Below are a list of questions or question variants that may appear in a cumulative quiz. The quiz will typically take place in the last 15 minutes of class on Thursday, and you will have to answer TWO randomly chosen questions. You will not be allowed to use any resources or notes. For each question, you will either answer to satisfaction or not to satisfaction, and no partial credit is given. There are revisions at the end of the semester as part of your third self-eval (3 corrections to equal one full CQ), and there are more opportunities to answer CQs than are needed for an 'A-range' course grade. Study and make sure you know how you would answer each question ahead of time to increase your chances of satisfactorily completing CQ questions.

Text shaded gray means I have asked that question for a CQ and (probably) will not ask it again.

Week 1

(Once two have been chosen, no more Week 1 questions will feature in a CQ)

For the chosen characteristic of big data in biology, provide the following three items:

a. An example of this characteristic in the context of big data in biology
b. An advantage of this characteristic in the context of big data in biology
c. A challenge/problem of this characteristic in the context of big data in biology

1-1. Volume

1-2. Velocity

1-3. Variability

1-4. Complexity

Week 2

1. Draw out the structure of a DNA molecule where one strand is 5' GT 3'. Phosphate groups can be given by circles, nitrogenous bases can be given by triangles, and hydrogen bonds can be designated by dashed lines. Make sure the following are clearly shown or labeled:
    a. A, G, C, and T are put in the triangles to indicate the type of nucleotide
    b. All 5' and 3' ends of the DNA molecule are indicated

2. What are the four major ingredients used in a polymerase chain reaction? List all four and for two of them (your choice), explain their role in the process.

3. Pick TWO aspects of DNA structure, describe that aspect of the structure, and explain how that aspect makes sequencing a challenge.

4. Explain in 1-2 sentences what Phred scores are (explain enough so that a person who hasn't heard of them will understand how they relate to sequencing), and why they are an important metric to examine in the sequencing process.

Week 3

1. You will be given a gel with the results of a Sanger Sequencing run visualized. An example image is shown below, but it is NOT the one I will give during the quiz. Answer the following two questions:
    a. What is the sequence of the TEMPLATE strand? If you have multiple DNA strands written down, make sure to clearly indicate which one you recognize as the template strand. Remember that a complete answer indicates orientation as well.
    b. Where is the negative charge when running the gel? Either 'top' or 'bottom' should be your answer. Make sure to explain your reasoning.

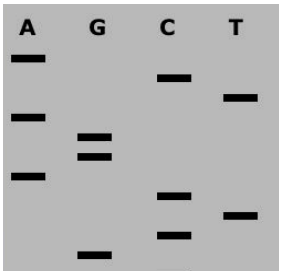


Figure 1. Gel electrophoresis showing the results of a Sanger Sequencing run for Sample X. The letters at the top indicate the beginning of the gel where the wells were located.

2. The Bajau live in Indonesia and are known for their extraordinary free diving ability, where they can swim as deep as 70 meters and hold their breath for up to 13 minutes. Researchers (Ilardo et al. 2018) were interested in this physiological adaptation, and they used an ultrasound to measure and examine differences in spleen size between the Bajau and neighboring populations who do not show this ability (i.e. the Saluans). They are interested in the spleen because it can contract to supply oxygenated red blood cells, important for people who dive for long periods while holding their breath. In the study, they observed that "We used these measurements to compare spleen sizes in the two populations…with the mean spleen size being higher among the Bajau. This difference was statistically significant (two-sample t test, p = 3.538e−07)." **Explain the meaning of the estimated p-value. If referring to the null or alternative hypotheses, make sure they are clearly described.**

3. Put the following steps of next generation sequencing in order (the letters in order will suffice, you don't need to write each sentence down again):

A	**Reversible chain terminating nucleotides are introduced and bind to their complementary base pairs.**	**D**	**After non-attached reversible chain terminating nucleotides are washed away, fluorescence is activated, allowing identification of the added base.**
B	**Adapters are added to ends of DNA fragments.**	**E**	**DNA strands are annealed to oligonucleotides on a flow cell.**
C	**Fluorescence and blocking groups are removed from the reversible chain terminating nucleotide, allowing more**	**F**	**DNA strands are amplified several times to create identical sets of the original DNA strand and its complementary strand.**

	nucleotides to be attached.		

4. Name three different benefits of the Human Genome Project.

Week 4

1. Define exon and intron, and indicate what step of the central dogma (there are three main steps of the central dogma) in which they are involved. In 2-3 sentences, indicate the purpose of that step in the central dogma, and what happens during that step.

2. In 2012, researchers sequenced an archaic human. They found that their raw data "provides about 31-fold (31x) coverage of the ~1.86 gigabases of the human autosomal genome to which short sequences can be confidently mapped" (Meyer et al, 2012).
    a. Explain what this sentence means, particularly the term '31-fold coverage'.
    b. Then, decide whether a final sequence determined from these data will be trustworthy for further data analysis, making sure to explain your reasoning.

3. Share the following about the different generations of sequencing.
    a. Describe **TWO differences** between first- and second-generation sequencing.
    b. Describe **ONE similarity** between first- and second-generation sequencing.
    c. Explain what makes third-generation sequencing different from the previous two.

4. In the Ice Age paper (Fu et al. 2016), there is a sentence in their supplement that says "We generated 2×75bp reads on Illumina HiSeq2500 or NextSeq500 instruments (p. 17)". This means their sequencing reads each consisted of 75 base pairs.

    Knowing this and the information from Extended Data Table 1 (slide 15 of Week 4R, I will also hand out a copy during the quiz), you should be able to determine the number of reads obtained for each ancient individual.

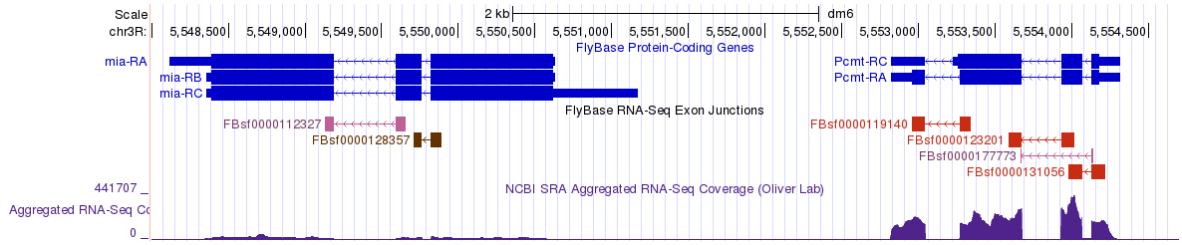
    For this question, I will name one ancient individual (from the 'Sample Code' column), and you will tell me the number of reads obtained for this individual. Show your work, including the formula you used.

Week 5

1. For the following question, I will provide an RNA sequence similar to the one below. You will draw the corresponding DNA, clearly indicate the name of each DNA strand (with regards to transcription), and then indicate the direction of transcription using an arrow.

   Example RNA Sequence: 5' UCCGAUUUGGGAA 3'

2. For the following question, I will provide an image from the Genome Browser. An example image is shown below, but it is NOT the one I will give during the quiz. For each gene in the image,
   a. Using the gene model, provide the gene symbol, the number of isoforms, and the number of exons and introns for one isoform of your choice.
   b. Indicate whether there is a difference in the protein sequence between any shown isoforms. Explain your reasoning. (If only one isoform, write n/a).
   c. Indicate whether you think the gene is expressed. Explain your reasoning.
   d. Indicate whether you think all isoforms are expressed. Explain your reasoning.



To view this image on the Genome Browser, choose the *D. melanogaster* Aug. 2014 assembly, and go to "chr3R:5,548,000-5,554,700".

3. For a research project, you decide to work on a protein called Zelda, which is expressed in *Drosophila melanogaster*. You collect processed mRNA for the *Zelda* gene from flies that are 2 hrs old, 6 hrs old, and 12 hrs old. After extracting the mRNA, you convert the mRNA to cDNA and amplify the cDNA. You then run the cDNA out on a gel and visualize the location of the cDNA after 30 minutes (Figure 1). Explain what bands at different locations mean in a gel, and then use that to help explain what is likely occurring for *Zelda* expression at each time point.

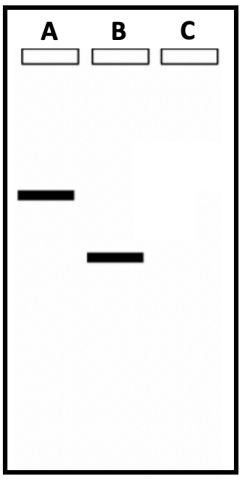
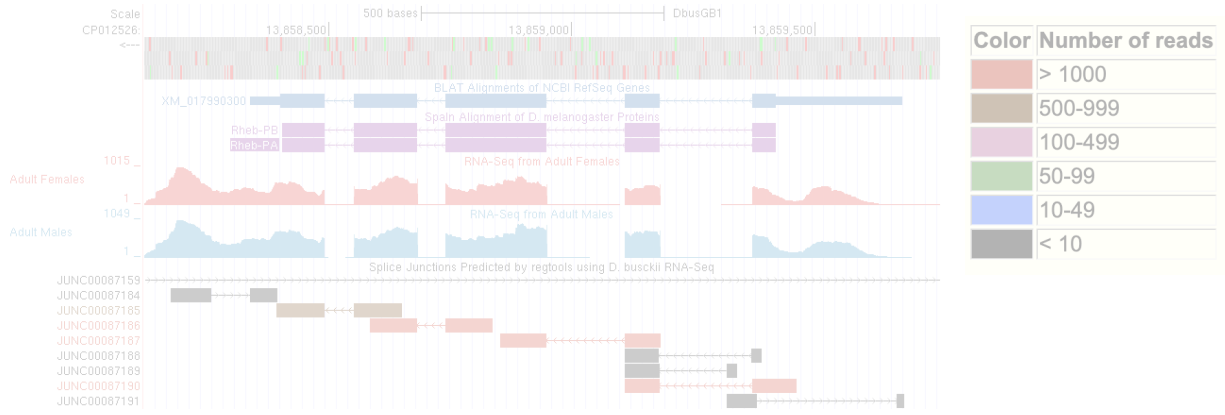


Figure 1. Gel electrophoresis results for processed mRNA for the *Zelda* gene in *D. melanogaster*. Zelda mRNA was retrieved from fly populations at (A) 2 hrs, (B) 6 hrs, and (C) 12 hrs. Amplified cDNA product was placed in wells at the top and run for 30 minutes.

4. While we are focused on mRNAs and the role they play in protein synthesis, there are several other roles that RNA can have. Name TWO of those types of RNAs and their general function in the cell.

Week 6



1. Explain how splice/exon junctions are determined, and why they are good evidence for validating a gene annotation. You may use the above example to help explain your response if helpful.

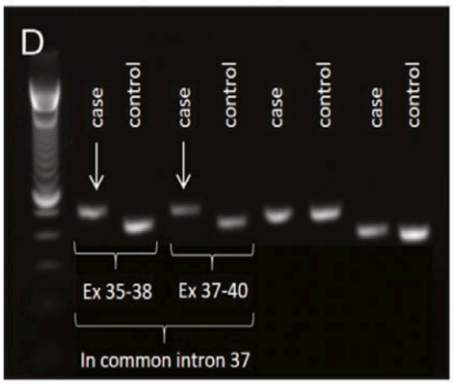
2. For the following question, I will provide a processed mRNA sequence (and a codon-amino acid table!!), for which you will determine the associated polypeptide sequence. Then, name the process in the central dogma that best describes the determination of a polypeptide sequence from mRNA. An example sequence is shown below, but it is NOT the one I will give during the quiz.

Example: 5' CACGGUCGAUGAGGUUAUAACGCCGCG 3'

3. In 2-3 sentences, describe the mutation observed in a patient with Duchenne's Muscular Dystrophy published in the study by Gonorzky et al (2015). Make sure to use the following SIX words *accurately* in your description (underline them to help me notice!): **intron, dystrophin, splice site, exon, stop codon, and translation**.

4. From Figure 1D in Gonorzky et al (2015, shown on right), explain the following:

a. The name of the method used to amplify the sequence of the original RNA from the *DMD* gene
b. The key result of the figure in **ONE** sentence
c. A clear explanation of why using gel electrophoresis would help lead to this result (i.e., explain how gels work)



Week 7

1. List all four forces of evolution we have discussed in class. Then, clearly define TWO forces of evolution (I will be picky on language, *make sure the definition of evolution is within your definition of the force*). Come up with 1-2 examples exhibiting these two forces that are not associated with any problems I have created. Make sure in your description to clearly explain **why** that force of evolution is involved (there can be two separate examples for each force, or one example involving 2 forces).

2. Define the term **fitness** as used in discussing evolutionary concepts. Then, give an example of a type of point mutation that would have NO fitness difference between the original allele and the new allele and explain your reasoning.

3. Explain why the following evolutionary misconceptions are incorrect:
   a. Individuals can evolve.
   b. Humans evolved from chimpanzees.
   c. Natural selection acts for the good of the species to give them what they need.

4. Explain why the following evolutionary misconceptions are incorrect:
   a. Evolution results in progress; organisms are always getting better.
   b. Species want to evolve.
   c. All mutations are bad.

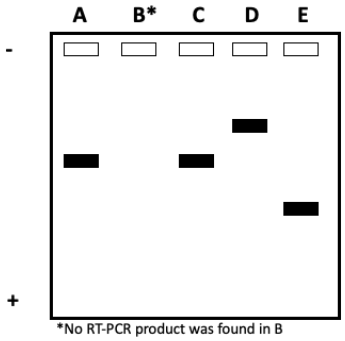
Week 8

1. For the following question, I will provide a processed mRNA sequence (and a codon-amino acid table!!), for which you will determine the associated polypeptide sequence. **Then, provide an example of a (a) silent, (b) missense, (C) nonsense, and (d) frameshift mutation.** An example sequence is shown below, but it is NOT the one I will give during the quiz.

Example: 5' CACGGUCGAUGAGGUUAUAACGCCGCG 3'

~~2. The *ADH1B* gene codes for an alcohol dehydrogenase, a protein that assists in metabolizing alcohol. An RT-PCR analysis was performed for five different people with five different phenotypes (listed below) related to this gene. Provide an example of a type of point mutation that could cause each of the following results in B-E - make sure to identify the type of mutation and why it would result in that phenotype (multiple answers could be correct). Note that A is the typical version of the ADH1B protein.~~

Removing question due to confusing language



*No RT-PCR product was found in B

Phenotypes relative to the Person represented in A
B. Very little or no metabolizing of alcohol is observed
C. Faster metabolization of alcohol
D. Low metabolization of alcohol
E. Typical metabolization of alcohol

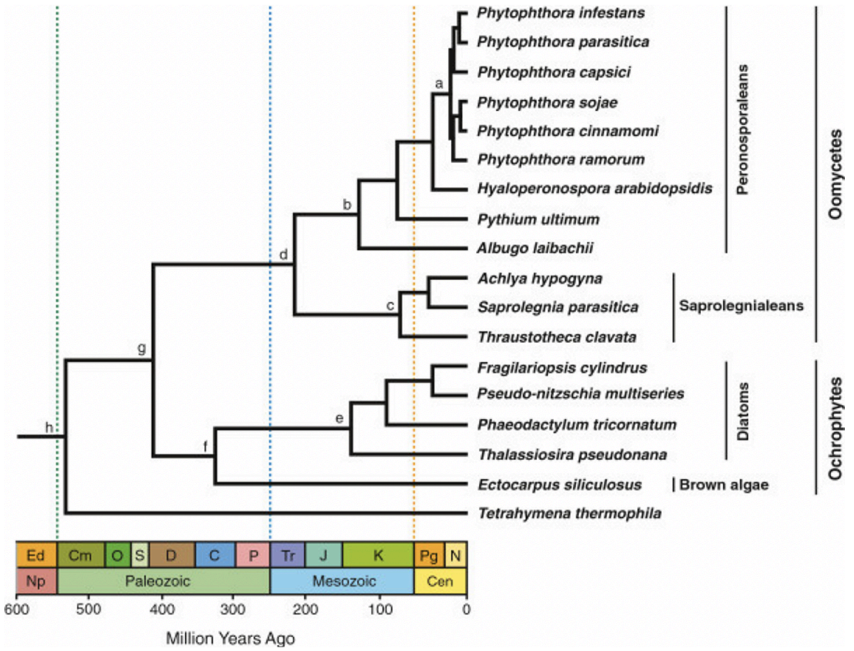
Consider the TWO scenarios described below in #3-4 and determine whether it describes evolution in process.

- **If YES, determine which two forces of evolution are *primarily* at play in the scenario. Define each force of evolution, and then explain your reasoning for why that force applies in the scenario.**
- **If NO, explain why the scenario is not describing evolution in process, making sure to clearly state the definition of evolution.**

3. In a study on humans, it is observed that those who primarily work outside tend to have darker skin color than those who primarily work inside, even after accounting for populations of different ancestry. It was found that the cause of darker skin color is related to increased production of proteins that assist in the process of forming melanin, a pigment molecule that helps to protect the body from damaging UV rays from the sun. Thus, a single person's skin color will change over their lifetime, depending on their exposure to the sun.

4. You have spent a long career studying stickleback fish living in pools of freshwater. In one pool, ~30% of fish show the presence of a dot pattern on their tail, whereas the other 70% do not. Over several generations, the dot pattern is found in ~50% of the fish in that same pool. The dot pattern was originally not observed in other pools, but over several generations, especially after periods of heavy rain mixing different pools of fish, nearby pools of stickleback fish begin to show a small frequency of fish with the dot pattern. There is no discernible effect of the dot pattern on fitness of stickleback fish.

Week 9

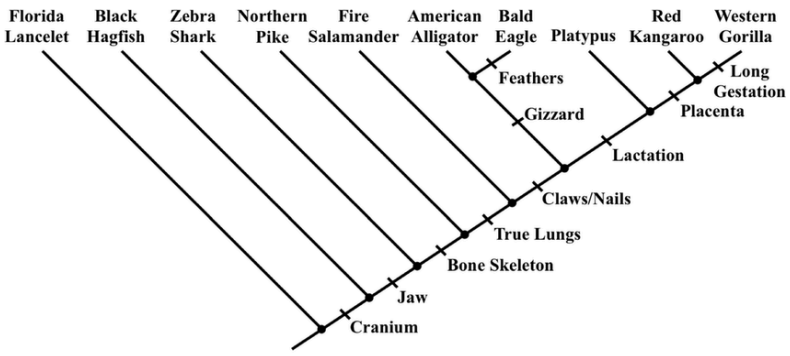


Questions 1 and 2 are associated with the phylogeny on the left.

1. I will ask you two questions on who is more closely related to who. One example shown here is the following question: who is more closely related to *A. laibachii*, *P. infestans* or *S. parasitica*?

2. *A. laibachii*, *A. hypogyna, S. parasitica*, and *T. clavate* all share the same mutation that is NOT in any of the other listed species. Give a date range when you think the mutation first arose. For the date range you choose, indicate the **<u>minimum</u>** number of times the mutation must have been gained and/or lost to see the observed mutation pattern across these species.

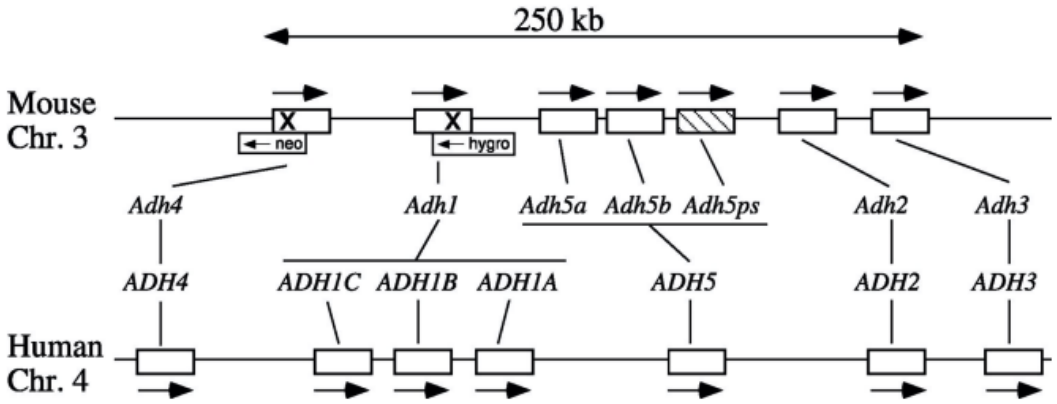
3. Based on the tree on the right, I will describe the presence/absence of three characters. You will then name one example of a species who should show that pattern. Note that the tick mark associated with a character indicates its first appearance and once it appears, it is not lost. For example, I could say: **jaw present, claws present, placenta absent**.



4. What are **<u>TWO</u>** pieces of evidence you might look for to demonstrate support for determining whether you have found a DNA region orthologous to your gene of interest? For each piece of evidence you provide, make sure to explain what the evidence is and why it can support your finding.

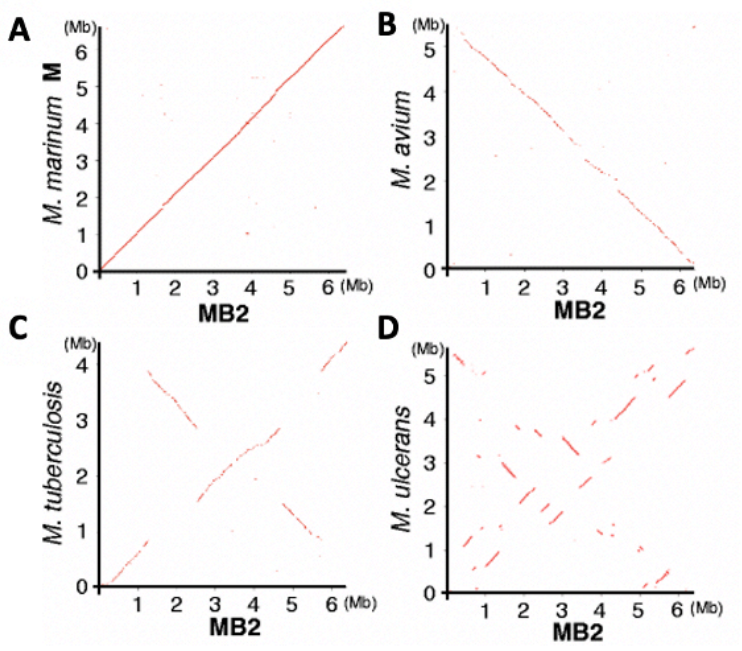
Week 10

1. Based on the following figure, clearly provide one example of a pair of **orthologous** genes, and one example of a pair of **paralogous** genes. Then, define what orthology and paralogy mean, making sure to highlight how they differ from each other.



2. You have several distantly related species, but all of them show high gene similarity for Gene *X*. Explain what is likely occurring in this scenario, using your understanding of evolutionary forces. If you use just 1-2 terms to describe what is occurring, make sure to define the term(s) sufficiently.

3. Consider that a biologically female individual does not have red-green colorblindness but carries the associated allele. This individual has a child with a biologically male individual who does not have red-green colorblindness. What is the chance the child has red-green colorblindness? Show your work.



4. The left figure compares an isolated bacterial strain (MB2) with four mycobacterial species. Answer the following two questions using this figure.

i. Which species is most closely related to MB2, and which species is least closely related to MB2? Explain your reasoning using your understanding of dot plots.

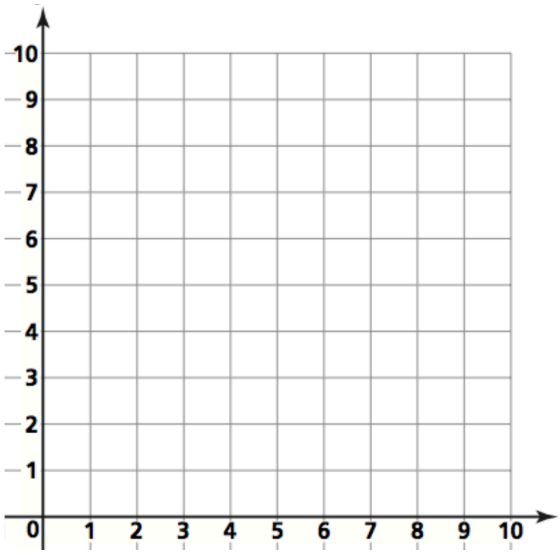
ii. What type of mutation is best represented by subfigure **B**? Explain your reasoning.

Week 11

1. For the following question, I will provide two amino acid sequences and a graph.

Draw a dot plot with **Sequence A on the x-axis and Sequence B on the y-axis (make sure your axes are labeled and have appropriate units)**.

An example pair of sequences is shown below, but it is NOT the one I will give during the quiz.
Sequence A: QLKFWMVN
Sequence B: MEQLMWFKVN

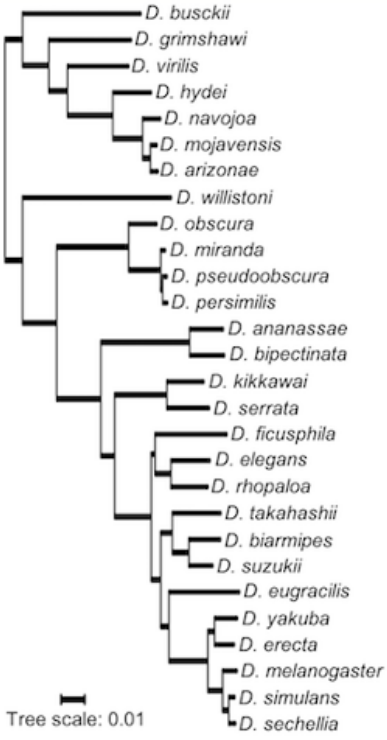


2. You have several distantly related species, but all of them show high gene similarity for Gene *X*. Explain what is likely occurring in this scenario, using your understanding of evolution. If you use just 1-2 terms to describe what is occurring, make sure to define the term(s) sufficiently.

3. You are a researcher studying a gene family consisting of two genes, Gene A and Gene B, in three species - wolves, dogs, and jackals, where wolves and dogs are more closely related to each other. These two genes are found in all three species. **Draw a tree showing what you would expect to observe if the duplication occurred once in the common ancestor of wolves and dogs, AND once in the ancestor of jackals.**

4. We have been annotating genes in the insulin signaling pathway and comparing these genes for different species to the ortholog in *D. melanogaster*.

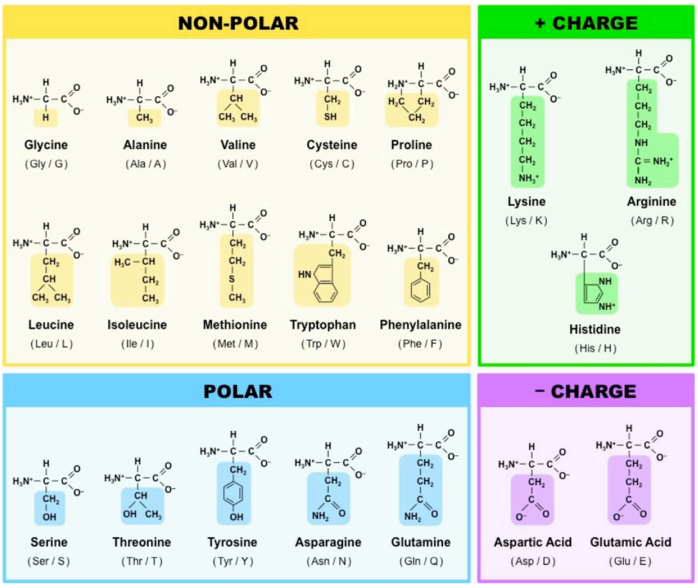
Consider for a Gene X that you examined two dot plots, one comparing *D. yakuba* to *D. melanogaster*, and another comparing *D. willistoni* to *D. melanogaster*. You find that *D. yakuba* shows many more mutations relative to *D. melanogaster* than *D. willistoni* shows. Use the phylogenetic tree and your understanding of mutational processes and evolution to explain what might have occurred at Gene X regarding selection processes.



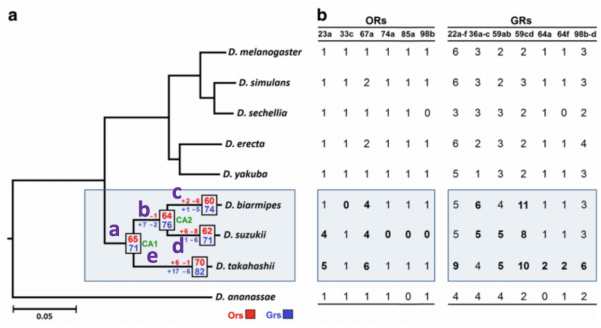
Week 12

1. What type of amino acid are you most likely to find on the inside of a protein? Explain your reasoning.

2. For the following question, I will provide a protein and indicate a mutation that has resulted in an amino acid change. Provide a hypothesis on whether the mutation has altered the function of the protein. Then, using your knowledge of amino acid interactions and the provided amino acid table, explain your reasoning for your hypothesis. An example of a protein and mutation is shown below, but it is NOT the one I will give during the quiz.



Example: ADH1B protein, ARG48HIS (Arginine mutated to Histidine at position 48 of the amino acid chain)



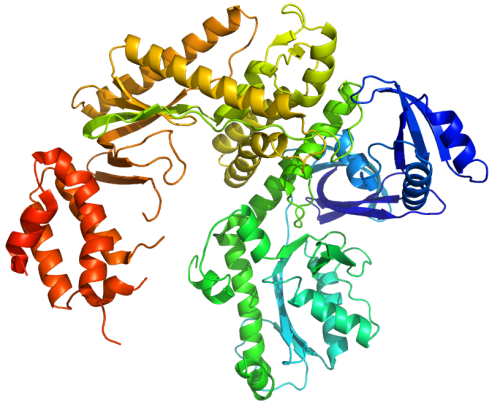
3. I will pick **one gene** from Fig. 4 from Hickner et al (2016). Develop a scenario consistent with the # of gene copies shown that indicates (a) the # of gene copies possessed by the MRCA of *D. biarmipes*, *D. suzukii*, and *D. takahashii* AND (b) when on branches a-e that gene loss/duplication occurred. Your scenario need not be the most parsimonious, but more highly parsimonious scenarios are better hypotheses.

4. We've talked a lot about selection acting on a gene. Consider a scenario **comparing the processed mRNA sequences corresponding to gene *β* (beta)** in three species (humans, chimps, and mice), where humans and chimps are more closely related to each other than they are to mice. I will pick one of the three hypotheses: (i) no selection in any species, (ii) negative selection in all three species, or (iii) positive selection acting on the chimp ***β***-gene. For the chosen hypothesis, provide the following predictions:

A. The expected dN/dS value (or range of values) for each species.
B. How the percent identity between different pairs of protein sequences might look for chimp vs human and mice vs human.
C. Two dot plot scenarios, one for chimp vs human and one for mice vs human (label axes appropriately, with units) - the types of mutation you draw do not matter, but I expect the conclusion based on the dot plots would clearly correspond to the chosen hypothesis

Week 13

1. For the following question, I will provide an image of a protein in its folded shape. I will ask you to:



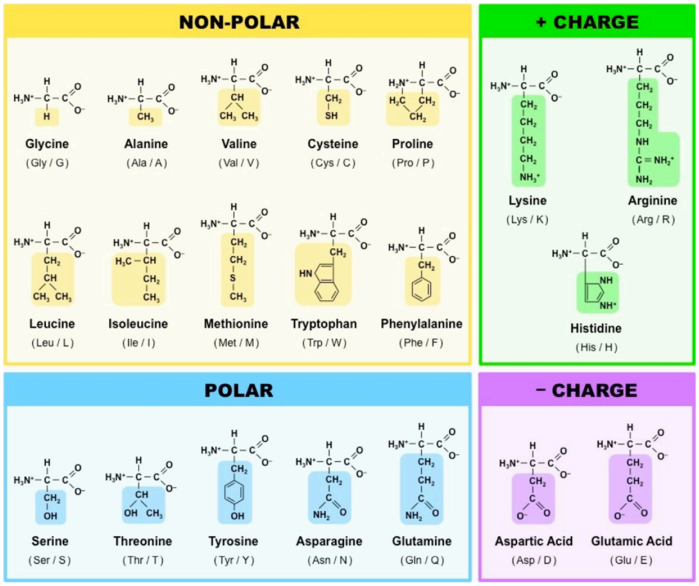
   A. Circle an example of an alpha-helix and a beta-pleated sheet and label each circle. If one is not present, then say so.
   B. Indicate the level of protein structure you are describing and the groups on the amino acid involved in this interaction.
   C. Indicate the type(s) of bonds formed between these groups.

2. In your own words, explain what mutation in what protein causes sickle cell disease (be specific!). Then, explain how that mutation makes a red blood cell take on a sickle shape, focusing on what happens to the related protein(s). Note that your response should refer to interactions with oxygen as well.

3. For the following question, I will provide **<u>THREE</u>** pairs of amino acids with an amino acid table like the one shown on the left. Name the most likely bond/interaction that will occur between these two amino acids if they were to interact with each other. If no bond/interaction occurs, state that.

An example pair of amino acids is shown below, but it is NOT the one I will give during the quiz.

E.g. Methionine and Alanine



4. (For Week 14) Of the four protein types we defined in Wednesday's lab **(kinase, phosphatase, transcription factor, receptor)**, indicate (i) which types can be assigned to the protein to which insulin binds (note more than one answer is needed!). (ii) For each of these types, explain their general function. (iii) Then, explain in your own words how this protein carries out these functions within the insulin signaling pathway after insulin binds.

This marks the last week of new CQs!