Below are answers to questions on the worksheet.

**Pre-Exercise Activity: Genome Annotation Slide Deck**

1. An annotation defines the location of items on a landscape. For a map, it may be roads. For genomes, it is where genes and other information is found in DNA.
2. Genomes are annotated according to structural versus functional information.
3. Structural annotation indicates where DNA features are located (e.g., start/stop codons for a protein-coding region) Functional annotation indicates where cellular information are coded (e.g., which cellular function the DNA region codes for).
4. DNA has two polymer strands with complementary (different) sequence. Each strand can have three coding regions because codons are three nucleotides. So starting with one nucleotide would give a different set of codons than starting with the next two adjacent nucleotides. Starting with the third nucleotide away give the same set of codons. Twice three is six.
5. An open reading frame is a (long) stretch of codons between a start and a single stop codon.
6. A start codon.
7. The Shine-Dalgarno sequence is a consensus sequence that is used for initial ribosome binding. It will indicate a nearby start codon.
8. AT rich region may indicate stop codons. GC rich region may indicate a coding region. Shine-Dalgarno sequence may indicate a start codon. AG rich region may indicate a Shine-Dalgarno sequence.
9. One can compare (BLAST) novel sequences to databases of known sequences.
10. 1) function of the protein, 2) cellular location of a protein, 3) cellular process the protein contributes to
11. Direct laboratory experiments; simple sequence comparisons
12. Both tRNA and rRNA sequences are highly conserved and relatively easy to find by simple sequence comparison. They are also numerous. Finding them first narrows down the parts of the genome to annotate.

**Part 1: Exploring and downloading data from a human microbiome database:**

1. 75160 (as of 12/30/2019) This number is under IMG Content on the left.
2. Joint Genome Institute (DOE – US Department of Energy)
3. Integrated Microbial Genomes & Microbiomes
4. West coast of North America, Caribbean Sea, Europe, Mediterrean Sea, Arctic, Antarctica, Africa, Pacific Ocean, Hawaii, Central and South America, South Pacific, etc. Click on the number to see the area it is counting.
5. J. Craig Venter Institute (JCVI) sequences metagenomes from samples typically from ocean environments.
6. Permanent draft and finished
7. Finished; no, annotation may not be complete, just that the confidence level of the genome sequence being correct is high
8. JCVI
9. October 1995 in Akershus Norway; from the periodontal pocket of a 76-year-old woman with marginal periodontitis
10. Facultative anaerobe; nitrate reducer; auxotroph for L-lysine, L-phenylalanine, L-tyrosine, L-histidine, L-serine; prototrophic for the remaining amino acids; non-selenocysteine synthesizer; makes and assimilates acetyl-CoA/CoA; glucose and fructose utilizing; bacilli; Gram positive; motile; found in soil and human digestive system (mouth); pathogenic; mesophile; sporulating; can produce protective endospores
11. 5, 588, 834 bp
12. 35.31%
13. 5,810 genes are protein coding (97.79%)
14. CP001283.1

**Part 2: Comparing gene regions on a genome browser (you will need to be logged into the IMG database for this part):**

1. *Helicoacter pylori, Helicobacter bizzozeronii, Campylobacter coli*
2. 1680 bp versus 1827 bp and 285 bp (*H. bizzozeronii*). It appears that the *H. bizzozeronii gene is a truncated form of the gene found in the other two species*
3. DNA binding domain, DNA primase activity (DNA-directed RNA polymerase activity), metal ion binding domain, nucelotidyltransferase activity, transferase activity, zinc ion binding domain
4. CHC2-type zinc finger, DNA primase, TOPRIM domain
5. No gene families had significant matches.
6. *H. pylori* seems well annotated whereas *H. bizzozeronii* needs more evidence for functional domains to match to gene families. The other genes in this region of the *H. bizzozeronii* genome are hypothetical proteins.