonds
5.5: Inter- and intra-molecular interactions
5.6: Covalent bonds
5.7: Bond Stability and Thermal Motion (a non-biological moment)
CRISPR
13: How Alleles Arise - Mutations
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
19.2: Some model systems
critical thinking
1.0: Introduction
1.2: Models, Hypotheses, and Theories
1.5: Understanding Scientific Ideas
crossing over
12.6: Linkage and haplotypes
14.4: Gene duplications and deletions
cyanobacteria
18.2: Dying for others – social interactions between "unicellular" organisms
cytokinesis
12.2: Monitoring cellular processes- mitosis
cytology
7.1: Discovering how nucleic acids store genetic information
cytoplasm
2.10: Life's impact on earth
6.13: Making a complete eukaryote
10.1: Targeting membrane proteins to where they need to be

D

Darwin
3.3: Fossils and family relationships- introducing cladistics
3.6: Darwin and Wallace's conceptual leap
4.4: Inclusive fitness, kin and group selection, and social evolution
13.2: Luria and Delbrück- Discovering the origin of mutations
de novo genes
7.13: The origin of new (de novo) genes
deep time
2.8: Mapping the history of life on earth
degradation rates
17.2: Steady state and changing molecular concentrations- synthesis and degradation
Delbrück
13: How Alleles Arise - Mutations
deleterious alleles
3.17: A brief reflection on the complexity of phenotypic traits
denatured
8.11: Diseases of folding and misfolding
deoxyribose
7.4: Unraveling Nucleic Acid Structure
development
17.4: Modeling gene expression
18.4: Transient and clonal ("true") metazoans
19.8: ESC and iPSC derived organoids
developmental biology
17.0: Introduction
19.0: Introduction
19.3: Frogs and fish
19.7: The Mouse
developmental processes
19.1: Model Systems
19.2: Some model systems
Dictyostelium discoideum
18.5: Unicellular to multicellularity to unicellularity
differentiation
4.7: The appearance of multicellular organisms
17.1: How do systems change at the molecular level?
17.5: Reversible, irreversible, and cascade effects
diffusion
6.3: Transport across membranes
dihybrid cross
15.4: Dihybrid crosses- linkage and recombination
dihybrid crosses
15: Mendel and Weldon - Contexts and their Effects on Phenotypes
15.5: Genetic complementation
diploid
4.8: Origins and implications of sexual reproduction
12: Asexual and Sexual Reproduction in Eukaryotes
12.1: Ploidy during the cell cycle
12.4: Steps in meiosis- from diploid to haploid
12.5: Recombination and independent segregation
13.1: Mutations into alleles
diploid organism
12.0: Introduction
dipole
5.8: Bond polarity, inter- and intramolecular interactions
directed selection
3.11: Types of simple selection
disaccharide
9.5: Adaptive behavior and gene networks- the lac response
disease
12.8: X-linked diseases and mono-allelic gene expression
14.0: Introduction
16.3: Using web-based bioinformatic tools- gnomAD
disease resistance
3.14: Population bottlenecks
disruptive selection
3.11: Types of simple selection
3.21: Sympatric speciation
dissociation
5.7: Bond Stability and Thermal Motion (a non-biological moment)
diversification
3.14: Population bottlenecks
diversity
2: Life's Diversity and Origins
3.0: Introduction
diversity of life
2.0: Introduction
3.24: Anti-evolution arguments
diversity of organisms
2.4: Spontaneous generation and the origin of life
3.18: Speciation and extinction

dizygotic twins
15.14: Concordance between monozygotic twins and genetic influence on a trait
DNA
2.1: What is life, exactly?
2.2: The cell theory and the continuity of life
3.7: Mutations and the origins of genotype-based variation
3.23: Homologies provide evidence for a common ancestor
7: The Molecular Nature of Heredity
7.0: Introduction
7.1: Discovering how nucleic acids store genetic information
7.4: Unraveling Nucleic Acid Structure
7.5: Discovering the structure of DNA
7.6: DNA, sequences, and information
7.7: Discovering RNA- structure and some functions
7.8: DNA Replication
7.10: Further replication complexities
7.11: Mutations, Deletions, Duplications and Repair
7.12: Genes and alleles
7.14: DNA repeat diseases and genetic anticipation
8.1: Specifying a polypeptide's sequence
8.3: Protein synthesis- transcription (DNA to RNA)
8.7: Getting more complex- gene regulation in eukaryotes
9: Genomes, Genes and Regulatory networks
9.0: Introduction
9.1: Locating information within DNA
13.0: Introduction
DNA dynamics
14.6: Transposons- moving DNA within a genome (and weird genetics)
DNA error repair
14.1: Rates and effects of somatic mutation
DNA Polymerase
7.8: DNA Replication
7.9: Replication machines
DNA replication
3.8: The origins of polymorphisms
7: The Molecular Nature of Heredity
7.9: Replication machines
10.2: Nuclear targeting and exclusion
11.0: Introduction
12.2: Monitoring cellular processes- mitosis
12.5: Recombination and independent segregation
14.1: Rates and effects of somatic mutation
DNA synthesis
12.1: Ploidy during the cell cycle
DNA transfer
11.1: What counts as sex in prokaryotes
DNA uptake
11.2: Other naturally occurring horizontal gene transfer mechanisms
domestication
3.4: Evolution theory's core concepts
dominance
15.13: Alleles, traits, and genetic diseases in humans.
dominant
15.1: How Mendel did what he did
15.2: Weldon's critique
15.4: Dihybrid crosses- linkage and recombination
15.11: Estimating the number of genes involved in a particular traits
dominant alleles
15.6: Interacting traits- synthetic lethality and co-dominance
16.3: Using web-based bioinformatic tools- gnomAD
dominant phenotype
13.1: Mutations into alleles
dorsal ventral
19.1: Model Systems
dosage compensation
12.7: X-inactivation and sex-linked traits
double stranded DNA
7.6: DNA, sequences, and information
Down Syndrome
14.2: Non-disjunction- aberrant chromosome segregation
Drew Berry
8.12: Molecular machines
Drosophila
15.5: Genetic complementation
20: Appendix- Muller's Morphs
Drosophila melanogaster
13.3: Forward and reverse genetics

E

E. coli
9: Genomes, Genes and Regulatory networks
9.4: E. coli as a model system
13.5: Longer term mutation and evolution studies
early Earth atmosphere
2.7: Experimental studies on the origins of life
Earth
1.4: Teaching and Learning Science
earth formation
2.8: Mapping the history of life on earth
earth history
2.9: Fossil evidence for the history of life on earth
ecological advantage
6.8: Oxygenic Photosynthesis
ecological interactions
4.7: The appearance of multicellular organisms
4.11: Curbing runaway selection
ecological niche
3.18: Speciation and extinction
4.0: Introduction
ecological niches
3.19: Mechanisms of speciation
3.21: Sympatric speciation
18.6: Evolutionary origins of clonal (permanent) multicellularity
ecological_niche
3.20: Mechanisms of reproductive isolation
ecology
9.4: E. coli as a model system
ecosystem interactions
10: Social Systems
ecosystems
9.3: Interaction networks and model systems
education
InfoPage
1: Understanding Science and Thinking Scientifically
8.12: Molecular machines
edwin hubble
2.8: Mapping the history of life on earth
electrical field
5.8: Bond polarity, inter- and intramolecular interactions
electrochemical gradients
6.7: Chemo-osmosis (an overview)
electromagnetic spectrum
1.0: Introduction
electron sharing
5.8: Bond polarity, inter- and intramolecular interactions
electron transport
6.5: Generating gradients- using coupled reactions and pumps
6.8: Oxygenic Photosynthesis
electron transport chain
6.7: Chemo-osmosis (an overview)
electronegativity
5.8: Bond polarity, inter- and intramolecular interactions
electrostatic interactions
5.10: Interacting with Water
embryogenesis
19.7: The Mouse
embryology
19.0: Introduction
19.6: The nematode Caenorhabditus elegans
embryonic axes
19.1: Model Systems
embryonic defense strategies
15.10: Imprinting- conflicts between mother, father, and fetus
embryonic development
7.12: Genes and alleles
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
12.7: X-inactivation and sex-linked traits
13.3: Forward and reverse genetics
15.8: Maternal and paternal effects
15.14: Concordance between monozygotic twins and genetic influence on a trait
17.5: Reversible, irreversible, and cascade effects
19.3: Frogs and fish
19.4: Chick and Quail
19.5: The fruit fly Drosophila melanogaster
19.7: The Mouse
embryonic manipulation
19.2: Some model systems
embryonic stem cells
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
emergent adaptations
9.6: Types of regulatory interactions
empathy
4.1: Selecting social (cooperative) traits
empirical evidence
1.2: Models, Hypotheses, and Theories
empirical observations
1.5: Understanding Scientific Ideas
endocrine signaling
10.3: Intercellular signaling- signals, receptors and responses
endocytosis
6.1: Defining the Cell's Boundary
6.3: Transport across membranes
endosymbiosis
2.10: Life's impact on earth
6.0: Introduction
energy
2.1: What is life, exactly?
5.5: Inter- and intra-molecular interactions
6.5: Generating gradients- using coupled reactions and pumps
6.6: Simple Phototrophs
6.10: Using the energy stored in membrane gradients
energy capture
6.0: Introduction

energy storage
6.0: Introduction
energy transfer
2.0: Introduction
5.1: Approaching Thermodynamics
5.2: Coupling thermodynamically favorable and unfavorable reactions
5.4: Coupling Reactions
enhancer element
16.1: Synteny examined using Genomicus
enhancers
9: Genomes, Genes and Regulatory networks
9.2: Enhancers and transcription start and stop sites
enthalpy
5.2: Coupling thermodynamically favorable and unfavorable reactions
entropy
5.2: Coupling thermodynamically favorable and unfavorable reactions
environmental changes
3.14: Population bottlenecks
18.1: How do unicellular organisms generate phenotypic diversity?
environmental damage
13.0: Introduction
environmental factors
3.5: Limits on populations
3.6: Darwin and Wallace's conceptual leap
13.2: Luria and Delbrück- Discovering the origin of mutations
16.5: Genome-wide Association Studies (GWAS)
environmental influence
15.2: Weldon's critique
environmental pressures
3.13: Genetic diversity, population size, founder effects and population bottlenecks
environmental sensing
18.5: Unicellular to multicellularity to unicellularity
environmental stress
4.3: Active (altruistic) cell death
18.2: Dying for others – social interactions between “unicellular” organisms
environmental variation
15.0: Introduction
enzyme
6.8: Oxygenic Photosynthesis
8.10: Allosteric and post-translational regulation
enzyme activity
7.8: DNA Replication
enzymes
5.3: Reaction rates
5.10: Interacting with Water
8.9: Regulating protein activity, concentrations, and stability (half-life)
10.4: Signaling molecules and receptors
epigenetics
9.0: Introduction
14.6: Transposons- moving DNA within a genome (and weird genetics)
equilibrium
5.1: Approaching Thermodynamics
5.2: Coupling thermodynamically favorable and unfavorable reactions
5.4: Coupling Reactions
5.7: Bond Stability and Thermal Motion (a non-biological moment)
equilibrium constant
5.3: Reaction rates
Erwin Chargaff
7.5: Discovering the structure of DNA
Escherichia coli
7.9: Replication machines
ethical considerations
19.0: Introduction
ethical issues
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
eukaryotes
2.2: The cell theory and the continuity of life
2.10: Life's impact on earth
6.8: Oxygenic Photosynthesis
6.11: Osmosis and living with and without a cell wall
6.13: Making a complete eukaryote
7.6: DNA, sequences, and information
7.10: Further replication complexities
8.3: Protein synthesis- transcription (DNA to RNA)
8.7: Getting more complex- gene regulation in eukaryotes
9.2: Enhancers and transcription start and stop sites
10.0: Introduction
12.3: Sex-determination and its chromosomal basis
eukaryotic
7.1: Discovering how nucleic acids store genetic information
eukaryotic lineages
4.7: The appearance of multicellular organisms
eukaryotic organisms
6.12: An evolutionary scenario for the origin of eukaryotic cells
eusocial organisms
4.0: Introduction
evidence
1.3: Science is Social
evidence evaluation
1.5: Understanding Scientific Ideas
evolution
2.0: Introduction
2.1: What is life, exactly?
2.3: The organization of organisms
3.0: Introduction
3.1: Organizing organisms (hierarchically)
3.2: Natural and un-natural groups
3.3: Fossils and family relationships- introducing cladistics
3.4: Evolution theory’s core concepts
3.10: Variation, selection, and speciation
3.16: Gene linkage- one more complication
3.18: Speciation and extinction
3.19: Mechanisms of speciation
3.20: Mechanisms of reproductive isolation
3.22: Signs of evolution- homology and convergence
3.23: Homologies provide evidence for a common ancestor
3.24: Anti-evolution arguments
4: Social Evolution and Sexual Selection
4.0: Introduction
4.1: Selecting social (cooperative) traits
4.4: Inclusive fitness, kin and group selection, and social evolution
4.11: Curbing runaway selection
7.0: Introduction
7.12: Genes and alleles
7.13: The origin of new (de novo) genes
8.0: Introduction
8.2: The origin of the genetic code
12.4: Steps in meiosis- from diploid to haploid
13: How Alleles Arise - Mutations
13.0: Introduction
13.2: Luria and Delbrück- Discovering the origin of mutations
13.5: Longer term mutation and evolution studies
15.13: Alleles, traits, and genetic diseases in humans.
18.6: Evolutionary origins of clonal (permanent) multicellularity
19.1: Model Systems
evolution theory
3: Evolutionary Mechanisms and the Diversity of Life
evolutionary adaptation
3.21: Sympatric speciation
18.4: Transient and clonal ("true") metazoans
evolutionary adaptations
18.0: Introduction
evolutionary biology
2.2: The cell theory and the continuity of life
3.15: Genetic drift
3.22: Signs of evolution- homology and convergence
5.0: Introduction
14.5: Orthologs and paralogs
15.0: Introduction
18.2: Dying for others – social interactions between “unicellular” organisms
18.3: Quorum effects (somewhat redundant)
19.8: ESC and iPSC derived organoids
evolutionary change
16.1: Synteny examined using Genomicus
evolutionary diversity
14.6: Transposons- moving DNA within a genome (and weird genetics)
evolutionary economics
4.9: Sexual dimorphism
evolutionary genetics
16.1: Synteny examined using Genomicus
evolutionary history
2.9: Fossil evidence for the history of life on earth
16.5: Genome-wide Association Studies (GWAS)
evolutionary mechanisms
3.12: Considering stochastic processes
3.19: Mechanisms of speciation
8.2: The origin of the genetic code

evolutionary origins
18: Supplemental Chapter - Social interactions between cells
evolutionary pathways
14.4: Gene duplications and deletions
evolutionary pressures
15.10: Imprinting- conflicts between mother, father, and fetus
15.15: Measuring evolution's impact on allele frequencies- Hardy-Weinberg
19.0: Introduction
evolutionary processes
1.1: The interconnectedness (self-consistency) of science
3.13: Genetic diversity, population size, founder effects and population bottlenecks
3.17: A brief reflection on the complexity of phenotypic traits
9.3: Interaction networks and model systems
evolutionary psychology
4.4: Inclusive fitness, kin and group selection, and social evolution
4.10: Sexual Selection
evolutionary relationships
7.2: Locating hereditary material within the cell
evolutionary significance
14.4: Gene duplications and deletions
evolutionary strategies
4.5: Group selection
evolutionary tools
16.0: Introduction
exocytosis
6.1: Defining the Cell's Boundary
6.3: Transport across membranes
expansion
1.4: Teaching and Learning Science
expansion diseases
15.16: Genetic anticipation
experimental analyses
19.2: Some model systems
experimental design
17.0: Introduction
experimental studies
2: Life's Diversity and Origins
experimentation
1.0: Introduction
extinct species
3.3: Fossils and family relationships- introducing cladistics
extinction
3.0: Introduction
3.5: Limits on populations
3.14: Population bottlenecks
3.18: Speciation and extinction
3.19: Mechanisms of speciation
extracellular
17.4: Modeling gene expression
extracellular signaling
10: Social Systems
extrinsic policing
4.6: Defense against social cheaters
eye color
13.3: Forward and reverse genetics
eye shape
13.3: Forward and reverse genetics

F

F factor
11.1: What counts as sex in prokaryotes
F plasmid
11.1: What counts as sex in prokaryotes
family relationships
3: Evolutionary Mechanisms and the Diversity of Life
fatty acids
6.2: The origin of biological membranes
feedback
17.4: Modeling gene expression
feedback loops
9.3: Interaction networks and model systems
fertilization
12.4: Steps in meiosis- from diploid to haploid
15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
19.4: Chick and Quail
fertilized chicken eggs
19.4: Chick and Quail
fertilized egg
15.8: Maternal and paternal effects
first organisms
6.12: An evolutionary scenario for the origin of eukaryotic cells
fluidity
6.1: Defining the Cell's Boundary
FMR1 gene
7.14: DNA repeat diseases and genetic anticipation
folding
8.8: Turning polypeptides into proteins
forward genetics
13.3: Forward and reverse genetics
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
fossil evidence
2: Life's Diversity and Origins
3.21: Sympatric speciation
fossil record
2.0: Introduction
2.9: Fossil evidence for the history of life on earth
3.3: Fossils and family relationships- introducing cladistics
fossils
2.9: Fossil evidence for the history of life on earth
3: Evolutionary Mechanisms and the Diversity of Life
3.3: Fossils and family relationships- introducing cladistics
founder population
3.13: Genetic diversity, population size, founder effects and population bottlenecks
Fragile X syndrome
7.14: DNA repeat diseases and genetic anticipation
fraternal twins
15.14: Concordance between monozygotic twins and genetic influence on a trait
freely available
InfoPage
Friedrich Wöhler
2.5: The Death of Vitalism
frog embryos
19.3: Frogs and fish
fruit flies
7.3: Identifying DNA as the genetic material
fruit fly
19.5: The fruit fly Drosophila melanogaster
function
8.0: Introduction
functional complex
15.6: Interacting traits- synthetic lethality and co-dominance

G

G2 phase
12.1: Ploidy during the cell cycle
galaxies
2.8: Mapping the history of life on earth
gametes
4.8: Origins and implications of sexual reproduction
4.9: Sexual dimorphism
12.4: Steps in meiosis- from diploid to haploid
12.5: Recombination and independent segregation
14.0: Introduction
15.1: How Mendel did what he did
15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
15.8: Maternal and paternal effects
gas
5.9: The implications of bond polarity
gastrointestinal
9.4: E. coli as a model system
gene
3.16: Gene linkage- one more complication
7.10: Further replication complexities
7.12: Genes and alleles
8.8: Turning polypeptides into proteins
gene deletion
14.4: Gene duplications and deletions
gene duplication
14.4: Gene duplications and deletions
14.5: Orthologs and paralogs
14.6: Transposons- moving DNA within a genome (and weird genetics)

gene expression
- 7.11: Mutations, Deletions, Duplications and Repair
- 8.3: Protein synthesis- transcription (DNA to RNA)
- 8.5: The translation (polypeptide synthesis) cycle
- 8.7: Getting more complex- gene regulation in eukaryotes
- 8.9: Regulating protein activity, concentrations, and stability (half-life)
- 9.0: Introduction
- 9.1: Locating information within DNA
- 9.2: Enhancers and transcription start and stop sites
- 9.3: Interaction networks and model systems
- 9.5: Adaptive behavior and gene networks- the lac response
- 9.6: Types of regulatory interactions
- 9.7: Final thoughts on (molecular) noise
- 10: Social Systems
- 10.1: Targeting membrane proteins to where they need to be
- 10.2: Nuclear targeting and exclusion
- 10.3: Intercellular signaling- signals, receptors and responses
- 10.4: Signaling molecules and receptors
- 12.1: Ploidy during the cell cycle
- 12.7: X-inactivation and sex-linked traits
- 14.2: Non-disjunction- aberrant chromosome segregation
- 14.4: Gene duplications and deletions
- 14.5: Orthologs and paralogs
- 14.6: Transposons- moving DNA within a genome (and weird genetics)
- 15.6: Interacting traits- synthetic lethality and co-dominance
- 15.10: Imprinting- conflicts between mother, father, and fetus
- 16: Tools for Studying Genes and Genomes
- 16.1: Synteny examined using Genomicus
- 16.2: Where is a gene expressed?
- 17.0: Introduction
- 17.1: How do systems change at the molecular level?
- 17.3: Direct and indirect cellular responses to signaling molecules
- 17.5: Reversible, irreversible, and cascade effects
- 18.1: How do unicellular organisms generate phenotypic diversity?
- 19.1: Model Systems
- 19.2: Some model systems
- 20: Appendix- Muller's Morphs

gene family
- 14.5: Orthologs and paralogs

gene interaction
- 15.5: Genetic complementation

gene mapping
- 13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies

gene name
- 16.1: Synteny examined using Genomicus

gene networks
- 9: Genomes, Genes and Regulatory networks

gene organization
- 16.0: Introduction

gene pathways
- 15.7: Interacting traits- epistasis

gene product
- 7.12: Genes and alleles

gene products
- 7.13: The origin of new (de novo) genes
- 14.4: Gene duplications and deletions
- 15.6: Interacting traits- synthetic lethality and co-dominance
- 15.7: Interacting traits- epistasis
- 16.0: Introduction

gene regulation
- 7.11: Mutations, Deletions, Duplications and Repair
- 8.3: Protein synthesis- transcription (DNA to RNA)
- 8.6: Effects of point mutations on polypeptides and proteins
- 9.1: Locating information within DNA
- 9.2: Enhancers and transcription start and stop sites
- 9.5: Adaptive behavior and gene networks- the lac response
- 9.7: Final thoughts on (molecular) noise
- 10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
- 12.1: Ploidy during the cell cycle
- 19.6: The nematode Caenorhabditus elegans
- 19.8: ESC and iPSC derived organoids

gene regulatory factors
- 17.2: Steady state and changing molecular concentrations- synthesis and degradation

gene regulatory sequences
- 9.2: Enhancers and transcription start and stop sites

gene splicing
- 16.2: Where is a gene expressed?

gene transfer
- 11.2: Other naturally occurring horizontal gene transfer mechanisms

genes
- 3.7: Mutations and the origins of genotype-based variation
- 3.9: Genotype-phenotype relationships- discrete and continuous traits
- 9.0: Introduction
- 9.1: Locating information within DNA
- 14.0: Introduction
- 16.0: Introduction
- 17.1: How do systems change at the molecular level?
- 17.4: Modeling gene expression

genetic
- 3.16: Gene linkage- one more complication
- 7.1: Discovering how nucleic acids store genetic information
- 15.5: Genetic complementation

genetic accuracy
- 7.9: Replication machines

genetic analyses
- 19.5: The fruit fly Drosophila melanogaster

genetic analysis
- 13.3: Forward and reverse genetics
- 15.7: Interacting traits- epistasis
- 15.8: Maternal and paternal effects

genetic background
- 3.8: The origins of polymorphisms

genetic changes
- 7.3: Identifying DNA as the genetic material

genetic code
- 2.1: What is life, exactly?
- 8.1: Specifying a polypeptide's sequence
- 8.2: The origin of the genetic code

genetic complementation
- 15.5: Genetic complementation

genetic concordance
- 15.14: Concordance between monozygotic twins and genetic influence on a trait

genetic contributions
- 15.14: Concordance between monozygotic twins and genetic influence on a trait

genetic data
- 6.12: An evolutionary scenario for the origin of eukaryotic cells

genetic differences
- 3.7: Mutations and the origins of genotype-based variation

genetic diseases
- 15: Mendel and Weldon - Contexts and their Effects on Phenotypes
- 15.11: Estimating the number of genes involved in a particular traits

genetic disorder
- 14.2: Non-disjunction- aberrant chromosome segregation

genetic diversity
- 3: Evolutionary Mechanisms and the Diversity of Life
- 3.4: Evolution theory's core concepts
- 3.7: Mutations and the origins of genotype-based variation
- 4.8: Origins and implications of sexual reproduction
- 12.5: Recombination and independent segregation

genetic drift
- 3.0: Introduction
- 3.12: Considering stochastic processes
- 3.13: Genetic diversity, population size, founder effects and population bottlenecks
- 3.15: Genetic drift
- 3.17: A brief reflection on the complexity of phenotypic traits
- 13.1: Mutations into alleles
- 13.5: Longer term mutation and evolution studies
- 15.15: Measuring evolution's impact on allele frequencies- Hardy-Weinberg
- 16.3: Using web-based bioinformatic tools- gnomAD

genetic elements
- 15.0: Introduction

genetic engineering
- 9.3: Interaction networks and model systems
- 11.2: Other naturally occurring horizontal gene transfer mechanisms
- 16.1: Synteny examined using Genomicus

genetic equilibrium
- 15.15: Measuring evolution's impact on allele frequencies- Hardy-Weinberg

genetic factors
- 3.14: Population bottlenecks

genetic genes
- 16.5: Genome-wide Association Studies (GWAS)

genetic information
- 7: The Molecular Nature of Heredity
- 7.2: Locating hereditary material within the cell
- 7.3: Identifying DNA as the genetic material
- 7.8: DNA Replication
- 11.0: Introduction

genetic interaction
- 15.6: Interacting traits- synthetic lethality and co-dominance

genetic linkage
- 3.16: Gene linkage- one more complication

genetic loci
- 15.6: Interacting traits- synthetic lethality and co-dominance

genetic manipulation
- 19.7: The Mouse

genetic markers
- 15.5: Genetic complementation

genetic material
- 6.13: Making a complete eukaryote
- 7: The Molecular Nature of Heredity
- 7.4: Unraveling Nucleic Acid Structure
- 7.11: Mutations, Deletions, Duplications and Repair
- 10.0: Introduction

genetic mechanisms
12.3: Sex-determination and its chromosomal basis
13.2: Luria and Delbrück- Discovering the origin of mutations
genetic mutation
18.3: Quorum effects (somewhat redundant)
genetic mutations
8.11: Diseases of folding and misfolding
15.8: Maternal and paternal effects
20: Appendix- Muller's Morphs
genetic polymorphisms
3.9: Genotype-phenotype relationships- discrete and continuous traits
genetic relatedness
18.6: Evolutionary origins of clonal (permanent) multicellularity
genetic relationship
4.4: Inclusive fitness, kin and group selection, and social evolution
genetic screens
13.3: Forward and reverse genetics
genetic selection
12.6: Linkage and haplotypes
genetic studies
19.3: Frogs and fish
genetic susceptibility
14.1: Rates and effects of somatic mutation
genetic variation
2.0: Introduction
3.11: Types of simple selection
3.18: Speciation and extinction
3.19: Mechanisms of speciation
7.12: Genes and alleles
13.1: Mutations into alleles
13.2: Luria and Delbrück- Discovering the origin of mutations
14.3: Genome dynamics
14.6: Transposons- moving DNA within a genome (and weird genetics)
15.2: Weldon's critique
15.13: Alleles, traits, and genetic diseases in humans.
genetic_variation
3.20: Mechanisms of reproductive isolation
genetics
3.9: Genotype-phenotype relationships- discrete and continuous traits
3.10: Variation, selection, and speciation
3.23: Homologies provide evidence for a common ancestor
7.0: Introduction
7.14: DNA repeat diseases and genetic anticipation
9.0: Introduction
12.8: X-linked diseases and mono-allelic gene expression
13: How Alleles Arise - Mutations
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
15: Mendel and Weldon - Contexts and their Effects on Phenotypes
15.0: Introduction
15.1: How Mendel did what he did
15.2: Weldon's critique
15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
15.4: Dihybrid crosses- linkage and recombination
15.13: Alleles, traits, and genetic diseases in humans.
15.14: Concordance between monozygotic twins and genetic influence on a trait
16.3: Using web-based bioinformatic tools- gnomAD
19.5: The fruit fly Drosophila melanogaster
19.6: The nematode Caenorhabditus elegans
genome
2.2: The cell theory and the continuity of life
3.7: Mutations and the origins of genotype-based variation
3.16: Gene linkage- one more complication
6.12: An evolutionary scenario for the origin of eukaryotic cells
7.6: DNA, sequences, and information
7.13: The origin of new (de novo) genes
14.0: Introduction
14.3: Genome dynamics
16.5: Genome-wide Association Studies (GWAS)
genome editing
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
genome evolution
14.5: Orthologs and paralogs
genome sequencing
3.23: Homologies provide evidence for a common ancestor
16.2: Where is a gene expressed?
genomes
7.0: Introduction
16.0: Introduction
genomic
13.0: Introduction
genomic analysis
16.1: Synteny examined using Genomicus
genomic differences
19.8: ESC and iPSC derived organoids
genomic DNA
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
genomic variations
3.8: The origins of polymorphisms
genomicus
16: Tools for Studying Genes and Genomes
genotype
2.2: The cell theory and the continuity of life
3.8: The origins of polymorphisms
3.15: Genetic drift
15.4: Dihybrid crosses- linkage and recombination
15.11: Estimating the number of genes involved in a particular traits
genotypes
3.7: Mutations and the origins of genotype-based variation
7.0: Introduction
12.7: X-inactivation and sex-linked traits
genus
3.2: Natural and un-natural groups
geological information
2.9: Fossil evidence for the history of life on earth
geology
2.6: Thinking about life's origins
2.9: Fossil evidence for the history of life on earth
georges lemaitre
2.8: Mapping the history of life on earth
germ cells
2.3: The organization of organisms
germ line
13.1: Mutations into alleles
germ line cell
12.5: Recombination and independent segregation
germ line cells
4.7: The appearance of multicellular organisms
germline mutations
14.1: Rates and effects of somatic mutation
Gibbs Free Energy
5.2: Coupling thermodynamically favorable and unfavorable reactions
5.3: Reaction rates
gnomAD
16: Tools for Studying Genes and Genomes
gradients
6: Membrane Boundaries and Capturing Energy
group selection
4.5: Group selection
growth
5.0: Introduction
6.0: Introduction
growth control
14.0: Introduction
guide RNA
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
GWAS
16: Tools for Studying Genes and Genomes

H

H+ gradients
6.7: Chemo-osmosis (an overview)
H.J. Muller
7.3: Identifying DNA as the genetic material
H2
6.9: Chemotrophs
hadean
2.9: Fossil evidence for the history of life on earth
Halobacterium
6.6: Simple Phototrophs
haploid
4.8: Origins and implications of sexual reproduction
12: Asexual and Sexual Reproduction in Eukaryotes
12.1: Ploidy during the cell cycle
12.4: Steps in meiosis- from diploid to haploid
haploid cells
14.4: Gene duplications and deletions
haploid gametes
12.0: Introduction
haploidy
12.8: X-linked diseases and mono-allelic gene expression
haploinsufficiency
20: Appendix- Muller's Morphs
haplotype
12.6: Linkage and haplotypes
height
15.11: Estimating the number of genes involved in a particular traits
hereditary changes
7.3: Identifying DNA as the genetic material
hereditary information
2.2: The cell theory and the continuity of life
7.2: Locating hereditary material within the cell
hereditary material
7: The Molecular Nature of Heredity
heredity
7.1: Discovering how nucleic acids store genetic information
heritability
3.9: Genotype-phenotype relationships- discrete and continuous traits
15.16: Genetic anticipation
16.5: Genome-wide Association Studies (GWAS)

hermaphrodite
12.3: Sex-determination and its chromosomal basis
hermaphrodites
19.6: The nematode Caenorhabditus elegans
heterozygote
15.15: Measuring evolution’s impact on allele frequencies- Hardy-Weinberg
heterozygous
4.8: Origins and implications of sexual reproduction
historical origins of genetics
15.0: Introduction
history
1.4: Teaching and Learning Science
history of life
2: Life’s Diversity and Origins
2.4: Spontaneous generation and the origin of life
history of science
1.0: Introduction
homeostasis
6.1: Defining the Cell’s Boundary
17.0: Introduction
Homo sapiens
19.0: Introduction
homologous chromosomes
14.4: Gene duplications and deletions
homologous genes
14.5: Orthologs and paralogs
homologous recombination
11.1: What counts as sex in prokaryotes
homologous trait
8.2: The origin of the genetic code
homologous traits
3.22: Signs of evolution- homology and convergence
homology
3.23: Homologies provide evidence for a common ancestor
homozygote
15.15: Measuring evolution’s impact on allele frequencies- Hardy-Weinberg
homozygous
4.8: Origins and implications of sexual reproduction
horizontal gene transfer
11.0: Introduction
11.1: What counts as sex in prokaryotes
11.2: Other naturally occurring horizontal gene transfer mechanisms
14.3: Genome dynamics
14.6: Transposons- moving DNA within a genome (and weird genetics)
human birth defects
19.0: Introduction
human development
12.8: X-linked diseases and mono-allelic gene expression
19.1: Model Systems
human evolution
3.15: Genetic drift
human genome
3.7: Mutations and the origins of genotype-based variation
human origins
3.24: Anti-evolution arguments
human population
3.5: Limits on populations
human reproduction
4.9: Sexual dimorphism
human traits
19.8: ESC and iPSC derived organoids
Huntington's disease
7.14: DNA repeat diseases and genetic anticipation
15.16: Genetic anticipation
hybrid sterility
14.3: Genome dynamics
hybridization
3.20: Mechanisms of reproductive isolation
hydrogen bonding
7.5: Discovering the structure of DNA
hydrogen bonds
5.8: Bond polarity, inter- and intramolecular interactions
7.6: DNA, sequences, and information
hydrophilic
5.10: Interacting with Water
hydrophobic
5.10: Interacting with Water
6.2: The origin of biological membranes
hydrothermal vents
2.7: Experimental studies on the origins of life
hypotheses
1: Understanding Science and Thinking Scientifically
1.2: Models, Hypotheses, and Theories
1.3: Science is Social
hypothesis testing
15: Mendel and Weldon - Contexts and their Effects on Phenotypes

I

identical twins
15.14: Concordance between monozygotic twins and genetic influence on a trait
immune system
4.3: Active (altruistic) cell death
13.2: Luria and Delbrück- Discovering the origin of mutations
immunocytochemistry
16.2: Where is a gene expressed?
imprinting
15.10: Imprinting- conflicts between mother, father, and fetus
in situ hybridization
16.2: Where is a gene expressed?
inclusive fitness
4: Social Evolution and Sexual Selection
4.1: Selecting social (cooperative) traits
4.4: Inclusive fitness, kin and group selection, and social evolution
4.5: Group selection
18.2: Dying for others – social interactions between “unicellular” organisms
individuals
3.8: The origins of polymorphisms
induced pluripotent stem cells
19.3: Frogs and fish
19.8: ESC and iPSC derived organoids
inheritable variation
3.7: Mutations and the origins of genotype-based variation
inheritance
3.4: Evolution theory’s core concepts
3.9: Genotype-phenotype relationships- discrete and continuous traits
7.0: Introduction
7.1: Discovering how nucleic acids store genetic information
12.8: X-linked diseases and mono-allelic gene expression
15.1: How Mendel did what he did
15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
15.4: Dihybrid crosses- linkage and recombination
inherited ataxias
15.16: Genetic anticipation
inherited diseases
7.14: DNA repeat diseases and genetic anticipation
inorganic molecules
6.9: Chemotrophs
intelligent design
2.6: Thinking about life’s origins
interaction networks
9: Genomes, Genes and Regulatory networks
9.3: Interaction networks and model systems
interactive textbooks
InfoPage
intercellular communication
10.3: Intercellular signaling- signals, receptors and responses
interconnectedness
1: Understanding Science and Thinking Scientifically
intermolecular forces
5.8: Bond polarity, inter- and intramolecular interactions
intermolecular interactions
5.6: Covalent bonds
5.7: Bond Stability and Thermal Motion (a non-biological moment)
intramolecular interactions
5.6: Covalent bonds
intrinsic policing
4.6: Defense against social cheaters
introduction
3: Evolutionary Mechanisms and the Diversity of Life
12: Asexual and Sexual Reproduction in Eukaryotes
16: Tools for Studying Genes and Genomes
inversions
14.3: Genome dynamics
ion transport
6.4: Channels and carriers
iPS cells
17.5: Reversible, irreversible, and cascade effects
iPSCs
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
isotope dating
2.9: Fossil evidence for the history of life on earth

J

Joachim Hammerling
7.2: Locating hereditary material within the cell
juxtacrine signaling
10.3: Intercellular signaling- signals, receptors and responses

K

kin selection
- 4: Social Evolution and Sexual Selection
- 4.1: Selecting social (cooperative) traits
- 4.5: Group selection
- 4.11: Curbing runaway selection

kinetic energy
- 5.7: Bond Stability and Thermal Motion (a non-biological moment)

L

lab research
- 19.6: The nematode Caenorhabditus elegans

lac operon
- 9.5: Adaptive behavior and gene networks- the lac response
- 9.7: Final thoughts on (molecular) noise
- 15.7: Interacting traits- epistasis

lactase
- 16.1: Synteny examined using Genomicus

lactose
- 9.5: Adaptive behavior and gene networks- the lac response

lactose permease
- 9.5: Adaptive behavior and gene networks- the lac response

lagging strand
- 7.8: DNA Replication

law of large numbers
- 3.12: Considering stochastic processes

leading strand
- 7.8: DNA Replication

learning
- 1: Understanding Science and Thinking Scientifically

LeChatelier
- 5.4: Coupling Reactions

LibreTexts
- InfoPage

life
- 1.1: The interconnectedness (self-consistency) of science

life cycle
- 19.6: The nematode Caenorhabditus elegans

life origin
- 2.8: Mapping the history of life on earth

life processes
- 2.5: The Death of Vitalism

life sciences
- 5.1: Approaching Thermodynamics

light absorption
- 6.6: Simple Phototrophs

light harvesting
- 6.8: Oxygenic Photosynthesis

light microscope
- 2.4: Spontaneous generation and the origin of life

linkage
- 12: Asexual and Sexual Reproduction in Eukaryotes
- 15: Mendel and Weldon - Contexts and their Effects on Phenotypes

linkage disequilibrium
- 12.6: Linkage and haplotypes

Linnaean
- 3.2: Natural and un-natural groups

Linnaeus
- 3.1: Organizing organisms (hierarchically)

lipid
- 6.1: Defining the Cell’s Boundary

lipid bilayer
- 6.3: Transport across membranes

lipids
- 6.2: The origin of biological membranes

liquid
- 5.9: The implications of bond polarity

lissencephaly
- 19.8: ESC and iPSC derived organoids

lithotrophs
- 6.9: Chemotrophs

living organisms
- 5.0: Introduction

logical basis
- 1.2: Models, Hypotheses, and Theories

LUCA
- 1.1: The interconnectedness (self-consistency) of science
- 2.7: Experimental studies on the origins of life
- 2.10: Life's impact on earth
- 8.2: The origin of the genetic code

Luria
- 13: How Alleles Arise - Mutations

M

macromolecular complexes
- 15.7: Interacting traits- epistasis

macromolecular synthesis
- 6.0: Introduction

macromolecules
- 5.6: Covalent bonds

Malthus
- 3.5: Limits on populations

mammal development
- 19.7: The Mouse

mass extinction
- 2.1: What is life, exactly?

mate choice
- 4.10: Sexual Selection
- 4.11: Curbing runaway selection

mate_selection
- 3.20: Mechanisms of reproductive isolation

maternal effect
- 15.9: Mitochondrial inheritance

maternal effect mutations
- 15.8: Maternal and paternal effects

maternal effects
- 15.10: Imprinting- conflicts between mother, father, and fetus

mating
- 19.5: The fruit fly Drosophila melanogaster

mating behavior
- 4.10: Sexual Selection

mating preference
- 3.21: Sympatric speciation

mating systems
- 4.9: Sexual dimorphism

matter
- 5.5: Inter- and intra-molecular interactions

medicine
- 16.3: Using web-based bioinformatic tools- gnomAD

meiosis
- 4.8: Origins and implications of sexual reproduction
- 12: Asexual and Sexual Reproduction in Eukaryotes
- 12.0: Introduction
- 12.3: Sex-determination and its chromosomal basis
- 12.4: Steps in meiosis- from diploid to haploid
- 12.5: Recombination and independent segregation
- 12.8: X-linked diseases and mono-allelic gene expression
- 13.1: Mutations into alleles
- 14.0: Introduction
- 15.1: How Mendel did what he did
- 15.3: Chi square (?2) analysis, hypothesis testing, and small numbers

meiotic alignment
- 14.4: Gene duplications and deletions

meiotic recombination
- 12.6: Linkage and haplotypes
- 14.3: Genome dynamics

melting curves
- 7.6: DNA, sequences, and information

melting points
- 5.9: The implications of bond polarity

membrane
- 6.1: Defining the Cell’s Boundary
- 6.5: Generating gradients- using coupled reactions and pumps

membrane dynamics
- 6.4: Channels and carriers

membrane permeability
- 6.4: Channels and carriers

membrane proteins
- 6.4: Channels and carriers
- 6.10: Using the energy stored in membrane gradients
- 10: Social Systems

membrane transport
- 6.3: Transport across membranes

membranes
- 6: Membrane Boundaries and Capturing Energy
- 6.2: The origin of biological membranes

Mendel
- 3.16: Gene linkage- one more complication
- 7.1: Discovering how nucleic acids store genetic information
- 15: Mendel and Weldon - Contexts and their Effects on Phenotypes
- 15.0: Introduction
- 15.1: How Mendel did what he did
- 15.2: Weldon's critique
- 15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
- 15.11: Estimating the number of genes involved in a particular traits

Mendelian Genetics
- 15.8: Maternal and paternal effects

Mendelian inheritance
- 16.5: Genome-wide Association Studies (GWAS)

messenger RNA
- 7.7: Discovering RNA- structure and some functions

metabolic pathway
- 9.5: Adaptive behavior and gene networks- the lac response

metabolism
- 5.4: Coupling Reactions

metamorphosis
- 19.5: The fruit fly Drosophila melanogaster

metazoans
- 18.4: Transient and clonal ("true") metazoans

methane
- 6.9: Chemotrophs

methods in biology
16.0: Introduction
micelles
6.2: The origin of biological membranes
microbes
2.4: Spontaneous generation and the origin of life
9.4: E. coli as a model system
microbial communities
10: Social Systems
11.2: Other naturally occurring horizontal gene transfer mechanisms
microbial ecology
18.3: Quorum effects (somewhat redundant)
microbial ecosystems
9.4: E. coli as a model system
microbial growth
2.5: The Death of Vitalism
microbial life
2.1: What is life, exactly?
microbiology
9.5: Adaptive behavior and gene networks- the lac response
microbiome
9.4: E. coli as a model system
microdissection
7.2: Locating hereditary material within the cell
microsatellite expansion
15.16: Genetic anticipation
microscopy
19.2: Some model systems
microsurgical manipulations
19.3: Frogs and fish
migration
3.13: Genetic diversity, population size, founder effects and population bottlenecks
misfolding
8.11: Diseases of folding and misfolding
missense mutation
15.12: On the nature of mutations (again)
mistranslation
8.2: The origin of the genetic code
mitochondria
2.10: Life's impact on earth
15.9: Mitochondrial inheritance
mitochondrial
6.13: Making a complete eukaryote
mitochondrial genomes
15.9: Mitochondrial inheritance
mitosis
12: Asexual and Sexual Reproduction in Eukaryotes
12.0: Introduction
12.2: Monitoring cellular processes- mitosis
12.5: Recombination and independent segregation
mobile genetic elements
11.1: What counts as sex in prokaryotes
model organism
19.5: The fruit fly Drosophila melanogaster
model organisms
9.3: Interaction networks and model systems
19.1: Model Systems
19.2: Some model systems
model systems
9: Genomes, Genes and Regulatory networks
19.0: Introduction
19.8: ESC and iPSC derived organoids
models
1: Understanding Science and Thinking Scientifically
modern physics
1.4: Teaching and Learning Science
molecular biology
3.7: Mutations and the origins of genotype-based variation
3.10: Variation, selection, and speciation
3.23: Homologies provide evidence for a common ancestor
7.5: Discovering the structure of DNA
7.8: DNA Replication
8.1: Specifying a polypeptide's sequence
8.5: The translation (polypeptide synthesis) cycle
8.12: Molecular machines
9.4: E. coli as a model system
16.0: Introduction
molecular cascades
10.4: Signaling molecules and receptors
molecular changes
19.0: Introduction
molecular components
5.2: Coupling thermodynamically favorable and unfavorable reactions
molecular dynamics
5.6: Covalent bonds
8.12: Molecular machines
molecular interactions
1.5: Understanding Scientific Ideas
5.0: Introduction
5.5: Inter- and intra-molecular interactions
5.9: The implications of bond polarity
9.7: Final thoughts on (molecular) noise
10: Social Systems
17.2: Steady state and changing molecular concentrations- synthesis and degradation
17.3: Direct and indirect cellular responses to signaling molecules
molecular interventions
19.2: Some model systems
molecular machines
8.12: Molecular machines
molecular mechanisms
2.4: Spontaneous generation and the origin of life
7.3: Identifying DNA as the genetic material
18.0: Introduction
molecular movement
5.1: Approaching Thermodynamics
molecular noise
9: Genomes, Genes and Regulatory networks
18.1: How do unicellular organisms generate phenotypic diversity?
molecular processes
17.0: Introduction
molecular shape
5.5: Inter- and intra-molecular interactions
5.6: Covalent bonds
molecular structures
5.6: Covalent bonds
molecular systems
15.7: Interacting traits- epistasis
molecular transport
6.4: Channels and carriers
molecular variation
15.0: Introduction
molecular visualization
8.12: Molecular machines
molecule solubility
5.10: Interacting with Water
molecules
5.5: Inter- and intra-molecular interactions
5.7: Bond Stability and Thermal Motion (a non-biological moment)
monohybrid crosses
15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
monozygotic twins
15.14: Concordance between monozygotic twins and genetic influence on a trait
morality
3.24: Anti-evolution arguments
morphogen gradient
17.4: Modeling gene expression
mosaicism
12.8: X-linked diseases and mono-allelic gene expression
14.2: Non-disjunction- aberrant chromosome segregation
mother fetus conflict
15.10: Imprinting- conflicts between mother, father, and fetus
mouse model
19.7: The Mouse
mRNA
7.7: Discovering RNA- structure and some functions
7.11: Mutations, Deletions, Duplications and Repair
8.3: Protein synthesis- transcription (DNA to RNA)
8.4: Protein synthesis- translation (RNA to polypeptide)
8.5: The translation (polypeptide synthesis) cycle
8.6: Effects of point mutations on polypeptides and proteins
10.1: Targeting membrane proteins to where they need to be
mRNA processing
8.7: Getting more complex- gene regulation in eukaryotes
mRNA synthesis
17.2: Steady state and changing molecular concentrations- synthesis and degradation
mtDNA
15.9: Mitochondrial inheritance
multicellular
2.3: The organization of organisms
18.6: Evolutionary origins of clonal (permanent) multicellularity
multicellular aggregates
18.5: Unicellular to multicellularity to unicellularity
multicellular organism
17.0: Introduction
multicellular organisms
4.0: Introduction
4.3: Active (altruistic) cell death
4.6: Defense against social cheaters
4.7: The appearance of multicellular organisms
9.2: Enhancers and transcription start and stop sites
10: Social Systems
18.0: Introduction
19.1: Model Systems
multicellularity
4: Social Evolution and Sexual Selection
4.1: Selecting social (cooperative) traits
18: Supplemental Chapter - Social interactions between cells
18.4: Transient and clonal ("true") metazoans
Mus musculus
19.7: The Mouse

muscular dystrophies
15.16: Genetic anticipation
mutagenesis
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
mutant alleles
13.3: Forward and reverse genetics
mutation
3.15: Genetic drift
7.0: Introduction
8.8: Turning polypeptides into proteins
13.1: Mutations into alleles
13.5: Longer term mutation and evolution studies
mutation rates
14.1: Rates and effects of somatic mutation
mutations
3.8: The origins of polymorphisms
3.9: Genotype-phenotype relationships- discrete and continuous traits
3.10: Variation, selection, and speciation
3.11: Types of simple selection
3.12: Considering stochastic processes
3.14: Population bottlenecks
3.17: A brief reflection on the complexity of phenotypic traits
7: The Molecular Nature of Heredity
7.3: Identifying DNA as the genetic material
7.10: Further replication complexities
7.11: Mutations, Deletions, Duplications and Repair
7.12: Genes and alleles
7.14: DNA repeat diseases and genetic anticipation
8.0: Introduction
8.6: Effects of point mutations on polypeptides and proteins
9.1: Locating information within DNA
12.4: Steps in meiosis- from diploid to haploid
13: How Alleles Arise - Mutations
13.0: Introduction
13.2: Luria and Delbrück- Discovering the origin of mutations
13.3: Forward and reverse genetics
14.0: Introduction
14.6: Transposons- moving DNA within a genome (and weird genetics)
15: Mendel and Weldon - Contexts and their Effects on Phenotypes
15.5: Genetic complementation
15.7: Interacting traits- epistasis
15.9: Mitochondrial inheritance
15.11: Estimating the number of genes involved in a particular traits
16.3: Using web-based bioinformatic tools- gnomAD
19.5: The fruit fly Drosophila melanogaster
MyoD
20: Appendix- Muller's Morphs

N

natural groups
3: Evolutionary Mechanisms and the Diversity of Life
3.2: Natural and un-natural groups
natural selection
3.0: Introduction
3.3: Fossils and family relationships- introducing cladistics
3.4: Evolution theory's core concepts
3.6: Darwin and Wallace's conceptual leap
3.10: Variation, selection, and speciation
3.11: Types of simple selection
3.18: Speciation and extinction
3.19: Mechanisms of speciation
3.24: Anti-evolution arguments
4.1: Selecting social (cooperative) traits
4.4: Inclusive fitness, kin and group selection, and social evolution
4.11: Curbing runaway selection
13.2: Luria and Delbrück- Discovering the origin of mutations
13.5: Longer term mutation and evolution studies
15.15: Measuring evolution's impact on allele frequencies- Hardy-Weinberg
naturalists
2.5: The Death of Vitalism
nebulae
2.8: Mapping the history of life on earth
necrosis
4.3: Active (altruistic) cell death
negative feedback
17.5: Reversible, irreversible, and cascade effects
negative selection
12.6: Linkage and haplotypes
neomorphic mutations
20: Appendix- Muller's Morphs
neurobiology
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
neurological disorder
8.11: Diseases of folding and misfolding
neutral mutation
15.12: On the nature of mutations (again)
nitrogen fixation
18.2: Dying for others – social interactions between "unicellular" organisms
nitrogenous base
7.4: Unraveling Nucleic Acid Structure
nuclear envelope
6.13: Making a complete eukaryote
nuclear transplant
19.3: Frogs and fish
nucleic acid
7.5: Discovering the structure of DNA
nucleic acid structure
7: The Molecular Nature of Heredity
nucleic acids
7: The Molecular Nature of Heredity
7.4: Unraveling Nucleic Acid Structure
7.7: Discovering RNA- structure and some functions
nucleotide
7.5: Discovering the structure of DNA
15.12: On the nature of mutations (again)
nucleotide sequences
7.6: DNA, sequences, and information
14.5: Orthologs and paralogs
nucleotides
7.4: Unraveling Nucleic Acid Structure
7.8: DNA Replication
9.0: Introduction
nucleus
2.10: Life's impact on earth
7.1: Discovering how nucleic acids store genetic information
10.0: Introduction
nucleus transplantation
7.2: Locating hereditary material within the cell
nutrition
9.4: E. coli as a model system

O

observation
1.0: Introduction
observations
1.3: Science is Social
1.4: Teaching and Learning Science
OER
InfoPage
offspring
19.5: The fruit fly Drosophila melanogaster
oocyte
15.9: Mitochondrial inheritance
open systems
5.1: Approaching Thermodynamics
organ formation
17.1: How do systems change at the molecular level?
organic chemistry
2.5: The Death of Vitalism
organic molecules
2.5: The Death of Vitalism
2.7: Experimental studies on the origins of life
organisms
1.1: The interconnectedness (self-consistency) of science
2: Life's Diversity and Origins
2.0: Introduction
2.3: The organization of organisms
3.0: Introduction
3.1: Organizing organisms (hierarchically)
3.3: Fossils and family relationships- introducing cladistics
organization
2: Life's Diversity and Origins
organizing organisms
3: Evolutionary Mechanisms and the Diversity of Life
organoids
19.8: ESC and iPSC derived organoids
organotrophs
6.9: Chemotrophs
origin of life
2.7: Experimental studies on the origins of life
6.12: An evolutionary scenario for the origin of eukaryotic cells
origin of universe
2.8: Mapping the history of life on earth
origins of replication
7.9: Replication machines
orthologs
14.5: Orthologs and paralogs
osmosis
6.3: Transport across membranes
6.11: Osmosis and living with and without a cell wall
osmotic movement
6.11: Osmosis and living with and without a cell wall
oxygen
6.8: Oxygenic Photosynthesis

P

paleontology
- 2.9: Fossil evidence for the history of life on earth

paracrine signaling
- 10.3: Intercellular signaling- signals, receptors and responses

paralogs
- 14.5: Orthologs and paralogs

parental care
- 4.9: Sexual dimorphism

parental investment
- 4.10: Sexual Selection

pathology
- 15.16: Genetic anticipation

pedigree analysis
- 16.3: Using web-based bioinformatic tools- gnomAD

permanent multicellularity
- 18: Supplemental Chapter - Social interactions between cells

permeability
- 6.2: The origin of biological membranes

phage resistance
- 13.2: Luria and Delbrück- Discovering the origin of mutations

phagocytosis
- 6.12: An evolutionary scenario for the origin of eukaryotic cells

phase change
- 5.7: Bond Stability and Thermal Motion (a non-biological moment)

phase transitions
- 5.9: The implications of bond polarity

phenotype
- 3.15: Genetic drift
- 4.8: Origins and implications of sexual reproduction
- 13.3: Forward and reverse genetics
- 15.1: How Mendel did what he did
- 15.4: Dihybrid crosses- linkage and recombination
- 15.8: Maternal and paternal effects

phenotype interaction
- 15.6: Interacting traits- synthetic lethality and co-dominance

phenotypes
- 3.9: Genotype-phenotype relationships- discrete and continuous traits
- 3.11: Types of simple selection
- 7.0: Introduction
- 13.5: Longer term mutation and evolution studies
- 15.2: Weldon's critique
- 15.3: Chi square (?2) analysis, hypothesis testing, and small numbers
- 15.5: Genetic complementation
- 15.7: Interacting traits- epistasis
- 15.9: Mitochondrial inheritance
- 15.11: Estimating the number of genes involved in a particular traits
- 15.13: Alleles, traits, and genetic diseases in humans.
- 15.16: Genetic anticipation
- 18.0: Introduction

phenotypic diversity
- 18: Supplemental Chapter - Social interactions between cells

phenotypic effect
- 15.12: On the nature of mutations (again)

phenotypic effects
- 3.17: A brief reflection on the complexity of phenotypic traits

phenotypic heterogeneity
- 18.1: How do unicellular organisms generate phenotypic diversity?

phenotypic ratio
- 15.4: Dihybrid crosses- linkage and recombination

phenotypic traits
- 16.5: Genome-wide Association Studies (GWAS)

phenotypic variation
- 3.4: Evolution theory’s core concepts
- 3.6: Darwin and Wallace's conceptual leap
- 3.10: Variation, selection, and speciation
- 3.18: Speciation and extinction

phenotypic variations
- 3.8: The origins of polymorphisms

phenotypic_distinction
- 3.20: Mechanisms of reproductive isolation

philosophy of science
- 1.0: Introduction

photoisomerization
- 6.6: Simple Phototrophs

photons
- 6.6: Simple Phototrophs

photosynthesis
- 2.10: Life's impact on earth
- 6: Membrane Boundaries and Capturing Energy
- 6.8: Oxygenic Photosynthesis

photosynthetic bacteria
- 6.0: Introduction

phototrophs
- 6.6: Simple Phototrophs
- 6.7: Chemo-osmosis (an overview)

phylogenetic relationships
- 3.22: Signs of evolution- homology and convergence

phylogenetic tree
- 3.1: Organizing organisms (hierarchically)

physical properties
- 5.9: The implications of bond polarity

physicochemical principles
- 1.1: The interconnectedness (self-consistency) of science

physics
- 1.1: The interconnectedness (self-consistency) of science
- 2.6: Thinking about life’s origins

physiological adaptations
- 3.6: Darwin and Wallace's conceptual leap

physiological similarities
- 19.7: The Mouse

pigments
- 6.6: Simple Phototrophs

placental mammals
- 15.10: Imprinting- conflicts between mother, father, and fetus

plant breeding
- 3.4: Evolution theory’s core concepts

plants
- 6.8: Oxygenic Photosynthesis

plasma membrane
- 6.11: Osmosis and living with and without a cell wall
- 10.0: Introduction

plasmid replication
- 11.1: What counts as sex in prokaryotes

plasmids
- 11.0: Introduction

pleiotropy
- 20: Appendix- Muller’s Morphs

ploidy
- 12: Asexual and Sexual Reproduction in Eukaryotes
- 12.1: Ploidy during the cell cycle

point mutation
- 15.12: On the nature of mutations (again)

point mutations
- 14.3: Genome dynamics

polar bond
- 5.8: Bond polarity, inter- and intramolecular interactions

polymerase
- 7.10: Further replication complexities

polymerase chain reaction
- 16.2: Where is a gene expressed?

polymorphism
- 3.8: The origins of polymorphisms

polymorphisms
- 3.7: Mutations and the origins of genotype-based variation

polypeptide
- 8.4: Protein synthesis- translation (RNA to polypeptide)
- 8.6: Effects of point mutations on polypeptides and proteins
- 8.8: Turning polypeptides into proteins
- 10.1: Targeting membrane proteins to where they need to be

polypeptide synthesis
- 8.3: Protein synthesis- transcription (DNA to RNA)
- 8.7: Getting more complex- gene regulation in eukaryotes
- 10.0: Introduction
- 10.2: Nuclear targeting and exclusion

polypeptides
- 8.0: Introduction
- 8.1: Specifying a polypeptide’s sequence

population
- 3.8: The origins of polymorphisms
- 3.14: Population bottlenecks
- 13.0: Introduction

population behavior
- 3.12: Considering stochastic processes

population bottleneck
- 3.13: Genetic diversity, population size, founder effects and population bottlenecks

population dynamics
- 3.10: Variation, selection, and speciation
- 3.13: Genetic diversity, population size, founder effects and population bottlenecks
- 18.2: Dying for others – social interactions between “unicellular” organisms
- 18.3: Quorum effects (somewhat redundant)

population genetics
- 3.15: Genetic drift
- 3.17: A brief reflection on the complexity of phenotypic traits
- 7.13: The origin of new (de novo) genes
- 13.1: Mutations into alleles
- 15.15: Measuring evolution’s impact on allele frequencies- Hardy-Weinberg

population growth
- 3.5: Limits on populations

positive selection
- 12.6: Linkage and haplotypes

prebiotic chemistry
- 2.7: Experimental studies on the origins of life

precision medicine
- 9.3: Interaction networks and model systems

predation avoidance
18.6: Evolutionary origins of clonal (permanent) multicellularity
predictive models
1.2: Models, Hypotheses, and Theories
pressure balance
6.11: Osmosis and living with and without a cell wall
primers
16.2: Where is a gene expressed?
prion
8.11: Diseases of folding and misfolding
prion diseases
8.11: Diseases of folding and misfolding
processivity
7.9: Replication machines
programmed cell death
4: Social Evolution and Sexual Selection
4.3: Active (altruistic) cell death
prokaryotes
2.2: The cell theory and the continuity of life
6.13: Making a complete eukaryote
7.10: Further replication complexities
8.3: Protein synthesis- transcription (DNA to RNA)
8.7: Getting more complex- gene regulation in eukaryotes
10.0: Introduction
10.1: Targeting membrane proteins to where they need to be
11.0: Introduction
proofreading
7.9: Replication machines
proteases
8.11: Diseases of folding and misfolding
protein
6.1: Defining the Cell’s Boundary
8.4: Protein synthesis- translation (RNA to polypeptide)
8.8: Turning polypeptides into proteins
8.11: Diseases of folding and misfolding
10.1: Targeting membrane proteins to where they need to be
protein activity
8.10: Allosteric and post-translational regulation
protein aggregation
17.3: Direct and indirect cellular responses to signaling molecules
protein assembly
10.2: Nuclear targeting and exclusion
protein carriers
6.3: Transport across membranes
protein degradation
8.9: Regulating protein activity, concentrations, and stability (half-life)
protein interactions
8.10: Allosteric and post-translational regulation
9.6: Types of regulatory interactions
protein localization
10.2: Nuclear targeting and exclusion
protein synthesis
7.11: Mutations, Deletions, Duplications and Repair
8.1: Specifying a polypeptide’s sequence
8.5: The translation (polypeptide synthesis) cycle
8.7: Getting more complex- gene regulation in eukaryotes
8.9: Regulating protein activity, concentrations, and stability (half-life)
8.10: Allosteric and post-translational regulation
17.2: Steady state and changing molecular concentrations- synthesis and degradation
protein transport
10.0: Introduction
proteins
3.23: Homologies provide evidence for a common ancestor
8.0: Introduction
8.9: Regulating protein activity, concentrations, and stability (half-life)
17.1: How do systems change at the molecular level?
proteolytic processing
8.9: Regulating protein activity, concentrations, and stability (half-life)
protocell
2.7: Experimental studies on the origins of life
pumps
6: Membrane Boundaries and Capturing Energy
6.5: Generating gradients- using coupled reactions and pumps
6.10: Using the energy stored in membrane gradients
Punnett square
15.4: Dihybrid crosses- linkage and recombination
purine
7.4: Unraveling Nucleic Acid Structure
pyrimidine
7.4: Unraveling Nucleic Acid Structure

Q

quail chick transplantation
19.4: Chick and Quail
quantitative trait
16.5: Genome-wide Association Studies (GWAS)
quantum physics
5.5: Inter- and intra-molecular interactions
6.6: Simple Phototrophs
quorum
18.5: Unicellular to multicellularity to unicellularity
quorum effects
18: Supplemental Chapter - Social interactions between cells
quorum sensing
4: Social Evolution and Sexual Selection
4.2: Quorum Sensing
18.0: Introduction
18.4: Transient and clonal ("true") metazoans
18.6: Evolutionary origins of clonal (permanent) multicellularity

R

radiation interactions
7.3: Identifying DNA as the genetic material
random events
3.12: Considering stochastic processes
reaction
5.3: Reaction rates
reaction rate
8.9: Regulating protein activity, concentrations, and stability (half-life)
reaction rates
5.3: Reaction rates
reactions
5.2: Coupling thermodynamically favorable and unfavorable reactions
5.4: Coupling Reactions
6: Membrane Boundaries and Capturing Energy
reasoning
1.2: Models, Hypotheses, and Theories
receptor proteins
10.3: Intercellular signaling- signals, receptors and responses
receptors
10.1: Targeting membrane proteins to where they need to be
10.4: Signaling molecules and receptors
recessive
15.1: How Mendel did what he did
15.2: Weldon's critique
15.4: Dihybrid crosses- linkage and recombination
15.11: Estimating the number of genes involved in a particular traits
15.13: Alleles, traits, and genetic diseases in humans.
recessive alleles
12.8: X-linked diseases and mono-allelic gene expression
15.6: Interacting traits- synthetic lethality and co-dominance
16.3: Using web-based bioinformatic tools- gnomAD
recessive mutation
13.1: Mutations into alleles
reciprocity
4.11: Curbing runaway selection
recombination
12: Asexual and Sexual Reproduction in Eukaryotes
12.0: Introduction
15: Mendel and Weldon - Contexts and their Effects on Phenotypes
redox reactions
6.7: Chemo-osmosis (an overview)
regulation
8.0: Introduction
regulatory interactions
9: Genomes, Genes and Regulatory networks
regulatory molecule
8.10: Allosteric and post-translational regulation
regulatory networks
9.0: Introduction
9.7: Final thoughts on (molecular) noise
17.0: Introduction
regulatory proteins
12.2: Monitoring cellular processes- mitosis
regulatory region
15.12: On the nature of mutations (again)
regulatory regions
9.2: Enhancers and transcription start and stop sites
regulatory sequences
9.1: Locating information within DNA
9.6: Types of regulatory interactions
religion
2.6: Thinking about life’s origins
repair
7: The Molecular Nature of Heredity
repeated sequences
7.6: DNA, sequences, and information
replication
5.0: Introduction
7.5: Discovering the structure of DNA
7.8: DNA Replication
7.10: Further replication complexities
13.0: Introduction
replication machinery
7.9: Replication machines
reproducibility
1.0: Introduction

reproduction
2.1: What is life, exactly?
2.3: The organization of organisms
12.4: Steps in meiosis- from diploid to haploid
reproductive barriers
12.3: Sex-determination and its chromosomal basis
reproductive compatibility
3.1: Organizing organisms (hierarchically)
reproductive interests
15.10: Imprinting- conflicts between mother, father, and fetus
reproductive isolation
3.21: Sympatric speciation
reproductive strategies
4.9: Sexual dimorphism
4.10: Sexual Selection
reproductive success
3.6: Darwin and Wallace's conceptual leap
3.10: Variation, selection, and speciation
3.11: Types of simple selection
3.13: Genetic diversity, population size, founder effects and population bottlenecks
3.16: Gene linkage- one more complication
3.17: A brief reflection on the complexity of phenotypic traits
3.18: Speciation and extinction
3.22: Signs of evolution- homology and convergence
4.4: Inclusive fitness, kin and group selection, and social evolution
4.5: Group selection
7.13: The origin of new (de novo) genes
15.8: Maternal and paternal effects
15.15: Measuring evolution’s impact on allele frequencies- Hardy-Weinberg
reproductive_isolation
3.20: Mechanisms of reproductive isolation
research methods
19.8: ESC and iPSC derived organoids
retinoblastoma
14.1: Rates and effects of somatic mutation
riboprotein
7.7: Discovering RNA- structure and some functions
ribose
7.4: Unraveling Nucleic Acid Structure
ribosome
7.7: Discovering RNA- structure and some functions
8.4: Protein synthesis- translation (RNA to polypeptide)
8.5: The translation (polypeptide synthesis) cycle
8.8: Turning polypeptides into proteins
ribosomes
8.7: Getting more complex- gene regulation in eukaryotes
10.1: Targeting membrane proteins to where they need to be
rigid cell wall
6.12: An evolutionary scenario for the origin of eukaryotic cells
RNA
3.23: Homologies provide evidence for a common ancestor
7: The Molecular Nature of Heredity
7.7: Discovering RNA- structure and some functions
7.10: Further replication complexities
7.11: Mutations, Deletions, Duplications and Repair
8.1: Specifying a polypeptide’s sequence
RNA polymerase
8.3: Protein synthesis- transcription (DNA to RNA)
9.1: Locating information within DNA
RNA primer
7.8: DNA Replication
RNA processing
10.2: Nuclear targeting and exclusion
RNA sequence
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
RNA sequencing
7.13: The origin of new (de novo) genes
16.2: Where is a gene expressed?
RNA synthesis
7.13: The origin of new (de novo) genes
10.0: Introduction
RNA transcription
10.2: Nuclear targeting and exclusion
Robert Brown
5.1: Approaching Thermodynamics
Rosalind Franklin
7.5: Discovering the structure of DNA
rotation
1.4: Teaching and Learning Science

S

S phase
12.1: Ploidy during the cell cycle
S. cerevisiae
9.6: Types of regulatory interactions
sampling
3.13: Genetic diversity, population size, founder effects and population bottlenecks
science
1: Understanding Science and Thinking Scientifically
1.0: Introduction
1.3: Science is Social
1.4: Teaching and Learning Science
2.6: Thinking about life’s origins
8.12: Molecular machines
scientific community
1.3: Science is Social
scientific discoveries
15.0: Introduction
scientific discussions
3.24: Anti-evolution arguments
scientific evidence
2.5: The Death of Vitalism
scientific experiments
2.4: Spontaneous generation and the origin of life
scientific explanations
1.2: Models, Hypotheses, and Theories
scientific ideas
1: Understanding Science and Thinking Scientifically
scientific inquiry
1.3: Science is Social
scientific knowledge
3.24: Anti-evolution arguments
scientific method
1.0: Introduction
1.2: Models, Hypotheses, and Theories
15.0: Introduction
scientific models
1.2: Models, Hypotheses, and Theories
scientific papers
1.3: Science is Social
scientific principles
1.5: Understanding Scientific Ideas
scientific reasoning
1.5: Understanding Scientific Ideas
scientific understanding
1.5: Understanding Scientific Ideas
Second Law of Thermodynamics
5.2: Coupling thermodynamically favorable and unfavorable reactions
sedimentary rocks
2.9: Fossil evidence for the history of life on earth
selection
3: Evolutionary Mechanisms and the Diversity of Life
3.11: Types of simple selection
13.0: Introduction
16.3: Using web-based bioinformatic tools- gnomAD
selection pressure
3.22: Signs of evolution- homology and convergence
selective advantage
3.14: Population bottlenecks
3.17: A brief reflection on the complexity of phenotypic traits
7.12: Genes and alleles
selective permeability
6.4: Channels and carriers
selective pressure
3.15: Genetic drift
selective pressures
3.21: Sympatric speciation
selective_pressure
3.20: Mechanisms of reproductive isolation
sex determination
12.3: Sex-determination and its chromosomal basis
12.7: X-inactivation and sex-linked traits
sexual behavior
4.9: Sexual dimorphism
sexual differentiation
4.9: Sexual dimorphism
sexual dimorphism
4.10: Sexual Selection
12.4: Steps in meiosis- from diploid to haploid
sexual reproduction
3.6: Darwin and Wallace's conceptual leap
3.15: Genetic drift
4: Social Evolution and Sexual Selection
4.0: Introduction
4.8: Origins and implications of sexual reproduction
12.0: Introduction
12.3: Sex-determination and its chromosomal basis
sexual selection
3.0: Introduction
4.9: Sexual dimorphism
4.10: Sexual Selection
4.11: Curbing runaway selection
15.10: Imprinting- conflicts between mother, father, and fetus
sexually dimorphic
15.10: Imprinting- conflicts between mother, father, and fetus
signaling
10.1: Targeting membrane proteins to where they need to be
17.4: Modeling gene expression
signaling molecules
4.2: Quorum Sensing
10.3: Intercellular signaling- signals, receptors and responses
17.1: How do systems change at the molecular level?

signaling pathways
10.4: Signaling molecules and receptors
signaling response
10.4: Signaling molecules and receptors
signaling systems
17.3: Direct and indirect cellular responses to signaling molecules
18.5: Unicellular to multicellularity to unicellularity
single cell
2.3: The organization of organisms
single cell RNA sequencing
16.2: Where is a gene expressed?
18.1: How do unicellular organisms generate phenotypic diversity?
single celled organisms
18.5: Unicellular to multicellularity to unicellularity
single nucleotide polymorphisms
16.5: Genome-wide Association Studies (GWAS)
single stranded DNA
7.6: DNA, sequences, and information
sliding clamp
7.9: Replication machines
SMAD
17.4: Modeling gene expression
small molecules
5.9: The implications of bond polarity
SNPs
20: Appendix- Muller's Morphs
social
1: Understanding Science and Thinking Scientifically
social behavior
4: Social Evolution and Sexual Selection
4.0: Introduction
4.1: Selecting social (cooperative) traits
4.4: Inclusive fitness, kin and group selection, and social evolution
4.5: Group selection
4.6: Defense against social cheaters
18.0: Introduction
18.3: Quorum effects (somewhat redundant)
social behaviors
18.2: Dying for others – social interactions between "unicellular" organisms
social cooperation
4.6: Defense against social cheaters
18.6: Evolutionary origins of clonal (permanent) multicellularity
social interactions
18: Supplemental Chapter - Social interactions between cells
18.4: Transient and clonal ("true") metazoans
social organisms
18.3: Quorum effects (somewhat redundant)
social slime molds
18.6: Evolutionary origins of clonal (permanent) multicellularity
soil nematode
19.6: The nematode Caenorhabditus elegans
solar system
1.4: Teaching and Learning Science
solid
5.9: The implications of bond polarity
somatic cell reprogramming
19.3: Frogs and fish
somatic cells
2.3: The organization of organisms
4.7: The appearance of multicellular organisms
somatic mutation
14.2: Non-disjunction- aberrant chromosome segregation
somatic mutations
14.1: Rates and effects of somatic mutation
specialization
2.3: The organization of organisms
specialized behaviors
18.4: Transient and clonal ("true") metazoans
speciation
3: Evolutionary Mechanisms and the Diversity of Life
3.18: Speciation and extinction
3.19: Mechanisms of speciation
3.22: Signs of evolution- homology and convergence
species
3.0: Introduction
3.1: Organizing organisms (hierarchically)
3.2: Natural and un-natural groups
3.14: Population bottlenecks
species origin
2.4: Spontaneous generation and the origin of life
sperm
15.9: Mitochondrial inheritance
splicing
15.12: On the nature of mutations (again)
spontaneous generation
2.4: Spontaneous generation and the origin of life
SRY gene
12.3: Sex-determination and its chromosomal basis
stability
6.2: The origin of biological membranes
17.2: Steady state and changing molecular concentrations- synthesis and degradation
stabilizing selection
3.11: Types of simple selection
statistical patterns
3.12: Considering stochastic processes
stochastic behaviors
9.7: Final thoughts on (molecular) noise
stochastic effects
9.3: Interaction networks and model systems
stochastic processes
3.12: Considering stochastic processes
18.1: How do unicellular organisms generate phenotypic diversity?
structural variants
14.3: Genome dynamics
structure
6.1: Defining the Cell's Boundary
6.2: The origin of biological membranes
8.8: Turning polypeptides into proteins
structures
8.0: Introduction
studies
13: How Alleles Arise - Mutations
study of life
2.0: Introduction
supernatural
2.6: Thinking about life's origins
sustainable population
3.5: Limits on populations
sympatric speciation
3.21: Sympatric speciation
symporter
6.10: Using the energy stored in membrane gradients
synonymous mutation
8.6: Effects of point mutations on polypeptides and proteins
15.12: On the nature of mutations (again)
synteny
16: Tools for Studying Genes and Genomes
synthesis
2.5: The Death of Vitalism
7.10: Further replication complexities
8.0: Introduction
8.4: Protein synthesis- translation (RNA to polypeptide)
8.8: Turning polypeptides into proteins
synthetic lethality
15.6: Interacting traits- synthetic lethality and co-dominance
synthetic phenotypes
15.6: Interacting traits- synthetic lethality and co-dominance

T

targeted mutagenesis
13.4: Generating specific mutations on demand - CRISPR-Cas9 and related technologies
taxonomic groups
3.2: Natural and un-natural groups
taxonomy
3.1: Organizing organisms (hierarchically)
teaching
1: Understanding Science and Thinking Scientifically
technology
13: How Alleles Arise - Mutations
temperature
5.7: Bond Stability and Thermal Motion (a non-biological moment)
5.9: The implications of bond polarity
territoriality
4.10: Sexual Selection
testable predictions
1.2: Models, Hypotheses, and Theories
TGFb family
17.4: Modeling gene expression
theories
1: Understanding Science and Thinking Scientifically
theory
3.24: Anti-evolution arguments
theory of evolution
1.3: Science is Social
3.4: Evolution theory's core concepts
thermal motion
5.7: Bond Stability and Thermal Motion (a non-biological moment)
thermodynamically unstable molecules
6.9: Chemotrophs

thermodynamics
2.1: What is life, exactly?
5.1: Approaching Thermodynamics
5.2: Coupling thermodynamically favorable and unfavorable reactions
5.4: Coupling Reactions
5.10: Interacting with Water
6.3: Transport across membranes
6.4: Channels and carriers
17.3: Direct and indirect cellular responses to signaling molecules
tissue formation
17.1: How do systems change at the molecular level?
tissue removal
19.4: Chick and Quail
totipotent cells
10.5: Cellular reprogramming- embryonic and induced pluripotent stem cells
Toxins
10: Social Systems
trait
15.11: Estimating the number of genes involved in a particular traits
trait comparison
3.22: Signs of evolution- homology and convergence
traits
3.1: Organizing organisms (hierarchically)
3.2: Natural and un-natural groups
3.9: Genotype-phenotype relationships- discrete and continuous traits
3.11: Types of simple selection
3.16: Gene linkage- one more complication
3.23: Homologies provide evidence for a common ancestor
4.1: Selecting social (cooperative) traits
7.0: Introduction
7.12: Genes and alleles
15.1: How Mendel did what he did
15.2: Weldon's critique
15.14: Concordance between monozygotic twins and genetic influence on a trait
transcription
7.11: Mutations, Deletions, Duplications and Repair
8.1: Specifying a polypeptide's sequence
8.3: Protein synthesis- transcription (DNA to RNA)
8.5: The translation (polypeptide synthesis) cycle
8.6: Effects of point mutations on polypeptides and proteins
8.7: Getting more complex- gene regulation in eukaryotes
9: Genomes, Genes and Regulatory networks
9.1: Locating information within DNA
17.2: Steady state and changing molecular concentrations- synthesis and degradation
17.4: Modeling gene expression
transcription factor
10.2: Nuclear targeting and exclusion
transcription factors
8.3: Protein synthesis- transcription (DNA to RNA)
8.9: Regulating protein activity, concentrations, and stability (half-life)
9.0: Introduction
9.1: Locating information within DNA
9.2: Enhancers and transcription start and stop sites
9.6: Types of regulatory interactions
9.7: Final thoughts on (molecular) noise
10.4: Signaling molecules and receptors
17.3: Direct and indirect cellular responses to signaling molecules
17.5: Reversible, irreversible, and cascade effects
transcription regulation
9.1: Locating information within DNA
transcriptional activator
9.2: Enhancers and transcription start and stop sites
transcriptional bursting
9.7: Final thoughts on (molecular) noise
transcriptional repressor
9.2: Enhancers and transcription start and stop sites
transduction
11.2: Other naturally occurring horizontal gene transfer mechanisms
transfer RNA
7.7: Discovering RNA- structure and some functions
transformation
11.2: Other naturally occurring horizontal gene transfer mechanisms
transgenic
19.4: Chick and Quail
transient metazoans
18: Supplemental Chapter - Social interactions between cells
transient multicellular
18.4: Transient and clonal ("true") metazoans
transition state
5.3: Reaction rates
translation
7.11: Mutations, Deletions, Duplications and Repair
8.1: Specifying a polypeptide's sequence
8.4: Protein synthesis- translation (RNA to polypeptide)
8.5: The translation (polypeptide synthesis) cycle
8.6: Effects of point mutations on polypeptides and proteins
8.7: Getting more complex- gene regulation in eukaryotes
8.8: Turning polypeptides into proteins
10.1: Targeting membrane proteins to where they need to be
17.2: Steady state and changing molecular concentrations- synthesis and degradation
translational bursting
9.7: Final thoughts on (molecular) noise
transplantation studies
19.4: Chick and Quail
transport
6: Membrane Boundaries and Capturing Energy
6.5: Generating gradients- using coupled reactions and pumps
transport mechanisms
6.4: Channels and carriers
transporters
6.10: Using the energy stored in membrane gradients
transposable elements
14.3: Genome dynamics
14.6: Transposons- moving DNA within a genome (and weird genetics)
transposons
14.6: Transposons- moving DNA within a genome (and weird genetics)
trinucleotide repeat
15.16: Genetic anticipation
trinucleotide repeat disorders
7.14: DNA repeat diseases and genetic anticipation
tRNA
7.7: Discovering RNA- structure and some functions
8.4: Protein synthesis- translation (RNA to polypeptide)
tumor
14.0: Introduction
tumor formation
14.1: Rates and effects of somatic mutation
twins
15.14: Concordance between monozygotic twins and genetic influence on a trait

U

unicellular
18: Supplemental Chapter - Social interactions between cells
18.4: Transient and clonal ("true") metazoans
18.6: Evolutionary origins of clonal (permanent) multicellularity
unicellular algae
4.7: The appearance of multicellular organisms
unicellular cooperation
18.0: Introduction
unicellular organisms
4.0: Introduction
4.2: Quorum Sensing
7.2: Locating hereditary material within the cell
18.1: How do unicellular organisms generate phenotypic diversity?
unique sequences
7.6: DNA, sequences, and information
universe
1.4: Teaching and Learning Science

V

van der Waals interactions
5.6: Covalent bonds
vertebrate development
19.4: Chick and Quail
Vesicle
6.1: Defining the Cell's Boundary
Vibrio fischeri
4.2: Quorum Sensing
video
8.12: Molecular machines
viral infection
4.3: Active (altruistic) cell death
18.2: Dying for others – social interactions between "unicellular" organisms
vitalism
1.1: The interconnectedness (self-consistency) of science
2.5: The Death of Vitalism

W

wallace
3.6: Darwin and Wallace's conceptual leap
water concentration gradient
6.11: Osmosis and living with and without a cell wall
water influx
6.11: Osmosis and living with and without a cell wall
Watson and Crick
7.5: Discovering the structure of DNA

X

X chromosome
12.3: Sex-determination and its chromosomal basis
12.7: X-inactivation and sex-linked traits
12.8: X-linked diseases and mono-allelic gene expression
Xenopus laevis
19.3: Frogs and fish

Y

Y chromosome

12.3: Sex-determination and its chromosomal basis

Z

zygote

15.9: Mitochondrial inheritance

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Your page has been created!

Remove this content and add your own.

- Edit page

  Click the Edit page button in your user bar. You will see a suggested structure for your content. Add your content and hit Save.

Tips:

- Drag and drop

  Drag one or more image files from your computer and drop them onto your browser window to add them to your page.

- Classifications

  Tags are used to link pages to one another along common themes. Tags are also used as markers for the dynamic organization of content in the CXone Expert framework.

- Working with templates

  CXone Expert templates help guide and organize your documentation, making it flow easier and more uniformly. Edit existing templates or create your own.

- Visit for all help topics.