2021

NHGRI Short Course in Genomics

Microbiome Virtual Lab Exploration!

Joshua Farr, M.Ed. NBCT

Science Department Team Lead
Mica Mountain High School

Sean P. Conlan, Ph.D.

Associate Investigator Microbial Genomics Section
Translational and Functional Genomics Branch
National Human Genome Research Institute

NIH National Human Genome Research Institute

The Forefront of Genomics®

Forward

"We can compare the gut of a person with inflammatory bowel disease to a dying coral reef or a fallow field: a battered ecosystem where the balance of organisms has gone awry."
*(***Ed Yong,** *I Contain Multitudes: The Microbes Within Us and a Grander View of Life)*

This quote from Ed Yong illustrates how the study of the human microbiome is comparable to the way in which ecologists study complex communities. While people are probably most familiar with the human microbiome, every ecological niche on Earth has its own collection of distinctive microorganisms. This includes coral reefs and barnyard soils, as well as hostile environments like hot springs or acidic outflows from mining operations. These microbial communities are critical to important planet-wide processes like carbon sequestration, nitrogen fixation and the breakdown of toxic chemicals. These wildly differing environments and functions reflect another important feature of the microbiome: diversity. Despite a bias for focusing on the bacterial component of the microbiome, microbial communities are complex and draw organisms from across the tree of life, including fungi, viruses, protists and metazoans.

Bacteria are often the focus of microbiome studies because microbiologists have developed robust tools for growing and identifying bacteria. Part of this stems from our frequently adversarial relationship with bacteria. Pathogens like *Yersinia pestis*, the causative agent of the Black Plague, have fundamentally shaped human history. The careers of famous microbiologists like Louis Pasteur (1822-1895), Robert Koch (1843-1910) and Alexander Fleming (1881-1955) were built on understanding and eradicating bacterial infections. While modern life wouldn't be possible without medical advances to treat bacterial infections, it's important to understand that bacteria far more often play a beneficial role in the world. Unlike pathogenic bacteria, which usually act alone, beneficial bacteria often function as a community made up of tens or hundreds of different microbes. Consortia of bacteria (and also fungi) are indispensable for making bread, cheese, yogurt, kimchi and many other foods. The microbiome is important to human health and the development of a strong immune system. Microbial communities are critical for nitrogen fixation and the health of our crops and livestock.

This collection of lesson plans is designed to introduce students to the microbiome through discussions of current science and hands-on experiments.

- **A Glimpse into the Microbiome provides an overview of the microbiome** starting with the ecological paradigms that defined early microbiome science and finishing with the impact of the microbiome on human health.

- **Exploring the Microbiome and Its Connection to Metabolic Syndrome** offers a detailed look at the current science linking the microbiome to a constellation of health problems called metabolic syndrome. Studies connecting obesity, diabetes and hypertension to the microbiome are some of the most frequently cited examples of microbiome-mediated health conditions.

- **You Are What You Eat — Exploring the Microbiome Through Inquiry-Based Labs** teaches students important concepts like the scientific method, sterile technique and reproducibility.

- **Microbiome Virtual Lab Exploration!** gives student the opportunity to analyze real microbiome data using web-based analysis tools.

Taken together, these lesson plans can be adapted to a variety of settings, student populations and educational goals. We hope students will be inspired to learn more about the hidden communities of microbes that shape human health and the environment.

Sean P. Conlan, Ph.D.
Associate Investigator
Microbial Genomics Section
Translational and Functional Genomics Branch
National Human Genome Research Institute

Microbiome Virtual Lab Exploration!

Timeframe

Time dedicated to this project varies based on the needs or desire to explore the content.

This project can be broken into multiple lessons/class periods of ~45-55 minutes each or introduced and then used during a single block period.

Requirements for successful implementation include

- Review of prior concepts: DNA, evolutionary evidence linking organisms, data/graphical representation, eukaryotic and prokaryotic cells, taxonomical hierarchy.
- Introduction to using PubMed and online databases.
- Introduction to database searching with RDP Classifier and BLAST. If students are more familiar with content, the teacher may adjust as desired to explore topics that may be of interest or yield metagenomic science studies.

Key concepts

- Students will be able to comprehend what the human microbiome is and how it differs between body sites. Analogy: different countries have different cultures. Different areas of the body are analogous. The term “culture” was meant to be a pun!
- Learn how DNA sequencing can be used to survey complex microbial communities.
- Additionally, help students develop questions for further research pertaining to their interest.

Learning objectives

- Discuss the complexity of the human microbiome.
- Explore a computational lab allowing students to explore rich microbiome data.
- Identify trends and make predictions based on information acquired from databases.
- Develop research questions for further investigation of the human microbiome.

Prerequisite knowledge

- Students should have a working knowledge of how DNA is a heritable molecule passed on from one parental generation to offspring. Specific genes yield traits that make organisms fit for their environment and those genes can be compared between or among species. This knowledge is the foundation for this advanced, yet achievable introduction to using databases to search “big data” for patterns and identifying unique attributes. The connection to the human genome is probably also going to raise many questions among students, which is exciting!
- After doing this activity, students should come up with research questions that they can explore. Based on the desires of the students and time alloted by the teacher, students could work in collaborative groups or solo to develop their questions. The teacher may want to take the time to practice writing research questions with students to help them hone in on what they are truly curious and specific enough about.

Standards (based on NGSS)

- Students who demonstrate understanding can:

  HS-LS1-2. Develop and use a model to illustrate the hierarchical organization of interacting systems that provide specific functions within multicellular organisms.
  - *Student-friendly language:* Students connect how organisms are related based on their genome.

HS-LS4-1. Communicate scientific information that common ancestry and biological evolution are supported by multiple lines of empirical evidence.

- *Student-friendly language:* Students should be able to use molecular evidence for comparing known and unknown organisms yielding taxonomic info (KPCOFGS).

HS-LS3-1. Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.

- *Student-friendly language:* Students connect that all organisms contain a genome made of DNA or RNA, and this commonality allows us to compare one organism to another.

HS-LS3-3. Apply concepts of statistics to explain the variation and distribution of expressed traits in a population.

- *Student friendly language:* Students should be capable of using statistical tools such as mean, median, mode, and standard deviation to graphically visualize data.

Materials and handouts

For the teacher

- Grice, E. and Segre, J. (2012). "The Human Microbiome: Our Second Genome." Annual Review of Genomics and Human Genetics 13. pp. 151-70. 10.1146/annurev-genom-090711-163814.
- Online access to databases: Teachers may need to ask their tech departments for access prior to attempting this lesson. Additional KWL charts may help students exhibit learned material compared to potential research project or extension projects.

For the student

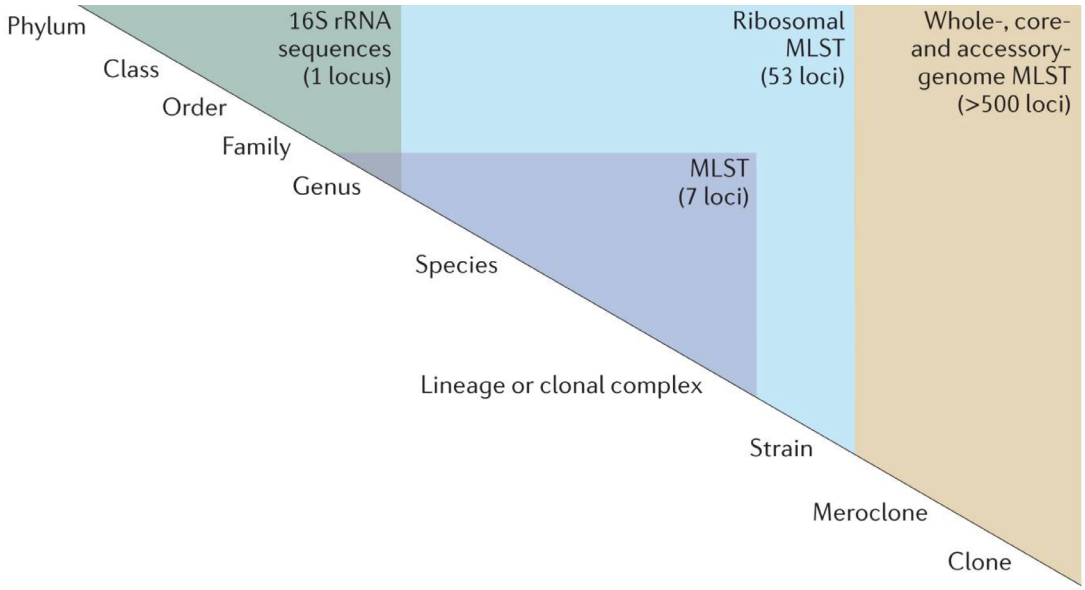
- Select an intro video about microbiomes or play both.
  - The Hidden World of Microbiomes
  - The Invisible Universe of the Human Microbiome NPR version
- "How to Sequence the Human Genome" TEDEd video, which may be useful for students. This activity is acts on the assumption that students have been exposed to or have prior knowledge of genome sequencing: https://ed.ted.com/lessons/how-to-sequence-the-human-genome-mark-j-kiel
- Figure 2 from “The human microbiome: our second genome" (Grice and Segre, 2012) printed in black and white ready for students to color code (possible pre-lab homework assignment to get students thinking about the diversity of organisms).
- Sequence data tool RDP Classifier: https://rdp.cme.msu.edu/classifier/classifier.jsp
- Access to PubMed: https://www.ncbi.nlm.nih.gov/pubmed/
- Access to BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi
- How to use BLAST video tutorial: https://www.youtube.com/watch?v=HXEpBnUbAMo

Activity background

One of the most basic questions a microbiome researcher can ask is: “What is in my sample?” There are many ways to answer this question, and the method you choose will determine how much biological detail you can elicit. The microbiome refers to the collection of bacteria, fungi, viruses, protists and metazoans in a sample, but researchers are often specifically interested in the bacterial component. Identification of bacteria is based on a taxonomic hierarchy. For instance, most people are familiar with the bacteria *Escherichia coli*. The bacteria *E. coli* is in the family Enterobacteriaceae and the phylum Proteobacteria; the full taxonomic hierarchy for *E. coli* is:

Bacteria **(Kingdom)**
Proteobacteria **(Phylum)**
Gammaproteobacteria **(Class)**
Enterobacterales **(Order)**
Enterobacteriaceae **(Family)**
Escherichia **(Genus)**
Escherichia coli **(Species)**

While taxonomic levels often stop at “species," additional taxonomic levels that allow scientists to categorize bacteria in finer detail (e.g., strains). The level of detail you can get from a microbiome experiment depends on the experimental method. For instance, figure 2 from Maiden et al., *Nature Reviews* 2013 (PMID: 23979428) (Maiden, Martin C. J., et al. (2012) “MLST revisited: the gene-by-gene approach to bacterial genomics.” *Nature Reviews Microbiology*, 11(10), pp. 728–736., doi:10.1038/nrmicro3093.) shows the level of resolution provided by various sequencing experiments.



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3980634/figure/F2/

In the experiment below, you will receive 16S rRNA sequence data. These sequences are derived from a fragment of the 16S rRNA gene and can be easily read with standard DNA sequencers. From the figure above you can see that short 16S rRNA sequences allow you to identify bacteria to the level of genus; full-length 16S rRNA sequences often allow classification to the species level.

For those who are interested, the 16S rRNA is an RNA molecule that forms part of the structure of the bacterial ribosome, the molecular machine used to synthesize protein from mRNA. The 16S rRNA has

conserved regions that don't vary much from bacteria to bacteria and variable regions that can be used as a sort of barcode or fingerprint for bacteria. The 16S rRNA is a remarkable molecule. Read more about it in the *Journal of Clinical Microbiology from a 2007 published article*.

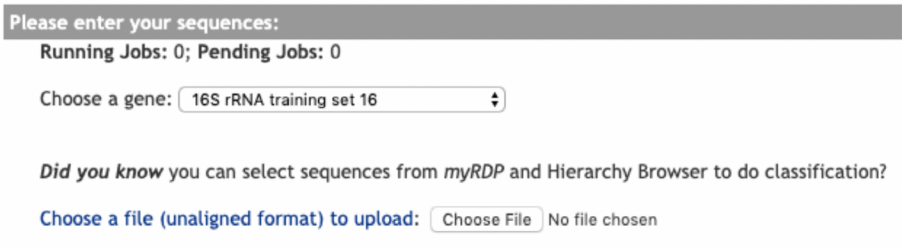
Each student (or group of students) will receive a file containing 100 short fragments of DNA collected from a specific human body site. These fragments represent a survey or census of what bacteria are present in the sample. Using figure 2 from "The human microbiome: our second genome" (Grice & Segre, 2012), you will use an online tool to categorize these sequences and make a guess where your body site collection of sequences came from.

Activity/lab protocol

1. Obtain assigned sequence set by the teacher. The file could be distributed by email, Google drive, USB drive, etc., DNA sequences are typically shared as text files in a format called FASTA. They are just text files, so you can open in GoogleDocs or similar if you want to look at the sequences.

    1. Sequence set: ba04826.sub100.fasta
    2. Sequence set: st06686.sub100.fasta
    3. Sequence set: to10842.sub100.fasta
    4. Sequence set: vf03604.sub100.fasta

2. Go to the RDP Classifier: https://rdp.cme.msu.edu/classifier/classifier.jsp
3. Under "Choose a file," select your file and hit "Submit."



4. On the results page, you will see a taxonomic breakdown showing how many sequences are assigned to each taxonomic level. Students should try to figure out which body site they have based on the bacteria found. You may have to adjust the "Display depth" dropdown menu on the results page to "10" to see detail at the genus level.

In-class example

This is an example of the RDP Classifier output for sequences from a soil sample (associated with this paper: https://www.ncbi.nlm.nih.gov/pubmed/17041161, sequence accessions DQ827724:DQ829627). Soil is complex and has a lot of different bacterial taxa. A small section of the RDP report is below

- phylum "Bacteroidetes" (50)
  - class Sphingobacteriia (19)
    - order "Sphingobacteriales" (19)
      - family Sphingobacteriaceae (1)
        - genus Sphingobacterium (1)
      - family Chitinophagaceae (17)
        - genus Ferruginibacter (1)
        - genus Filimonas (1)
        - genus Parasegetibacter (1)
        - genus Terrimonas (4)
        - unclassified_Chitinophagaceae (10)
      - family "Saprospiraceae" (1)
        - unclassified_"Saprospiraceae" (1)
  - class Cytophagia (9)
    - order Cytophagales (9)
      - genus Chryseolinea (1)
      - family Cytophagaceae (6)
        - genus Adhaeribacter (4)
        - genus Sporocytophaga (2)
        - unclassified_Cytophagales (2)
  - class Flavobacteriia (12)
    - order "Flavobacteriales" (12)
      - family Flavobacteriaceae (12)
        - genus Flavobacterium (12)
  - unclassified_"Bacteroidetes" (10)
- phylum "Chloroflexi" (18)
  - class Chloroflexia (2)
    - unclassified_Chloroflexia (2)
  - class Caldilineae (3)
    - order Caldilineales (3)
      - family Caldilineaceae (3)
        - genus Litorilinea (2)
        - unclassified_Caldilineaceae (1)
  - class Ktedonobacteria (4)
    - order Ktedonobacterales (1)
      - unclassified_Ktedonobacterales (1)
    - unclassified_Ktedonobacteria (3)
  - class Anaerolineae (6)
    - order Anaerolineales (6)
      - family Anaerolineaceae (6)
        - unclassified_Anaerolineaceae (6)
  - unclassified_"Chloroflexi" (3)
- phylum "Armatimonadetes" (14)
  - genus Armatimonadetes_gp5 (1)
  - class Armatimonadia (1)
    - order Armatimonadales (1)
      - family Armatimonadaceae (1)
        - genus Armatimonas/Armatimonadetes_gp1 (1)
  - genus Armatimonadetes_gp4 (8)
  - class Chthonomonadetes (4)
    - order Chthonomonadales (4)
      - family Chthonomonadaceae (4)
        - genus Chthonomonas/Armatimonadetes_gp3 (4)

While this may look somewhat complicated, focus on the genus-level classifications. The sequences for the full analysis are included in the DOK03 file should you want to use this example in class.

There are many options for adding to this:

1. Students can use PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) to download the article instead of giving it to them. Literature searching and familiarity with PubMed is an important skill. A teacher may decide to take more time with students to do article searching, or use it as a follow-up lesson building on the research questions developed by students.
2. Students can BLAST some sequences to try to figure out what they are below the genus level. For instance, the sequences below are full-length 16S rRNA sequences from examples of human pathogens. Access https://blast.ncbi.nlm.nih.gov/Blast.cgi, click Nucleotide BLAST, paste one of the sequences from below into the query box and hit “BLAST."

>S000645022
ttttatggagagtttgatcctggctcaggatgaacgctggcggcgtgcctaatacatgcaagtcgagcgaacggacgaga
agcttgcttctctgatgttagcggcggacgggtgagtaacacgtggataacctacctataagactgggataacttcggga
aaccggagctaataccggataatattttgaaccgcatggttcaaaagtgaaagacggtcttgctgtcacttatagatgga
tccgcgctgcattagctagttggtaaggtaacggcttaccaaggcaacgatgcatagccgacctgagagggtgatcggcc
acactggaactgagacacggtccagactcctacgggaggcagcagtagggaatcttccgcaatgggcgaaagcctgacgg
agcaacgccgcgtgagtgatgaaggtcttcggatcgtaaaactctgttattagggaagaacatatgtgtaagtaactgtg
cacatcttgacggtacctaatcagaaagccacggctaactacgtgccagcagccgcggtaatacgtaggtggcaagcgtt
atccggaattattgggcgtaaagcgcgcgtaggcggtttttaagtctgatgtgaaagcccacggctcaaccgtggaggg
tcattggaaactggaaaacttgagtgcagaagaggaaagtggaattccatgtgtagcggtgaaatgcgcagagatatgga
ggaacaccagtggcgaaggcgactttctggtctgtaactgacgctgatgtgcgaaagcgtggggatcaaacaggattaga
taccctggtagtccacgccgtaaacgatgagtgctaagtgttaggggtttccgcccccttagtgctgcagctaacgcatt
aagcactccgcctggggagtacgaccgcaaggttgaaactcaaaggaattgacggggacccgcacaagcggtggagcatg
tggtttaattcgaagcaacgcgaagaaccttaccaaatcttgacatcctttgacaactctagagatagagccttcccctt
cgggggacaaagtgacaggtggtgcatggttgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgcaacgagcgc
aacccttaagcttagttgccatcattaagttgggcactctaagttgactgccggtgacaaaccggaggaaggtggggatg
acgtcaaatcatcatgccccttatgatttgggctacacacgtgctacaatggacaatacaaagggcagcgaaaccgcgag
gtcaagcaaatcccataaagttgttctcagttcggattgtagtctgcaactcgactacatgaagctggaatcgctagtaa
tcgtagatcagcatgctacggtgaatacgttcccgggtcttgtacacaccgcccgtcacaccacgagagtttgtaacacc
cgaagccggtggagtaacctttaggagctagccgtcgaaggtgggacaaatgattggggtgaagtcgtaacaaggtagc
cgtatcggaaggtgcggctggatcacctcctttct

>S000006023
atcatkgctcabgatgaacgctggcggcgtgcctaacacatgcaagttgagcgatttacttcggtaaagagcggcggacg
ggtgagtaacgcgtgggtaacctaccctgtacacacggataacataccgaaaggtatgctaatacgggataatatatttg
agaggcatctcttgaatatcaaaggtgagccagtacaggatggacccgcgtctgattagctagttggtaaggtaacggct
taccaaggcgacgatcagtagccgacctgagagggtgatcggccacattggaactgagacacggtccaaactctacggga
ggcagcagtggggaatattgcacaatgggcgaaagcctgatgcagcaacgccgcgtgagtgatgaaggccttcgggtcgt
aaaactctgtcctcaaggaagataatgacggtacttgaggaggaagccccggctaactacgtgccancagccgcggtaat
acgtaggggctagcgttatccggatttactgggcgtaaagggtgcgtaggcggtctttcaagtcaggagtgaaaggcta
cggctcaaccgtagtaagctcttgaaactgggagacttgagtgcaggagaggagagtggaattcctagtgtagcggtgaa
atgcgtagatattaggaggaacaccagttgcgaaggcggctctctggactgtaactgacgctgaggcacgaaagcgtggg
gagcaaacaggattagataccctggtagtccacgctgtaaacgatgagtactaggtgtcgggggttacccttcggtgcc
gcactaacgcattaagtactccgcctgggaagtacgctcgcaagagtgaaactcaaaggaattgacggggacccgcacaa
gtagcggagcatgtggtttaattcgaagcaacgcgaagaaccttacctaagcttgacatcccaatgacatctccttaatc
ggagagttcccttcggggacattggtgacaggtggtgcatggttgtcgtcagctcgtgtcgtgagatgttgggttaagtc
ccgcaacgagcgcaacccttgtctttagttgccatcattaagttgggcactctagagagactgccagggataacctggag
gaaggtggggatgacgtcaaatcatcatgccccttatgcttagggctacacacgtgctacaatgggtagtacagagggtt
gccaagccgtaaggtggagctaatcccttaaagctactctcagttcggattgtaggctgaaactcgcctacatgaagctg
gagttactantaatcgcaatcaaaatgctgcggtgaatgcgttcccggggtcttntacacaccgcccgtcacaccacggg
agttggaaacgcccgaagccgaattatctaacctttggaanaantcstcgaaagtggaatcaataacttgggtnaantc
gtaacaaggtaaccgtatcggaaggt

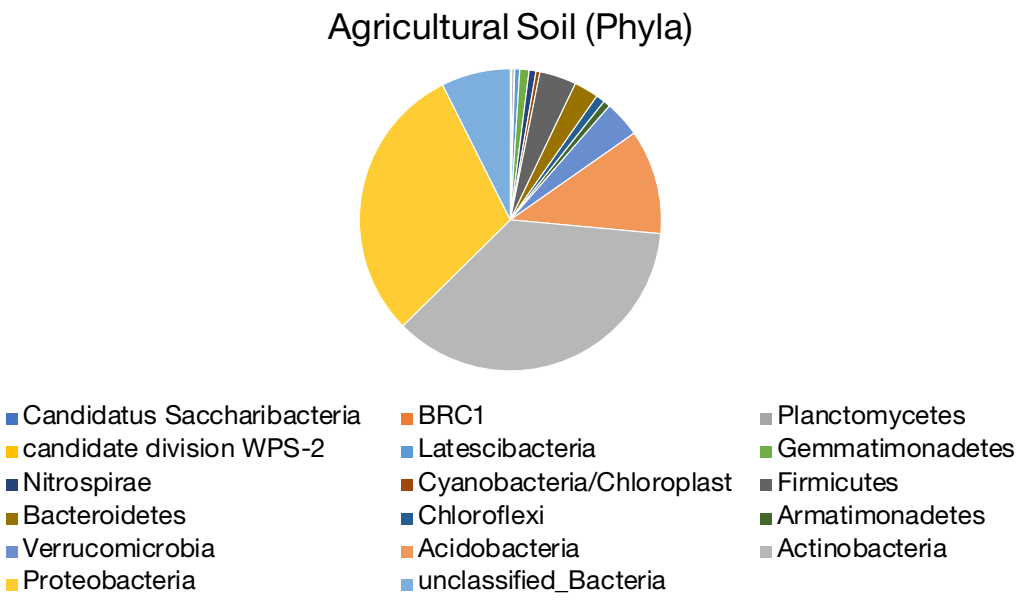
You should notice two things about the blast results:

- There is a ranked list of strong matches attached to annotated bacterial reference sequences.
- For the first sequence, it is hard to say exactly what strain it is, but the bacterial species should be obvious. The second sequence only has a single match for a possible strain identification.

Both of these bacteria can be pathogenic if found in the wrong body location. Students can familiarize themselves with both of these organisms at the Centers for Disease Control (outlets provided by the teacher). This could also be an expanded conversation. What characteristics make websites reliable?

- https://www.cdc.gov/hai/organisms/staph.html
- https://www.cdc.gov/hai/organisms/cdiff/cdiff_infect.html

3. Students can explore other opportunities for calculations and data visualization, including, plotting, diversity calculations, etc., For instance, one could make pie charts showing the abundance of bacteria at various taxonomic levels. Taking the soil example from above., RDP classify, change the Display Depth dropdown to 3 to get a manageable list of taxa and plot something like:



4. Students can develop researchable questions with their new database skills. Questions must be written, edited and improved by other students. GoogleDocs is great for this, as a teacher could have a set of questions from one period and share it with another period (or class, school or country!) to get feedback.
   - **Questions cannot be:** yes or no, answered by a simple Google search, unethical or use human subjects. The teacher can modify and adjust the timeframe for such an exploration as desired.

Closure

Students have been asked to connect many things and explore many new ideas.

Ideas for closure:

- Using a set of newly learned terminology, link them into a written “Exit Ticket” to leave class: bacteria, genome, database, microbiome, etc. The teacher can decided or ask students to come up wtih a list that they write on a board. Students can then provide input about how to connect these terms as a class or individually. The teacher may want to create a short list of words to practice with all students and then a “homework list” for students to work on and practice with outside of class. Encouragement should be placed on connections, use of arrows, multiple colors, emojis or illustrations to help the students make sense of these complex terms.
- Visually draw out and explain on a whiteboard (or create a short video) a flow chart of how to use databases and compare organisms.
- Explain three things you’ve learned about the human microbiome.
- Write down a question about something you still don't know about the human microbiome.
- These last few items could be combined in a Know, Want to Know and Learned chart (often called a KWL chart).

Handout options are attached below:

- Students can use the black and white version. Students could color code specific regions and have a record of a visual in their class notebook.
- Teachers can use the color version to project or put on screen after students color code or to help guide students with understanding.

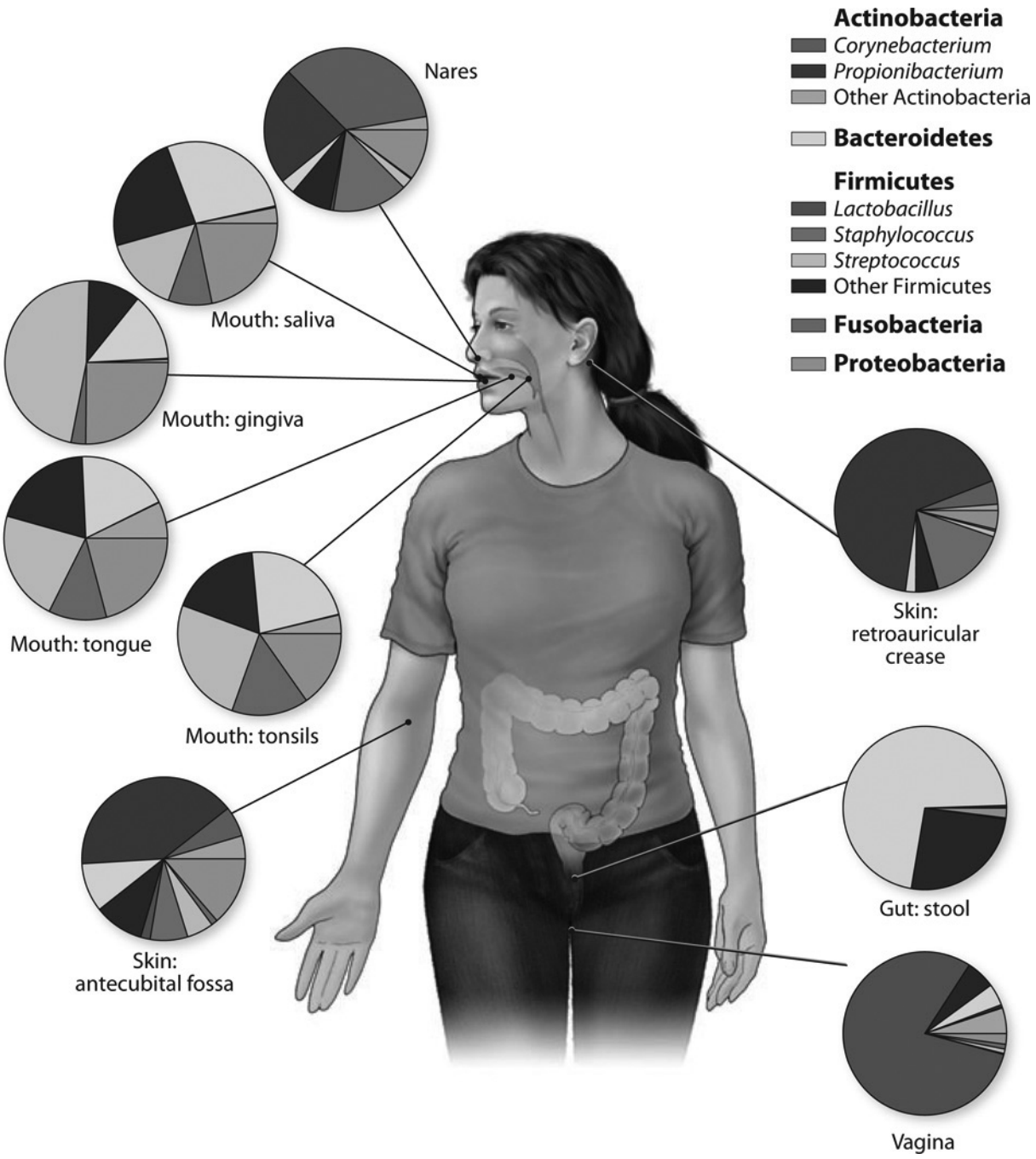


Figure 2. Genus- and phylum-level classification of bacteria colonizing a composite subject, showing that human microbiome diversity is dependent on the site sampled. Sites in the oral cavity share greater similarity than other types of sites, such as the skin, vagina, and gut. Data derived from the NIH Human Microbiome Project study (http://commonfund.nih.gov/hmp).

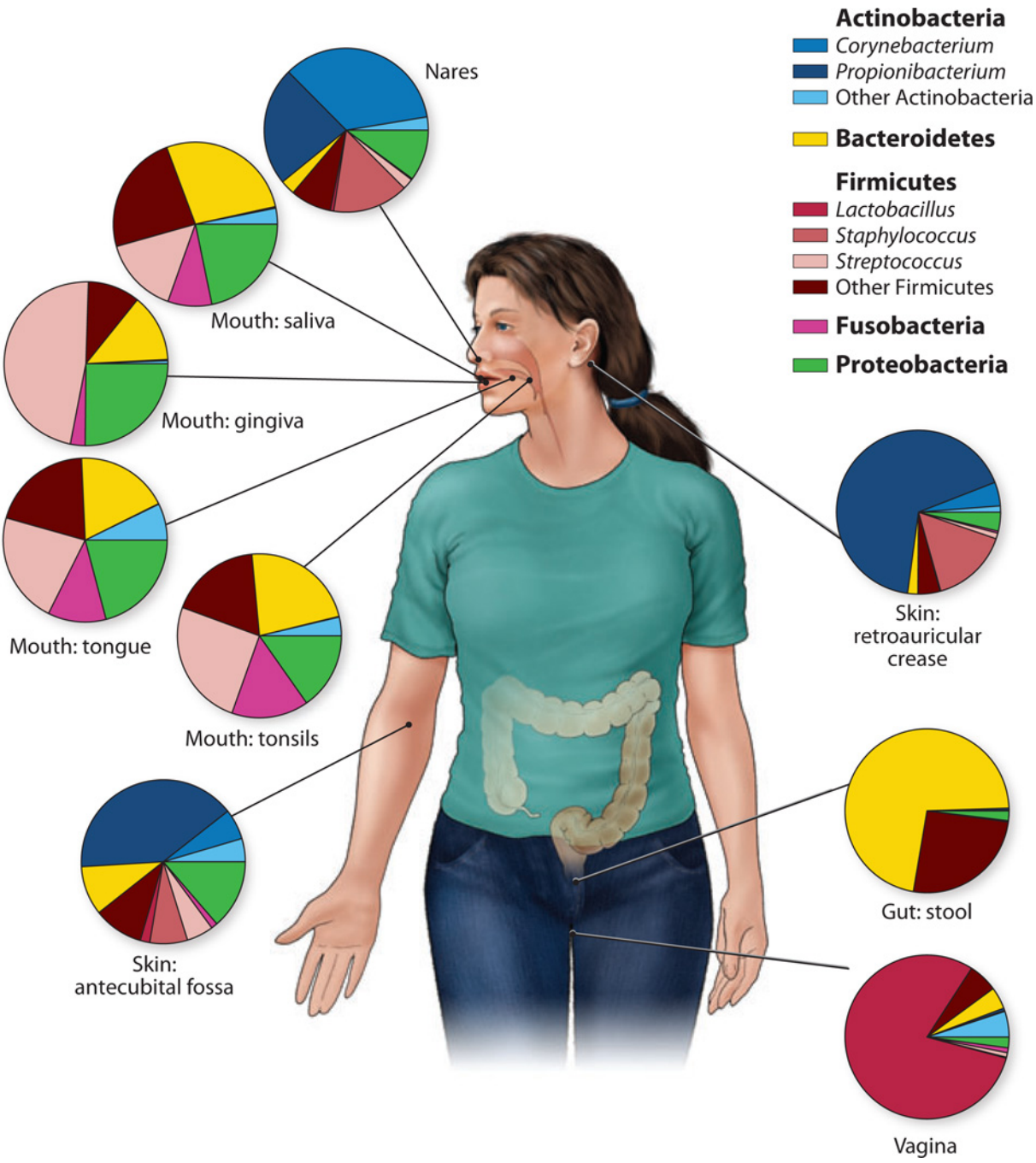


Figure 2. Genus- and phylum-level classification of bacteria colonizing a composite subject, showing that human microbiome diversity is dependent on the site sampled. Sites in the oral cavity share greater similarity than other types of sites, such as the skin, vagina, and gut. Data derived from the NIH Human Microbiome Project study (http://commonfund.nih.gov/hmp).


The Forefront
of Genomics®