Suggested answers and notes to questions in **Databases: A Study of Influenza**

1. Massively parallel sequencing due to improvements of technology – next generation sequencing and computational programs.
2. Computer technology improvements helped with data analysis.
3. Bacterial phenotypes: Cell shape, motile/nonmotile, pathogenic/nonpathogenic, Gram negative/Gram positive

Mammals/birds: beak shape/size, facial features, size, limb type, body shape, hair, feathers, inside body versus outside body development, coloration,

1. A microbiome is an “aggregate of all microbial genomes” or bacteria in a sample/location. Microbiomes are in the intestines, soil, ocean, oral cavity, nasal cavity, and on skin surface
2. Only about 1% of the microbiome can be cultured. This means that 99% of the microbiota cannot be studied as pure cultures in the laboratory. When DNA sequencing of the microbiota of a microbiome, one does not have to purified the bacteria first. The new problem is to sort out the sequences to determine the unique ones. Also there is a problem with assembling the sequences into a genome.
3. Biological databases are a repository of biological data including DNA and protein sequences, taxonomic classification of organisms, etc. 1) DNA or protein sequences mainly, 2) features of the sequences and its role in the cell, 3) the database must have a web-based interface or “genome browser” so scientists can access and analyze the data
4. USA – GenBank; UK – EMBL; Japan – DDBJ
5. The accession number of the DNA sequence.
6. It has information about the sequence in organized columns.
7. Mutation to change the DNA sequence, inheritance to pass down the specific sequence, and evolution to cull the sequences
8. More differences imply different lineages and more distantly related. You have to go further back in generations to find the most recent common ancestor.
9. Expression of genes that lead to morphology is based on environmental factors and not only on inheritance of genes.
10. Influenza A’s genetic material is linear RNA molecules – 8 of them.
11. Hemagglutinin (HA) helps the virus attach to host cells. Neuramidase (NA) cleaves the viral particles from the host cell surface releasing the viral particles (students have to Google this). Both NA and HA are found on the surface of the Influenza. There are 15 HA and 9 NA subtypes.
12. Antigenic drift is the mutation of NA and HA genes causing different amino acids to be on their surface. This change renders the Influenza virus undetectable by the host immune system (antibody). The mutations occur readily because RNA polymerase that copies the Influenza genetic material does not have good proofreading function.
13. Observations may include the subtype of virus that recombined to make the 2009 H1N1, the sources of the viruses, the year of the viral sample.
14. Hypotheses will depend on the answers to question #16. Its general form is: The 2009 H1N1 virus came from the reassortment of …
15. Each entry (a-g) should have a specific choice. Write down the actual choices made on the website.