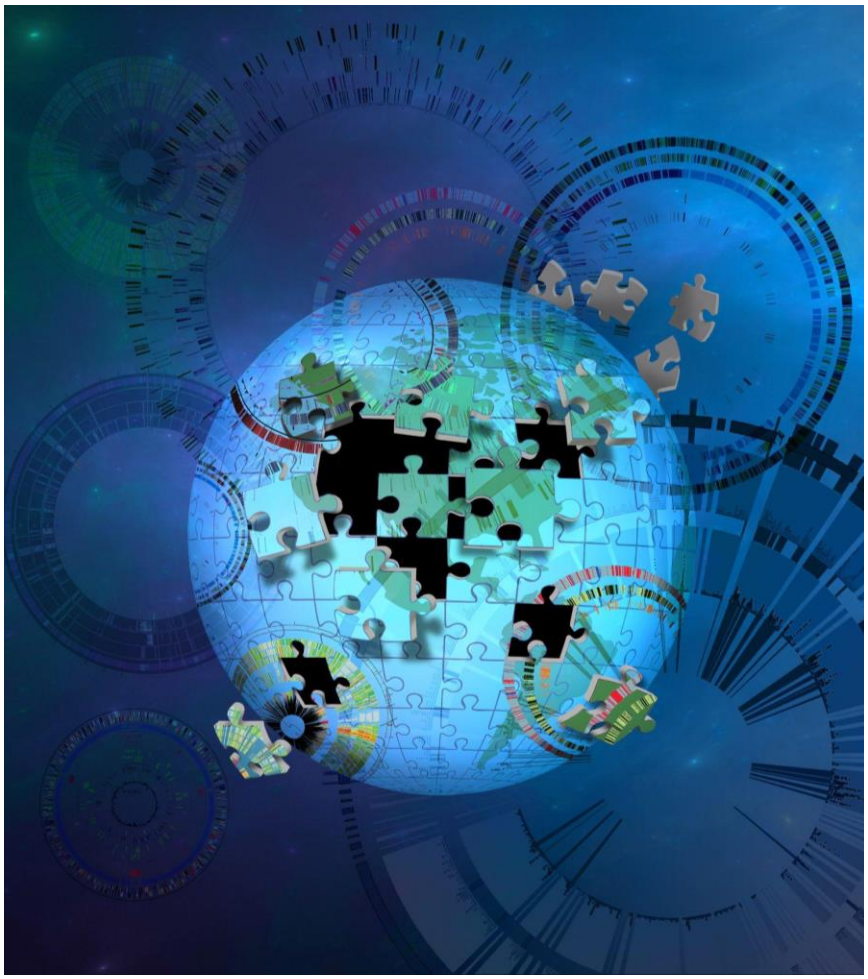
A child-centric microbiology education framework

The Global Microbial Metagenome

(An invisible genomic puzzle across the Earth)

I just love puzzles….
but how do I put one together if I can't see the pieces?



Matthew Kellom & Emiley Eloe-Fadrosh

DOE Joint Genome Institute, Berkeley, CA, USA

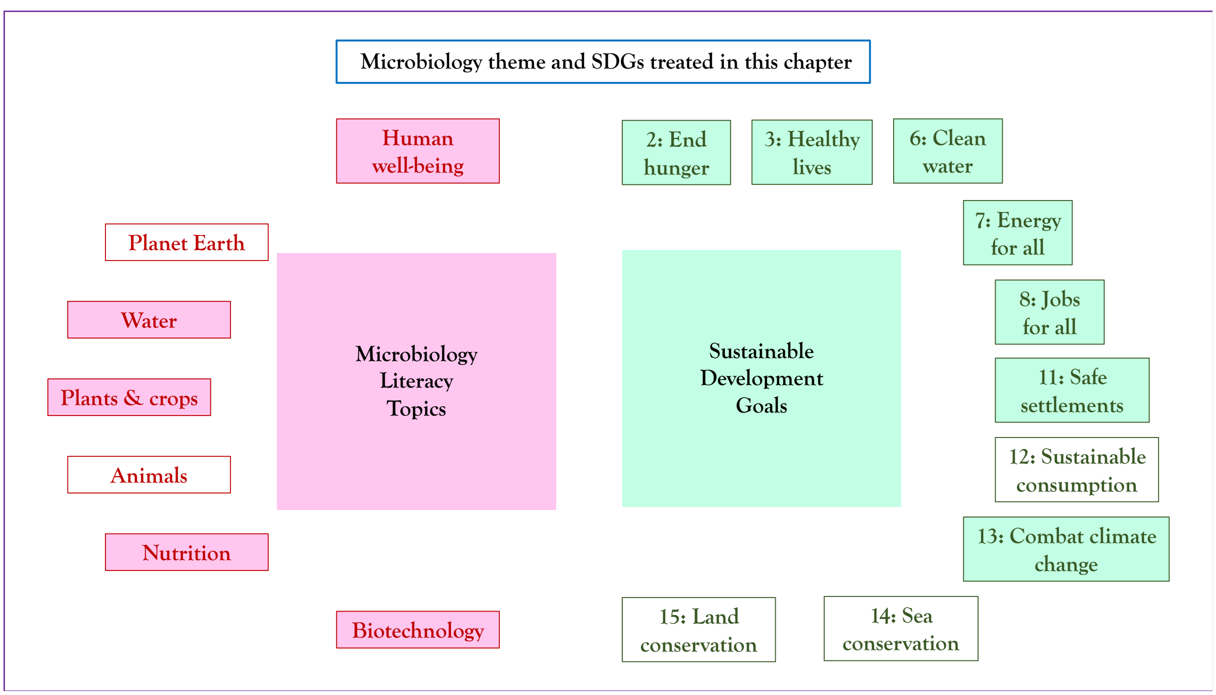
The Global Microbial Metagenome

Storyline

Microbes greatly impact our everyday lives and yet we rarely have an opportunity to observe them around us. Incredibly, the number of diverse microbes living on our planet is estimated to be larger than the number of stars in the universe. Each microbe has its own unique genome encoding genes for all functions needed to eat and grow. One way to better understand these microbes is to use metagenomics, which is the study of genomic material collected directly from organisms in an environment. In this microbiology educational chapter, we outline the advantages that studying the global microbial metagenome has on our society, including advances in medicine, energy production, pollution reduction, and agriculture. Understanding Earth's microbial metagenomes has multiple consequences for *Sustainable Development Goals.*

The Microbiology and Societal Context

The microbiology: metagenomics; microbiomes. *Sustainability issues: health*: understanding healthy and unhealthy human microbiomes, discoveries in treatment and preventative medicine; *energy*: renewable biofuel production; *environment*: carbon capture technologies to reduce carbon dioxide emissions, pollution reduction; *agriculture*: improving crop production.



The Global Microbial Metagenome: the Microbiology

1. ***Microbes are everywhere, but what are they doing?*** Microbes are often thought of as germs or parasites, but the vast majority of microbes do not cause disease. Most microbes live in a healthy relationship with their particular environment, which includes inside and outside of the human body, the waters of boiling hot springs, under freezing glaciers, deep within the Earth's subsurface, and even floating amongst the clouds. But how do we know which microbes are in which environments and what they are doing there?

2. ***First: collect your sample.*** The first step in understanding microbial life of any environment is sample collection. Whether it's a handful of dirt or a glass of pond water, a sample of our world will be teeming with microbes. This is often a very exciting (sometimes the **most** exciting) part of microbiology, because we are often interested in exotic environments, like the polar regions, tropical rain forests, hot springs, saline lakes, the deep sea, space, etc. So, collecting samples not only involves serious logistical planning (more, usually much more, than we need to do when we go on holiday), but also making trips to exciting, sometimes difficult to access sites. This can make sampling an adventure of discovery (e.g. imagine diving in a tiny submersible to sample a black smoker 7 km under the sea!).

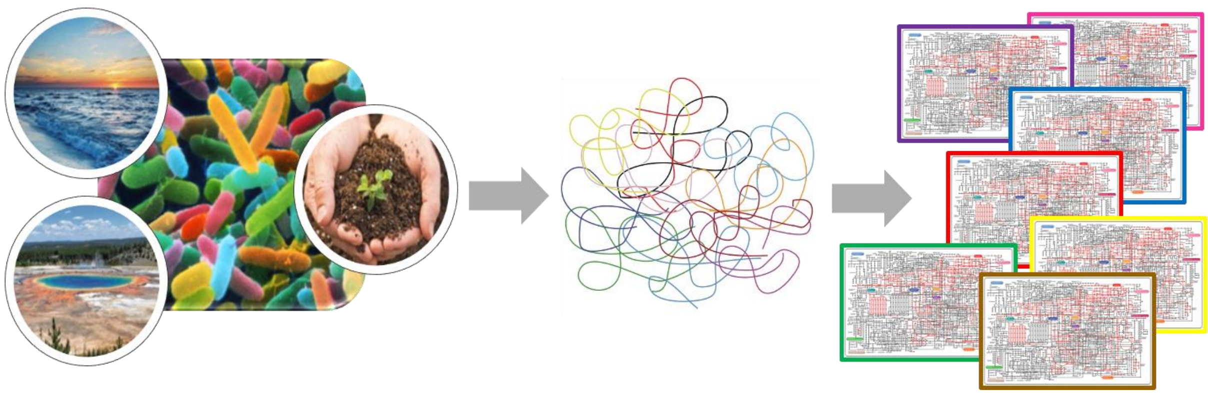
Soil sampling at the Blodgett Forest Research Station in Northern California. Credit: Roy Kaltschmidt, Berkeley Lab.

3. ***Now: let's look at the microbes in the sample.*** Individually, microbes are too small to see with the naked eye but you may notice them when conditions are right for them to grow, multiply and form large colonies. A classical method of identifying microbes is to grow (cultivate, like we cultivate plants in the garden) these colonies and perform tests in the laboratory to characterize them based on how they grow.

While in the past, cultivation was the only method available to obtain microbes for study, it recently became clear that many microbes require very specific and thus far undiscovered conditions to grow in a laboratory environment. We have only been able to cultivate a tiny fraction of microbes from the environment, perhaps only 1%, so we were missing the 'silent majority', and hence obtaining a highly distorted picture of which microbes are in a sample and what they are doing. So how can we capture 100%, or close to it?

4. ***Time to look at the microbial genomes in the sample.*** Another method to identify microbes is called **metagenomics**, which is the study of genomic information from all the microbes living in our samples. We collect this genomic information by extracting **DNA (deoxyribonucleic acid)** from the cells in our samples and sequencing it: gathering the information encoded by the genomes. However, this extracted DNA represents a mix of all the organisms within a sample, so it is a challenge to identify who is who.

To make matters worse, the limits of current **sequencing technology** doesn't allow us to obtain full **genomes** directly from sequencing machines. Instead, we obtain millions, or potentially billions, of relatively small genome fragments from this mixed community. Piecing together all of the small genomic fragments and studying the metagenome of the sample is like putting together an invisible genomic puzzle.



Metagenomics is the study of genetic material collected directly from organisms in an environment. This genetic material (DNA) is a mix of all the diverse microbial genomes represented by the different colored squiggles in the middle. The challenge with current sequencing technologies is to put these pieces together, just like a puzzle, to understand which microbes are from that environment and what functions they perform (e.g., eat sugars, photosynthesize, or even breathe metal).

5. ***Genome assembly: time for the computer.*** The global metagenomic puzzle is huge, complex, and to a great extent unknown. Unlike normal puzzles, we cannot put the pieces together by eye. The use of metagenomics relies heavily on computers and software to sort the data produced by sequencing machines, assemble sequences of small genome fragments into larger fragments, through identifying sequence overlaps, and ultimately into complete genomes. But this is on an individual sample, so we need to repeat with millions of samples to get an idea of the global microbiome. On the other hand, some environments, like marine systems and the atmosphere, are highly connected, which means that some microbes are widely distributed, so a few samples are able to capture the more abundant microbes that live in these major global habitats.

Advanced sequencing machines and computers are crucial tools that make metagenomics possible. Genetic material (DNA or RNA) must be extracted from the sample and properly prepared before it can be loaded into the sequencing machine. Pictured above, a team of researchers are loading a prepared DNA sample into a sequencer at the Department of Energy Joint Genome Institute (JGI). Credit: Thor Swift, Berkeley Lab.

6. ***There are a lot of microbes (microbial genomes) on the planet.*** The number of diverse microbes living on our planet is estimated to be larger than the number of stars in the universe. Each microbe has its own unique genome encoding **genes** for all functions needed to eat and grow.

7. ***What is the importance of metagenomics?*** Over Earth's history, microbes have collectively driven massive global changes, such as adding oxygen to our atmosphere. If scientists can solve the metagenomic puzzle, then it will help us understand which microbes are in different environments and what they are able to do.

Metagenomic discoveries could lead to the development of methods for **carbon capture** and contaminant remediation, or production of **biofuels**. Metagenomics could help us understand the microbes that live within and on us to improve our health and prevent or treat diseases. Soil and plant-associated metagenomics could aid agricultural crop yields to feed more people around the world. Thus, solving the invisible puzzle that is the global metagenome has multiple societal benefits.

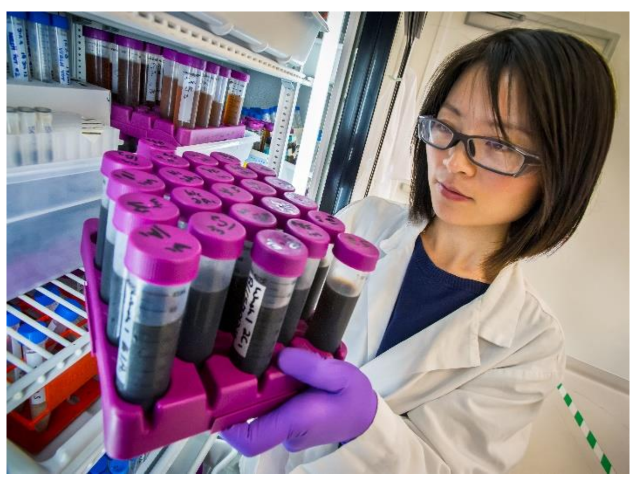
Metagenomics is at the center of discovery and development of natural products that can be used to create consumer goods, reduce negative environmental impacts, and improve our well-being. In these societal context examples, we will introduce a handful of ways in which metagenomics is used to understand the microbiology associated with human activities. The wide scope of challenges that metagenomics is applied to in these examples highlights the broad impact it has had in modern biological research and society in general.

a. *Our microbiomes.* The human body hosts a wide range of microbes, whether it's the dense community in our large intestine or the dry, scaly top of our heads. These microbes impact our health in unique ways that scientists are only beginning to understand. A big surprise when scientists first started studying microbes from the human body was that there are as many or more microbial cells than there are human cells. And in fact, there are even more microbial genes compared to human genes, when we look at the human genome compared to the human microbial metagenome. Thus, we are more microbial than we are human!

Most of these microbes are beneficial or neutral to human health, but some cause disease if able to grow unchecked or grow in areas they aren't supposed to be. With metagenomics, we are able to capture 'snapshots' of the microbes inhabiting human bodies from a range of healthy and disease states, allowing us to find microbial differences between each. With enough metagenomes we can discover what healthy microbial communities have in common, enabling us to find patterns that may be a cause or result of disease. But we already know that microbiomes significantly influence our physical and mental health.

b. *Disease prevention and therapy.* Demystifying human microbial communities is not the only medical advancement that can come from metagenomics. Discovery of microbial **metabolites** that aid in disease prevention and treatment has been a very active area of research since the first inoculations to prevent smallpox and the isolation of antibiotic penicillin from mold colonies. The "microbial arms race" that has occurred over billions of years of evolution has yielded an array of molecules that microbes use to deter the growth of adversaries, and that *we* can now exploit to deter the growth of disease-causing microbes. Metagenomics allows us to take advantage of evolutionary history, examining molecules that have medicinal properties to prevent and treat disease.

c. *Energy: renewable biofuels.* The decline of fossil fuel usage and subsequent production of biofuels as a more sustainable energy option is directly affected by metagenomic research. Two keys to biofuel production are the identification of plant **biomass** sources that are more easily broken down, and increasing the efficiency of that degradation process with microbial metabolisms.



Preparing plant degradation samples for analysis. Credit: Roy Kaltschmidt, Berkeley Lab.

One such example is the metagenomic sequencing of microbial communities growing on rice straw, a type of agricultural waste. Identifying the combination of microbes and **enzymes** that were very successful growing on this crop waste, and converting it to biofuel precursors, could lead to the creation of microbial consortia that enhance biofuel production technologies.

Other types of agricultural waste such as wheat straw, corn stover, and switchgrass are also potential sources of biomass that can be degraded by microbial consortia determined with the help of metagenomics. Using agricultural waste as a biomass source for biofuel production has the added benefits of not requiring additional land usage and reducing pollutants.

d. *Carbon capture technologies to reduce industrial carbon footprints.* While much of Earth's oxygen is produced by land plants, at least 50% of atmospheric oxygen is produced by marine **phytoplankton**, a collection of single-celled photosynthetic microbes. Prochlorococcus, a highly abundant type of photosynthetic bacteria found in the global oceans, produces up to 20% of atmospheric oxygen (see the NOAA Ocean Oxygen page for more information). These microbes produce oxygen as part of their metabolism, which consumes carbon dioxide and converts it into biomass. We can take advantage of these metabolisms to capture carbon dioxide produced by many industrial processes, using microbes and carbon capture technology, before it gets a chance to reach the atmosphere. Metagenomic analysis could reveal highly efficient enzymes or microbial communities that mitigate carbon emissions. These carbon dioxide consuming microbes could then in turn be used to create biofuels, reducing the demand of some carbon emitting industrial processes.

Preparation for aerial testing for carbon pollution in the atmosphere above San Francisco Bay Area. Credit: Roy Kaltschmidt, Berkeley Lab.

e. *Pollutant reduction.* **Bioremediation**, the biological degradation of environmental contaminants, can be an environmentally-friendly alternative to chemical and physical methods of contaminant degradation and sequestration. In many cases, microbes that degrade a contaminant are already present in the environment, but may have an abundance that is too low to mitigate harmful effects in both the short-term and long-term. With metagenomic analysis, we can learn how to temporarily increase the abundance of a desired microbe by providing nutrients to help them grow. Or, non-native microbes or enzymes that have the right type of metabolism for bioremediation can be added to the environment in an effort to speed up contaminant degradation. Introducing non-native microbes is less than ideal, so monitoring their growth with follow-up metagenomic analysis may be necessary.

f. *Improving plant health and crop yields.* The **rhizosphere**, the region of soil and microbes that are in close contact with plant roots, is critical for the health and productivity of our crops. The microbes within the rhizosphere accelerate the degradation of dead plant matter and nutrients released by plant roots, converting this biomass into **bioavailable compounds** and creating a nutrient-rich soil. Despite their importance, the microbial components that contribute to this nutrient cycling are largely unidentified. Soils that are abundant in these beneficial microbes may also keep harmful pathogenic bacteria away from plant roots. Metagenomic analysis of rhizosphere microbes will help us better understand the metabolic processes involved in nutrient cycling and lead to strategies for enhancing the production of nutrient-rich soils.

Experimental setup for rhizosphere imaging.
Credit: Marilyn Sargent, Berkeley Lab.

Relevance for Sustainable Development Goals and Grand Challenges

The global microbial metagenome relates to several SDGs (*microbial aspects in italics*), including:

- **Goal 2.** End hunger, achieve food security and improved nutrition and promote sustainable agriculture (*end hunger and malnutrition, improve agricultural sustainability and productivity*). The understanding of rhizosphere and soil microbes will positively impact crop yields, providing increased world food supply. Using metagenomics to enhance the processes of soil nutrient cycling will make farming a more sustainable practice by reducing the need for fertilizer supplementation and the ecological damage that is caused by agricultural runoff.
- **Goal 3.** Ensure healthy lives and promote well-being for all at all ages (*improve human health and well-being, prevent and treat disease*). In addition to improving global crop supplies and nutrition, metagenomics is heavily involved in understanding human health. The interplay between the human body and the microbes that inhabit it plays an important role in both staying healthy and diagnosing disease, and metagenomics is a useful tool for understanding the processes in each. In addition, metagenomics can be used to discover molecules from natural environments that help prevent or treat disease.
- **Goal 6.** Ensure availability and sustainability management of water and sanitation for all (*Ensure safe drinking water, improve water quality, reduce pollution, protect water-related ecosystems, and improve water sanitation treatment*). To ensure that communities have safe and sustainable water sources, water is treated so that it can be used as drinking water or reclaimed for irrigation. Metagenomics is a tool that can be used to monitor the quality of water and assess any microbe-related health risks in either municipal tap water or reclaimed irrigation water. Metagenomics can also be used to study microbial communities that help to break down waste before recirculation back into the environment. Agricultural runoff, sewage, landfill seepage, industrial waste, and many other pollutants contribute to the contamination of water sources and the right combination of microbes could help reduce the environmental impact.
- **Goal 7.** Ensure access to affordable, reliable, sustainable, and modern energy for all (*Ensure access to renewable and sustainable energy*). With the help of metagenomics, advancement of biofuel production technologies will lead to more sustainable and affordable energy sources. Without the need to extract source material for fossil fuels, biofuels can be produced in a wider range of locations and require less transportation, opening access to reliable energy sources globally.
- **Goal 8.** Promote sustained, inclusive and sustainable economic growth, full and productive employment, and decent work for all (*Promote economic growth, innovation, and job creation*). Metagenomics is an important tool of the growing biotechnology industry. Innovations in medicine, biofuel production, and product development discovered through metagenomics research will lead to the creation and expansion of businesses, and create jobs. Further, workforce development and training opportunities are increasingly available as the field of metagenomics expands across academic, government, and industry sectors.
- **Goal 11.** Make cities and human settlements inclusive, safe, resilient, and sustainable (*Effective bioremediation, water and sewage treatment, improved crop yields*). Improvements to both food production and waste processing will help improve growing city populations around the world, making them safer and more sustainable.

- **Goal 13.** Take urgent action to combat climate change and its impacts (*Reduce greenhouse gas emissions, produce renewable energy sources*). Metagenomics has revealed a wide diversity of biogeochemically relevant functions from uncultivated microbes, including new enzymes involved in nitrogen and carbon cycling. These new pieces can help scientists better understand climate mitigation and adaptation, and predict future climate scenarios.

Potential Implications for Decisions

1. Individual

   a. We know that our gut microbiomes influence our physical and mental health. Should we have a metagenome analysis of our gut microbiota? (Hint: do we know enough about which microbes are responsible for which conditions for routine metagenomic analyses to be useful?)

   b. Metagenomics has shown that what we eat influences our gut microbiome. If we want to have a healthy and diverse gut microbiome, we should eat diverse healthy foods.

2. Community policies

   a. A diverse microbiome is important for human health. Provision of green spaces encourages people to spend time outdoors which exposes them to soil and air microbes and enriches their microbiomes.

   b. Providing clean, safe drinking water requires regular monitoring. Metagenomic analyses provide detailed information on microbial quality and can inform about potential contaminants.

   c. Metagenomic analysis of wastewater treatment plant inflows and outflows can inform about infections and pathogen loads (e.g. SARS-CoV-2) in the community, and about pathogen loads released into the environment, both of which are important for public health agencies.

3. National policies

   a. Metagenome-informed healthcare policies: development of monitoring systems and regulatory practices using metagenomics in clinical settings.

   b. Policies to maintain, improve and restore quality of agricultural soils that are informed by an analytical toolbox that includes metagenomics

   c. Climate policies guided by metagenome data-integrated climate and Earth systems models.

   d. Public health infectious disease surveillance through metagenome-based monitoring of national networks of wastewater treatment plants.

Pupil Participation

1. Class discussions

   a. What is the potential to discover new microbes using metagenomics?

b. Can you think of how metagenomics could be used in your daily life or as a class?
c. Can you think of new applications for metagenomics?
d. Metagenomics has a variety of applications spanning human health and environmental sciences. What other useful information would you want to know about microbes beyond using metagenomics to access their genomes?

2. ***Pupil stakeholder awareness***
   a. Your favorite plant in the garden is growing poorly but has no obvious signs of attack by insects or plant pathogens. Do you think you might learn the reason by carrying out a metagenomic analysis of the rhizosphere microbes?
   b. A new investigation reports the movement of a pollutant in groundwater towards a drinking water supply. What do you think we should do and what role might metagenomics play in this?
   c. Global warming is one of the major problems facing humanity. How can metagenomics contribute to mitigation strategies?
   d. You like to bake sourdough bread. One day, the dough looks funny. Wouldn't it be interesting to compare its metagenome with that of your normal dough?
   e. Your compost pile decomposes at different speeds when you have it enclosed or out in the open. Can metagenomics help you understand why, and maybe choose the better composting method?
   f. The nearby pond smells like sulfur in the Summer. How would you use metagenomics to see how the pond is different from the Winter?

The Evidence Base, Further Reading and Teaching Aids

- Podcasts: The Genome Insider and Natural Prodcast
- Webinars: Microbial Genomes & Metagenomics Webinar Series
- NIH Tools for Teaching the Microbiome
- DOE Joint Genome Institute Blog and Science Highlights

Glossary

Metagenomics: the study of genetic material (genomes) recovered directly from a mixed community of organisms from an environmental sample
DNA (deoxyribonucleic acid): chemical name for the molecule that encodes genetic instructions for all living things. A DNA molecule consists of two strands that form a double helix with four bases, adenine (A), cytosine (C), guanine (G), and thymine (T)
Genome: the entire set of genetic instructions found in a cell
Gene: the basic physical unit of inheritance that contains the information needed to specify traits or functions
Sequencing technology: technique for determining the DNA sequence of an organism's genome

Carbon capture: process of capturing carbon dioxide (CO_2) before it enters the atmosphere and storing it (carbon sequestration)
Biofuel: any fuel that is derived from biological material, primarily plant or algae
Metabolite: a chemical made or used when a cell breaks down food
Biomass: the total mass of organisms in a given area or volume
Enzyme: a biological catalyst, almost always a protein
Phytoplankton: photosynthesizing microbes that inhabit the upper sunlit layer of almost all oceans and bodies of freshwater
Bioremediation: the use of either naturally occurring or introduced microbes to consume and break down environmental pollutants
Rhizosphere: the region of soil surrounding plant roots where the chemistry and microbiology is influenced by their growth, respiration, and nutrient exchange