Genomes and Genomics: the instruction plan for life on our planet

Mom, I have brown eyes like you and Linnea is tall like Dad because we carry your genes. Can any of these genes help us cure diseases or clean up the environment?

Photographic composition where half a portrait of a girl is shown on the left and half a portrait of her biological mother is shown on the right. The resemblance between them is evident by genetic inheritance. Courtesy of iStock.com/Batirtze Arrien Crespo.

Robert M. Bowers and Tanja Woyke

Department of Energy Joint Genome Institute, Berkeley, CA, USA

A child-centric microbiology education framework

Storyline

The diversity of life on planet Earth is immense. It ranges from the smallest bacterium only visible with a microscope, to single-celled organisms that can be seen with the naked eye, to multicellular organisms, such as plants and animals, including us humans. Life has evolved across nearly all of Earth's diverse landscapes over billions of years, displaying diverse metabolic strategies and stunning morphologies. Furthermore, some organisms inhabit environments we perceive as extreme, including acidic hot springs, the subseafloor, and even the atmosphere. With our own eyes, we can directly observe the astonishing diversity of the macroscopic world that surrounds us, from tropical rain forests to coral reefs. However, the overwhelming extent of biological diversity resides within the world of single cells, a world that we knew little about until the development of the microscope, and more recently, the decoding of microbial genomes.

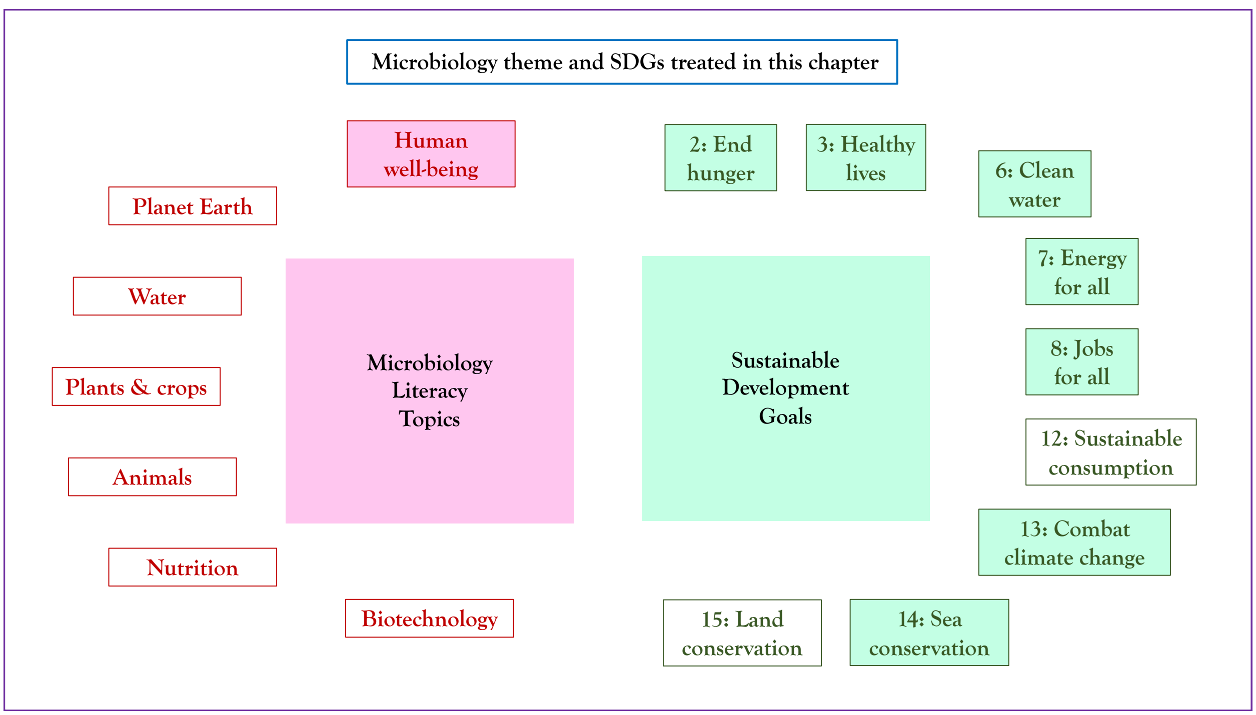
Panoramic view of the Grand Prismatic Spring, a hot spring in Yellowstone National Park, Wyoming, USA, where microbial life can thrive under the most extreme environments. Genomes of these microorganisms hold clues as to how they adapted to life in these extremes. Courtesy of iStock.com/ VirtualVV.

We can begin to make sense of this vast diversity by studying the instruction plan of life, which lies in the genomes of organisms. Once genomes are decoded, their information can be related to the traits of the organisms and the environments that the organisms inhabit. The field of genomics interprets this instruction plan of an organism to understand how and why cells differ from one another, and how they make a living. Genomics provides us with a blueprint for how an organism grows, develops, and finds food or avoids death. For example, think about a caterpillar that becomes a butterfly. Inside the caterpillar, the genome instructs each cell on how to change and collectively reorganize themselves to make a butterfly. The genome also has instructions for how the butterfly will fly, find its food and its mate, and how to migrate. Scientists read genomes as though they are instruction manuals for living organisms, to help study the way organisms work, how they came to look the way they do, and how they are related to one another. Being related means that living organisms share parts of their instruction plans (a.k.a. genomes). Over the short term (20-40 years for humans), parts of these instruction plans get handed down to the next generation. For example, you have some of your grandparents' and parents' instruction plans, or genomic parts, and therefore have some of their traits; maybe you have your Grandma's and Mom's brown eyes. Over the long term (billions of years), all living organisms are related to each other through evolutionary time. Even humans and bacteria share some very similar instructions.

On a much broader level, scientists use genomics to address issues related to human health, ponder the evolution of life, and tackle some of the largest environmental and economic challenges currently impacting our planet, including clean energy, climate change, and food security. Genomics is perhaps the most useful and widely used methodology in the biomedical sciences. It can be used to detect pathogens that cause illness, to develop cancer therapies, and can even be used to develop personalized nutrition. Genomics can help us in all these areas, by providing us with the blueprint of life.

The Microbiology and Societal Context

The microbiology: Genomes-genomics; gene expression; Human Genome Project; CRISPR; COVID-19; bioinformatics. *Sustainability issues:* plant stress tolerance; infection prevention and therapy; water purification, wastewater treatment, bioremediation; bioenergy; bioeconomy and employment creation; carbon sinks; eutrophication.



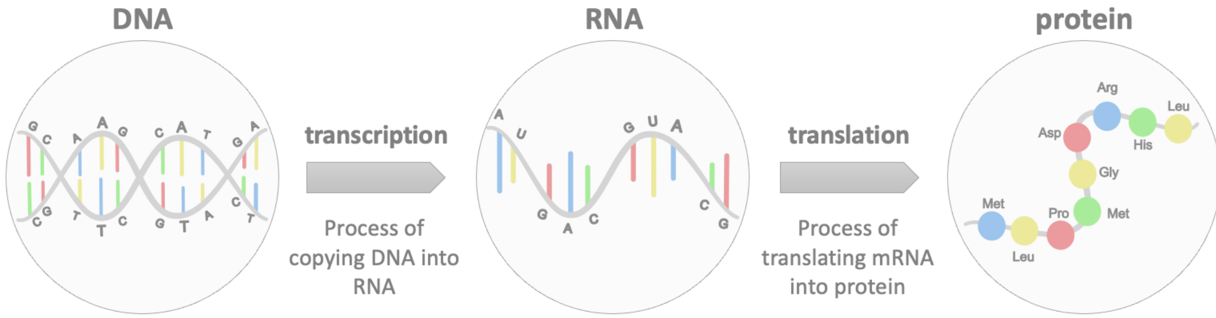
Genomes and Genomics: the instruction plan for life on our planet

1. ***What is genomics?*** **Genomics** is a field of biology that studies an organism's genes, collectively called the organism's **genome**. Genomics is used to address human health related issues, to study our own evolutionary history, the evolutionary history of life itself, and to tackle some of the largest challenges currently impacting our planet including clean energy and food security. However, before jumping into the applications of genome sequencing, let us first review some of the key concepts and terminology in molecular biology to see why genomics is relevant to our everyday lives.

2. ***How is a genome defined?*** A **genome** can be defined as the complete genetic makeup of an organism, which includes all the genes. All organisms have a genome, and the genome is composed of all cellular DNA (deoxyribonucleic acid) that is passed from one generation to the next. Genomes are often described as the genetic blueprint of a cell. Different regions of the genome can be used for distinct purposes and can even become activated during different phases of development. However, genomes should not be thought of as static, as the word *blueprint* might imply. The structure of the genome is highly dynamic. Some regions are tightly packed and inaccessible while others are unwound and ready for use. Some regions of the genome are perpetually active, containing genes with housekeeping functions, while other regions remain relatively quiet, utilized on a more as-needed basis.

3. ***How do the genes fit into the genome?*** Individual **genes** represent certain subsets of the genome. They are composed of strings of DNA bases, also known as deoxyribonucleotides, of a certain length. DNA sequences are represented by combinations of the letters A, G, T and C and can be viewed on a sheet of printed paper, or more commonly on a computer screen. One of the smallest known genes, coding for an antibiotic protein (to fight off other bacteria) in *Escherichia coli* is so small that it fits on half a line in this document: ATGCGTACTGGTAATGCAAACTAA, only 24 bases long! However, the full genome of an *E. coli* cell would fill roughly 250 pages of a book, and the human genome, nearly 250,000 pages!

4. ***How does the cell make something useful out of the DNA?*** The **central dogma** of molecular biology states that genetic information flows in one direction, from DNA to ribonucleic acid (RNA) to protein. The genome is essentially the starting point for the flow of information from the genome blueprint to cellular activities. The blueprint becomes activated when a gene from the genome is transcribed into an RNA molecule, which is then translated into protein. Proteins are often described as the 'workhorse' of life, as protein molecules provide the structure and function necessary for the daily activities of a cell, including the metabolism of food, production of waste, movement, etc.



Schematic depiction of the central dogma of molecular biology. Modified from iStock.com/ Rungnaree Jaitham.

5. ***What have we learned from peering into our own genome?*** The current state of genomics is the direct result of a massive effort undertaken by an international consortium of scientists, spanning academics, industry and government, to sequence the human genome. **The Human Genome Project** (HGP) sparked the invention and development of whole genome sequencing technologies that continue to advance to this day. We have learned much since the human genome was first sequenced 20 years ago. We now know that our 3 billion base pair human genome is ~99.9% identical between all humans. This means that ~0.1% (~ 300 million base pairs) of the genome typically differs between individual humans; these differences have provided clues as to how and when our ancestors migrated out of Africa. We also learned more about what genes and gene mutations cause specific genetic disorders such as schizophrenia, Huntington's disease, obesity, and certain cancers. Through genome sequencing, scientists can now probe deeper into our evolutionary past, and health-related applications are currently transitioning to the clinical setting where the sequencing of individual human genomes now costs as little as $100 (compared to ~$3 billion for the original HGP). Affordable genomes will likely help us to diagnose and treat a myriad of human diseases more rapidly and help prescribe medicines that work best in the context of each person's genetic makeup, so-called *precision* or *personalised medicine*. While we have learned a lot since the release of the human genome sequence, there is much that we can still learn, and with the sequencing of more human genomes, there are ample opportunities for students and young scientists for decades to come.

6. ***What about microbial genomes? Why should we care about the life that we cannot see?*** Sequencing of the human genome had a very large impact on the field of microbial genomics, as the HGP played a critical role in fast-tracking the development of advanced, high throughput sequencing technologies that can be readily applied to any organism with DNA, i.e., all life as we currently know it. Moreover, microbes have been relevant to medicine for over 150 years, ever since Pasteur linked microbes to disease (a.k.a. germ theory), and more recently through studying the human microbiome. Plus, their utility to industry for fermentation (beer, wine, cheese, etc.) and for the production of antibiotics and other pharmaceuticals is well known. Importantly, scientists have recognized that microorganisms are key to all sorts of environmental processes, especially given that they are such a dominant component of the overall biosphere. As such, microbial genomes provide

a valuable source of information that can be leveraged to address not only human but also environmental challenges.

7. ***I've heard a lot about CRISPR in the news. What is it, and how will it help scientists address human and environmental issues?*** CRISPR is a much easier way to say **clustered regularly interspaced short palindromic repeats**, which is a mouthful. The reason this acronym has become so important to scientists and society alike is because the application of the CRISPR system to biotechnology in the form of gene editing has rapidly become one of the most important scientific breakthroughs in recent history. However CRISPRs did not evolve to be gene editors; they are actually an immune system for microbes that detects and destroys invading DNA that usually arrives in the form of viruses that infect the microbes. CRISPR is one of the tools that microbes use to fight the viruses that are trying to infect them. The CRISPR system is encoded within microbial genomes where short sequences of DNA that correspond to previously infecting viruses are stored. When the microbe comes into contact with the same virus again, the DNA that matches previously infecting viruses is expressed into RNA, then together with the CRISPR associated proteins, is guided to the viral target, where the virus is then degraded (i.e., killed), thus providing immunity against the invading virus. Beyond a microbial defense system, as mentioned above, CRISPR is used by scientists as a gene-editing tool, and makes the previously difficult and laborious task of modifying a gene for biotechnology purposes far easier. However, given the power that this new technology has, the ethics of its application will also be something to pay attention to.

8. ***What about COVID-19? Does genomics play a role in keeping us safe from human pathogens?*** Absolutely! Sequencing the DNA of human pathogens, or the RNA in the case of RNA viruses, is rapidly becoming one of the most important methods of pathogen detection. Pathogen genome sequencing is used for pathogen surveillance and outbreak monitoring, as well as for the development of vaccines. Genome sequencing of SARS-CoV-2, the virus that causes COVID-19, is helping scientists to understand how the virus is spreading and changing over time. The field of pathogen detection has expanded from cultivation-based approaches to using the much faster **PCR** (polymerase chain reaction) technique, to most recently, making use of direct whole genome sequencing without cultivation, a.k.a., metagenomics. The application of whole genome sequencing to pathogen detection has facilitated more detailed epidemiological studies, providing a way to study pathogen evolution in real time. Combining pathogen genomics with epidemiology for diseases such as tuberculosis, for example, has greatly facilitated public health efforts to minimize disease transmission and spread.

9. ***Is it possible for me to analyze the genomes of the organisms that I'm interested in?*** Of course! Efforts have been made to make much of the genome sequence data that have been obtained over the last 20 plus years available to the public. For example, the website for the National Center for Biotechnology Information (NCBI: https://www.ncbi.nlm.nih.gov) is a wonderful open resource and starting point, as genomes and genes can be explored with a click of the mouse. Moreover, as computer-based skills become readily incorporated into high school curricula, they can be directly applied to the exploration of genomes. For example, basic programming skills can be used to download and investigate the DNA sequences of many genomes at a time using the power of laptop

computers. For this reason, we hope that students and teachers alike will begin to explore the fascinating world of life through genomics, especially as more opportunities become available.

Relevance for Sustainable Development Goals and Grand Challenges

Genomes and genomics relate to several SDGs, including

- **Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture** (*end hunger and malnutrition, increase agricultural productivity).* Understanding the genomes of plants, specifically food crops, the microorganisms associated with them, as well as microbes in the surrounding soil, will enhance agricultural productivity and sustainability. For example, we can decipher, through genomics, how plants can tolerate stress such as drought or nutrient limitation, or how microbes can help the plant access certain soil nutrients and promote plant growth.
- **Goal 3. Ensure healthy lives and promote well-being for all at all ages** (i*mprove health, reduce preventable disease and premature deaths).* In addition to contributing to agricultural sustainability and productivity, genomics has been key to our understanding of human health. The Human Genome Project, the decoding of the human genome, was one of the greatest scientific endeavors in history. Deciphering the human instruction plan has provided most valuable insights into human diseases and has greatly improved medical practices and human health overall. Moreover, studying the genomes of the microorganisms associated with humans (a.k.a. the human microbiome, which includes pathogens), has further accelerated our understanding of human health and disease, and greatly improved disease treatment options.
- **Goal 6. Ensure availability and sustainable management of water and sanitation for all** (a*ssure safe drinking water, improve water quality, reduce pollution, protect water-related ecosystems, improve water and sanitation management).* Microbes in sewage plants help treat and purify water, and by studying their genomes, scientists can make microbial contributions to water purification more efficient. Furthermore, some microbes degrade hydrocarbons (i.e., oil). Understanding which microbes can do this, and how, can facilitate the use of microbes to clean spills (a.k.a. bioremediation), following events such as the deep horizon oil spill in the Gulf of Mexico. Understanding the function of these unique microbes often begins with decoding and interpreting their genomes.
- **Goal 7. Ensure access to affordable, reliable, sustainable and modern energy for all** (e*nsure access to clean, renewable and sustainable energy, and increase energy use efficiency).* Sustainability requires the increasing use of renewable resources. Biofuels are fuels derived directly from biomass as opposed to fossil fuels, such as oil. Alternatives to fossil fuels are needed as the formation of fossil fuels typically takes millions of years. As a result, fossil fuels are limited in supply, and their extraction and use negatively impacts air quality and climate change. While biofuels are thought to be a more sustainable alternative, they are nevertheless difficult to produce. Scientists continually work on increasing the efficiency of biofuel production. For example, they study microorganisms (fungi, bacteria) that most efficiently break down complex plant biomass, while also genetically modifying the plant to be more easily broken down by the microbes. The field of genomics plays an important role in all these efforts,

allowing us to learn which microorganisms break down the plant material and how, and by modifying the instruction plan of the plant to make the job easier for the microbes. This makes genomics an important tool in our search for alternative and more sustainable forms of energy.

- **Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all** (p*romote economic growth, productivity and innovation, enterprise and employment creation).* Genomics is a tool that has been applied across a broad range of sectors; it is critical in biotechnology, parts of the energy sector, agriculture and importantly the health sector, including the pharmaceutical industry. Government labs employ genomics as a tool for basic research, the results of which often find applications in the industrial sectors.
- **Goal 13. Take urgent action to combat climate change and its impacts** *(reduce greenhouse gas emissions, mitigate consequences of global warming, develop early warning systems for global warming consequences, improve education about greenhouse gas production and global warming).* An important step towards combating climate change is to understand how organisms and whole ecosystems respond to rising temperatures. This is important for identifying resilient crops as well as for identifying microbial consortia that either enhance or perturb the resiliency of important agricultural crops. Moreover, we are just beginning to uncover the vast microbial diversity residing in the soils beneath our feet. Provided that a tremendous amount of carbon is cycled through soils each year, much of it by the microbes, perhaps the microbes can teach us how to best keep the carbon sequestered as opposed to releasing it to the atmosphere in the form of harmful greenhouse gases. Currently, there is much effort going into the sequencing of microbial genomes extracted directly from the environment via metagenomics. These collective genomes can provide metabolic snapshots of whole communities, which can subsequently be used in climate models to better understand the effects of particular changes to our environment.
- **When Goal 14. Conserve and sustainably use the oceans, seas and marine resources for sustainable development** *(reduce pollution of marine systems by toxic chemicals/agricultural nutrients/wastes like plastics, develop mitigation measures for acidification, enhance sustainable use of oceans and their resources).* Excess nitrogen and phosphorus inputs into our oceans, as a result of current agricultural practices, are known to significantly alter the marine ecosystem. For example, algal blooms can occur if anthropogenic nitrogen and phosphorus inputs are not controlled, which can lead to dead zones when the algal biomass is degraded by bacteria that, in the process, consume all available dissolved oxygen required for a healthy marine ecosystem. One possible solution to this important problem is to apply less fertilizer while, at the same time, increasing crop production. Genomics is important for both the understanding of the crop plants themselves and to better understand the microbial consortia that can act as beneficial symbionts to the agricultural crops. Moreover, researchers are actively studying the genes of microbes that are responsible for the degradation of plastics, with the goal of creating a sustainable approach to clean up some of our polluting waste. Such bioremediation holds great promise in removing these harmful polymers from aquatic and other ecosystems.

Potential Implications for Decisions

1. Individual

a. Hygiene and disease: Future applications of genomics for monitoring human health and the spread of disease.

b. Composting & gardening: A genomics approach to better understand the symbiotic relationships between plant growth promoting microbes and their host plants.

2. Community policies

a. Public health: Genomics-informed healthcare monitoring for personalized medicine.

b. Water/sewage treatment: Using genomics to increase sewage treatment efficiencies.

3. National policies

a. Genome-informed healthcare practices: Development of monitoring systems using genomics within clinical settings.

b. Agricultural monitoring using genomic profiling of host plants, their commensals, and pathogens.

c. Integration of genomics facilitated metabolic profiles into climate models.

Pupil Participation

1. Class discussions

a. Can you think of key events in human history where genomics played a large role?

b. What are some of the main implications of genomics on human health?

c. Can you think of examples of the use of genomics in our everyday lives?

d. Are there any future applications of genomics you could envision?

e. Are these any areas not discussed on class where you think genomics might be useful and how (hint: space science)?

2. Pupil stakeholder awareness & Exercises

a. Genomics is relevant to distantly related fields, from human health to climate change. Can you describe how the knowledge of a genome sequence may be important to your own life, and to those of future generations?

b. Can you describe how a genome sequence might be used to help fight climate change, or help clean up the environment?

c. Imagine a new plant pathogen emerges that destroys certain food crop plant, posing a threat to our food supply. How might genomics help in fighting this pathogen?

d. What additional types of information might add context to a given genome sequence?

The Evidence Base, Further Reading and Teaching Aids

Podcasts: The Genome Insider and Natural Prodcast
NIH Brief Guide to Genomics
Book chapter: Del Giacco, L., Cattaneo, C. (2012). Introduction to Genomics. In: Espina, V., Liotta, L. (eds) Molecular Profiling. Methods in Molecular Biology, vol 823. Humana Press. https://doi.org/10.1007/978-1-60327-216-2_6

Glossary

Biotechnology: Utilization of biological process for industrial purposes.
Biodegradation: Breakdown of organic matter by microorganisms.
Biofuel: Fuels produced from renewable resources, including plants and algae.
Bioremediation: Making use of microbes to degrade environmental pollutants.
Commensals: Relationship among organisms where one gains a benefit, while the other remains neutral; neither harmed nor benefits from the relationship.
COVID-19: A human respiratory disease caused by the SARS-CoV-2 virus.
CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats, which are part of a bacterial defense system that forms the basis for CRISPR-Cas9 genome editing.
DNA (deoxyribonucleic acid): The hereditary material in humans and almost all other organisms.
Fermentation: Extraction of energy in the absence of oxygen.
Genomics: A sub-field of molecular biology concerned with the structure, function and evolution of genomes.
Genome(s): The complete set of genetic information in an organism.
Gene: Considered the basic unit of inheritance.
Gene editing: A method that enables scientist to alter the DNA of many organisms.
Germ theory: States that microbial pathogens or 'germs' can cause disease.
Greenhouse gas: A gas contributing to the greenhouse effect by absorbing infrared radiation, i.e. CO2, CFCs and methane.
Human microbiome: The aggregate of all microbes residing within and on the surface of people.
Hydrocarbons: Any of a class of organic compounds consisting only of carbon and hydrogen.
Metagenomics: The study of genetic material recovered directly from the environment.
Molecular biology: Branch of biology dealing with the structure and function of the molecules necessary for life, i.e. nucleic acids and proteins.
Pathogens: A microorganism that can cause disease.
PCR: Polymerase Chain Reaction, a technique used to amplify DNA.
Polymer: A substance composed of many repeating subunits.
Protein: Large, complex molecules that play many critical roles in an organism.
RNA (ribonucleic acid): A nucleic acid residing in living cells, having many biological roles including coding, decoding, gene regulation and gene expression.
SARS-Cov-2: The causative agent of COVID-19, and a member of a family of viruses called the coronaviruses.
Sequencing: The acquisition of the exact sequences of nucleotides or bases in a DNA molecule.