

1. Compare the % sequence divergence observed in two genes in a pair of salamander species. The histone protein requires a precise structure to function properly, while the structure of the hormone receptor gene depends on only a few key amino acids. You can assume that most of the observed variation in these genes has no effect on fitness.
 - a. Fill in the blanks below with $>$, $<$ or $=$ to indicate the relative divergence expected in these comparison.

3 rd codon positions in the histone gene	$>$	1 st codon positions in histone gene
3 rd codon positions in the histone gene	$</=$	3 rd codon positions in the hormone receptor gene
1 st codon positions in the histone	$<$	1 st codon positions in the hormone receptor gene

- b. Explain the reasoning that you used in filling in the blanks above.
 1. **The histone (H) gene is under negative selection, so you would expect more silent mutations than replacement mutations to be fixed. Since changes in the third codon position are more likely to be silent, more of them will be fixed, resulting in more divergence in those positions 2 and 3.**
 - 2 and 3. **The entirety of the H gene is under very strict negative selection, while only part of the hormone receptor (HR) gene is under strict selection. Because of this, you would expect replacement changes to be neutral in some of the HR gene sequence. Thus more 1st codon position changes would go to fixation in the HR gene sequence (3rd part). Since some 3rd codon position changes are replacement changes, you would expect to see slightly more 3rd position changes in the HR gene sequence, but an equal number was also accepted (2nd part).**
- c. Suppose that you discover a gene in which most mutations that change the protein sequence increase fitness. How would you expect the rate of divergence at 1st codon positions in the gene under positive selection compare to the rate of divergence at 1st codon positions in the histone and hormone receptor genes? Why?

You would expect the most divergence in the gene under positive selection, then the HR gene, and then the H gene. This is because in a gene under positive selection, changes in the protein sequence are beneficial, so they go to fixation a lot faster than neutral changes, and much much faster than negative changes, thus leading to more divergence.

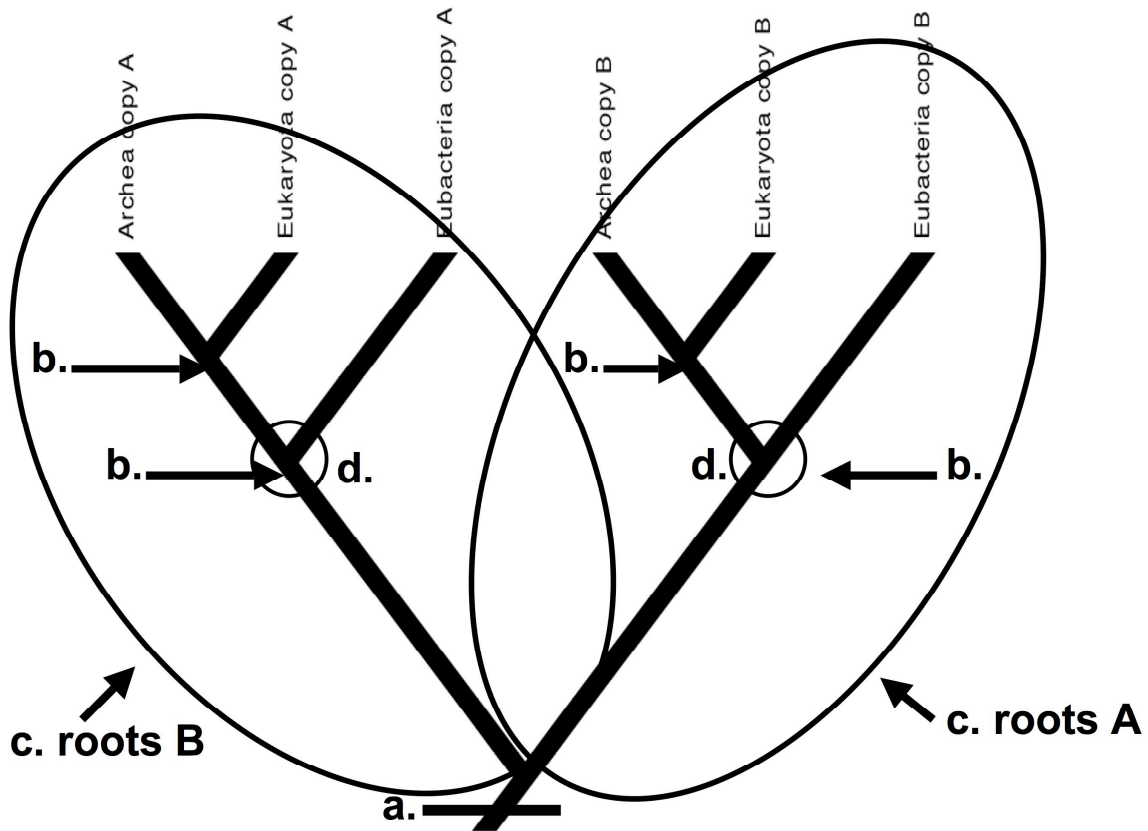
- d. How would you expect the rate of divergence at 1st codon positions to compare to the rate of divergence at 3rd codon positions in the gene under positive selection? Why?

More at the 1st codon position, see 2c for explanation

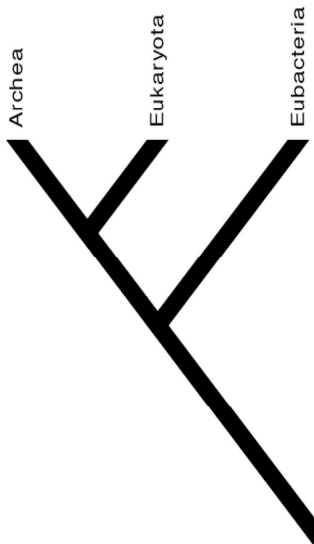
2. Why is the rate of fixation of neutral mutations (μ) independent of population size?

It is independent of population size because the rate of fixation = rate of production ($2N\mu$)*probability of fixation ($1/2N$). The $2N$'s cancel out. To clarify: In small populations the rate of production is small, but the probability of fixation is large. In large populations the rate of production is large, but the probability of fixation is small.

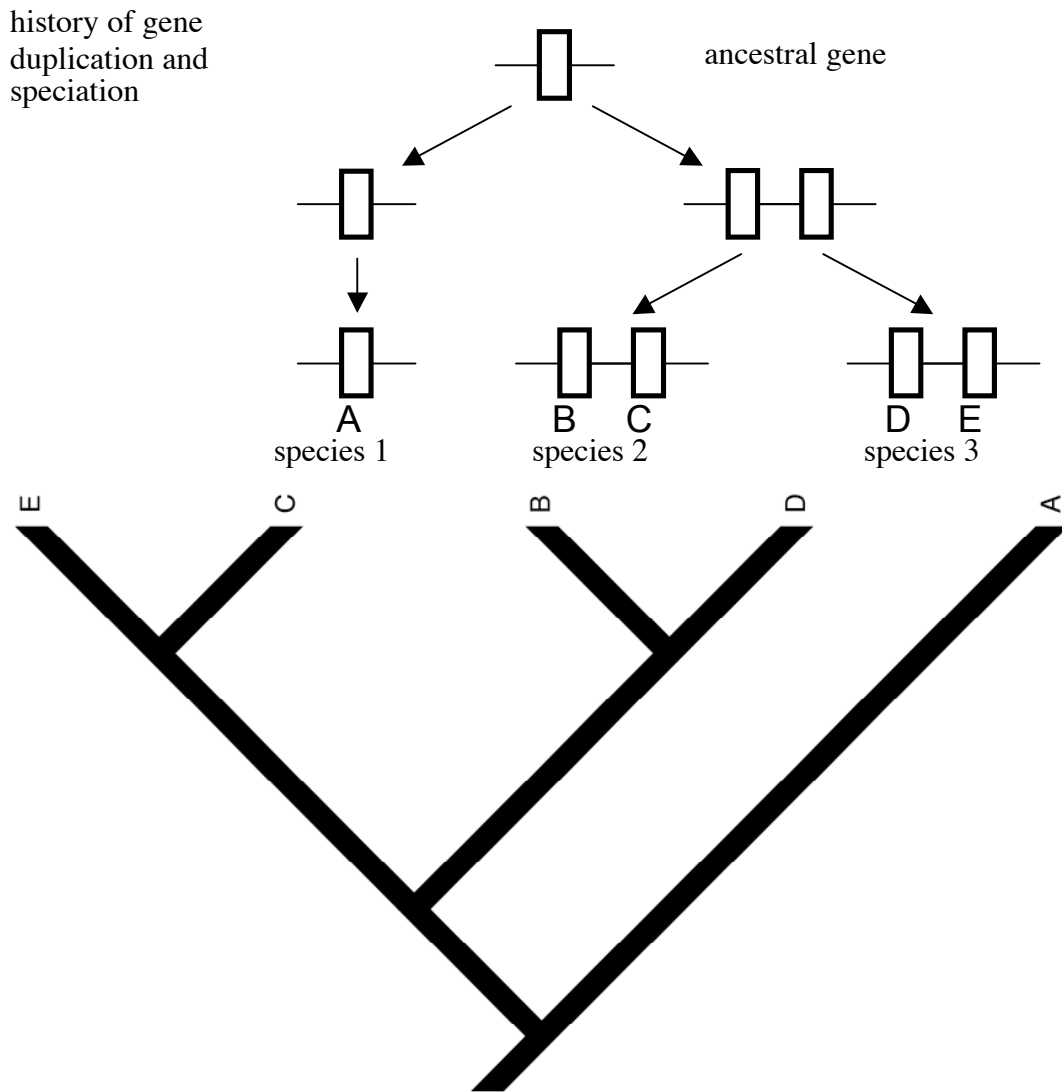
3. In order to root the tree of life an outgroup is required. Obviously, the entire tree of life isn't going to have an outgroup; all life is part of the ingroup. Scientists circumvented this problem by using duplicate genes that are present in all three major domains of life. The following questions refer to the tree below.
- a. Label on the tree where the duplication event took place that produced copy A and copy B.
- b. Label on the tree where speciation events occurred.
- c. Circle the part of the tree that is acting as the root for gene copy A. Circle the part of the tree that is acting as the root for gene copy B.
- d. Circle on the tree the most recent common ancestor of gene copy A. Circle on the tree the most recent common ancestor of gene copy B.



e. The tree below is a phylogenetic tree showing the relationships of different genes to each other. Based on this gene tree what would the phylogenetic tree depicting the relationships between these three taxa look like?



4. The diagram below shows the evolutionary history that generated five copies (labeled A-E) of a single ancestral gene in three species.



5. Although Samoan fruit bats (*Pteropus samoensis*) and Anetian fruit bats (*P. anetianus*) are more closely related to each other than either are to Tongan fruit bats (*P. tonganus*), the protein sequence of the Pap2L gene in *P. anetianus* is more similar to *P. tonganus* than it is to *P. samoensis*. Mutants for the Pap2L gene in *P. samoensis* are unable to detect papaya groves when foraging at night, whereas *P. samoensis* individuals with normal Pap2L can detect papaya from long distances after dark.

Most people had a very hard time with this question. It was completely analogous to the example given in class, so reviewing your notes from 4/14 should help to clarify

- a. Identify one prediction of the neutral theory that is violated in this scenario.

The one prediction that is violated by this scenario is that the amount of divergence will be proportional to the amount of time past. The *P. sam.* is sister to *P. anet.* (and so less time has passed since these taxa separated) but *P. anet* and *P. tong.* had more similar Pap2L sequences.

- b. What might have caused Pap2L divergence in *P. samoensis*, and what sort of evidence would you use to support your claim? (Think about, but don't talk about, our discussion of the FoxP2 gene)

There are four reasons why there might be increased divergence. In this scenario, two are possible. Lets go through all four and see which you can rule out.

- 1. More time. This is not possible because *P. sam* is the same age as *P. ana*, they are sister species.**
- 2. Higher overall mutation rate. If this were the case, there would be more silent divergence in addition to the replacement divergence. We have not shown that the silent divergence is less between the more closely related taxa, so this is a possible answer.**
- 3. Relaxed selection. If this were the case, a high proportion of the mutations would have no fitness effect, but we told you that mutants for the Pap2L gene cannot find the papaya trees in the dark, thus they cant eat, thus they are unfit.**
- 4. Positive selection. If we can show that the other three are not true, then we can conclude that it is positive selection. We could also conclude that it is positive selection if we know that the ratio of the rates of replacement substitutions/silent substitutions is >1 .**

Answers were scored correct if they identified one of the two possible correct explanations, and showed how they would go about defending them.

6. . Lynn Margulis put forth the endosymbiotic theory in 1970 in her important work, "The Origin of Eukaryotic Cells", providing a plausible explanation for the existence of organelles in eukaryotic cells.

- a) Provide two examples of endosymbiotic events and describe the evidence that supports the validity of the occurrence of each.(5 pts.)

The endosymbiotic theory suggests that eukaryotic organelles, such as mitochondria and chloroplasts, were once free-living bacteria that were engulfed by proto-eukaryotes. It is hypothesized that mitochondria evolved from a free-living bacterial form known as purple bacteria, and that chloroplasts evolved from a free living photosynthetic cyanobacteria. The following evidence supports these endosymbiotic events: both of these organelles have their own circular DNA in addition to the eukaryotic DNA in the nucleus of the eukaryotic cell, chloroplast DNA and mitochondrial DNA are more similar in sequence to

bacterial DNA than to other eukaryotic DNA and, both of these organelles have at least a double membrane, suggesting one membrane is the original membrane of the free-living bacteria, and the second membrane is from the vacuole of the host cell.

b) What basic advantages does this type of mutualistic association provide for each of the partners involved? (5 pts.)

Mutualistic associations require that both parties partners in the association receive benefits from the relationship. Endosymbiosis provides an advantage for the free-living bacteria because they are provided with a protected environment in which to live. In turn, the host cell colonized by mitochondria-like free living bacteria, benefit from the energy provided by aerobic respiration, and those colonized by chloroplast-like free living bacteria benefit from the photosynthetic capacity of these organisms which allows them to produce sugars from sunlight.

7. While investigating the evolutionary dynamics of a particular gene in a lineage a researcher discovers that synonymous and non-synonymous sites are evolving at close to the same rate. This surprised the researcher because this gene is known to have an important function and mutants for this gene normally have low fitness.

a. Why is the researcher surprised that synonymous and non-synonymous sites are evolving at similar rates?

Because the gene is known to have an important function with mutants normally having lower fitness, the researcher would expect this gene to be under stabilizing selection. While under purifying selection, synonymous sites are expected to evolve at a faster rate than non-synonymous sites because non-synonymous changes don't change the amino acid sequence of protein.

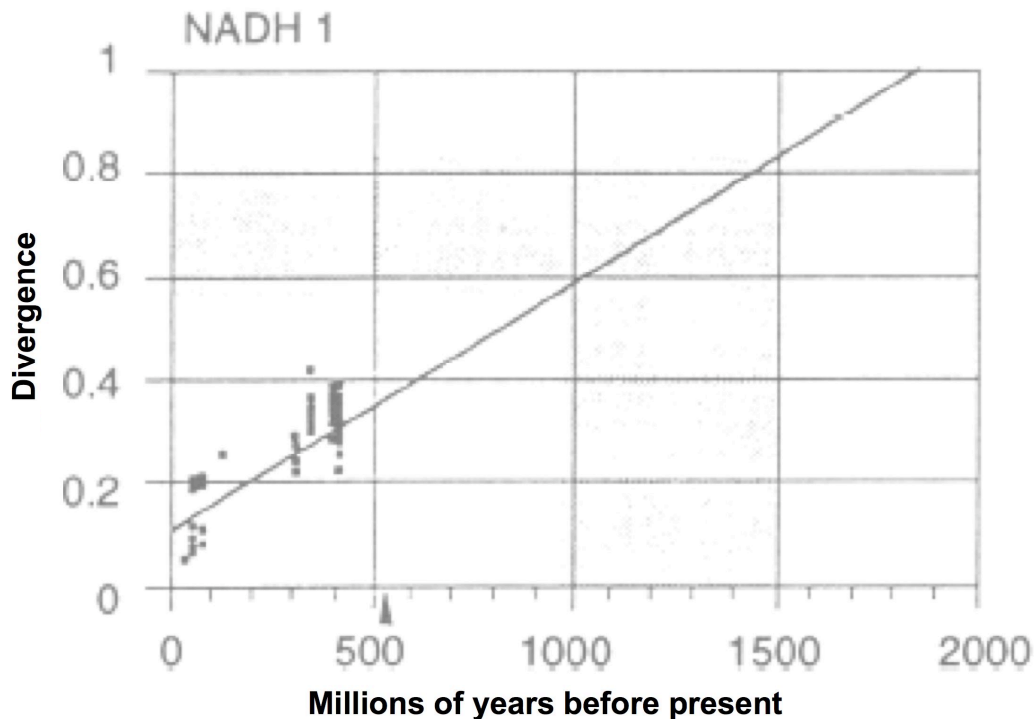
b. What is the expected dN/dS ratio for a gene under purifying selection?

<1

c. What factor could account for the low dS rate in this case? Explain.

Assuming this gene is under purifying selection, which research suggests is true, one way to account for the similar substitution rate between non-synonymous and synonymous sites is codon bias. If in this lineage there is extreme codon bias so that only one codon is normally associated with each amino acid, then there can be purifying selection acting on synonymous sites which could account for the similar rates of change.

8. **Ugh! I forgot to tell you the divergence between protostomes and deuterostomes. It's 0.65. Now you can do part a.** In 1996 researchers attempted for the first time to use the molecular clock to estimate when protostomes and deuterostomes diverged from each other. The graph below represents a molecular clock based on the gene NADH 1 and calibrated using known divergence dates based on the vertebrate fossil record.



- a. How long ago does this molecular clock suggest protostomes and deuterostomes diverged from each other? **~ 1,200 mya.**
- b. The fossil record, on the other hand, suggests that protostomes and deuterostomes diverged no earlier than 600 million years ago. The molecular clock method predicts that the amount of sequence divergence is equal to the rate of sequence change \times time since divergence. Assuming that protostomes and deuterostomes did diverge from each other within the last 600 million years, how might you explain the estimates derived using the molecular clock approach? Clue: think about how they calibrated the clock. How might this bias the results? **This molecular clock was calibrated using the vertebrate fossil record, which is well resolved. But, there is reason to believe that molecular evolution is occurring at a slower rate in vertebrates than in invertebrates. If this is true then estimated dates would be earlier (ie. further in the past) than actual divergence dates.**
9. Three hypotheses have been presented regarding the relationship of the Ediacaran fauna to modern fauna.
- H1: The Ediacaran fauna include ancestors or close relatives of modern phyla.
- H2: The Ediacaran fauna represent a separate experiment in animal life, but are closely related to animals.

H3: The Ediacaran fauna are only distantly related to animals and shouldn't be considered part of the animal kingdom.

Two representatives of Ediacaran organisms are shown below. The first, *Fractofusus andersoni*, shows bilateral symmetry. The second, *Tribrachidium*, shows 3-fold symmetry.



Fractofusus andersoni



Tribrachidium

For each hypothesis, place *Fractofusus* and *Tribrachidium* on the tree somewhere that would support the hypothesis. Two important innovations in the animal kingdom have been included on the tree.

There are multiple possible answers for this question.

For H1 if these organisms are closely related to modern phyla, than they must fall within the animal part of the tree. *Tribrachidium* may be sister to the Echinodermata. *Tribrachidium* has 3-fold symmetry, which may represent the ancestral condition of Echinodermata. *Fractofusus* shows bilateral symmetry so might fall within the clade of bilaterally symmetrical animals.

For H2 the Ediacaran organisms are closely related to animals, but are a separate but aren't closely related to modern phyla. This hypothesis suggests that the Ediacaran organisms are the sister group to the rest of the animals or possibly nested somewhere between sponges and the rest of the animals.

For H3 the Ediacaran organisms are not closely related to animals. For this hypothesis Ediacaran organisms may be the sister group to either plants or fungi or possibly diverge from the base of the tree. They could also be sister to the fungi/animal clade.

