

The Process of Molecular Phylogenetics Answer Key and Instructor Heuristics

Before the Exercise Begins: 0:15 – 0:25

- Start with a brief primer on phylogenetics. It is similar to cladistics and has the same goal – determining the shared ancestry of related species. The difference is that is no concentration on “features” or anatomy or fossils. The analysis is purely molecular – DNA sequences, protein sequences, etc.
- Cladistics and Phylogenetics are not in competition – they work together. The multiple approaches complement each other and scientists in different fields learn from one another.
- Give a brief lecture on the day’s topics based on the student handout – DNA sequences, pseudogenes, protein sequences, chromosome structure, etc. No need to spend much time – they can and should read their handout for more context. Ideally, they will have read the handout before they come to class.

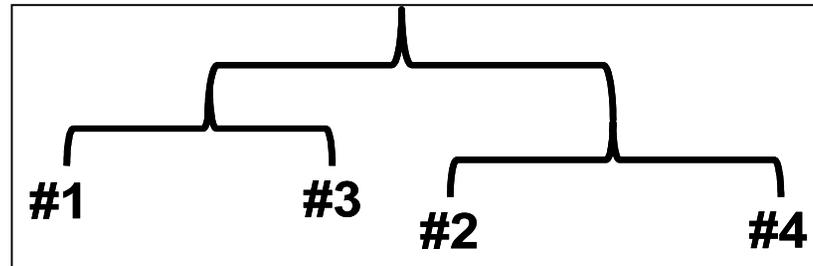
Exercise #1:

- Before setting the students loose on the activity, lead a small discussion of what a “pseudogene” is and how a gene could become “broken” through random mutations. Here are some good questions to pose to students:
 - Q:** If the GULO gene is needed for the synthesis of Vitamin C, and Vitamin C is so important to life, how would the mutation destroying this gene have been tolerated? In other words, why would it not have been quickly eliminated from the population by natural selection?
 - A:** If the mutation occurred in an individual or population that already had an abundance of citrus fruit in their diet (say, in Africa, or other tropical regions), there would be no real disadvantage of a “broken allele” for the “vitamin C gene.” With no selection against this allele, the DNA would persist in the population as a pseudogene.
 - Q:** Can you name a species with a functional GULO gene and some with a nonfunctional version? How can you tell?

- Thus, if they were to build a phylogenetic tree, the rule of parsimony (without doing the full calculations) would dictate that the divergence between the ancestor of #1 and #3 and the ancestor of #2+#4 occurred before the further divergence of these species. And the hypothetical tree that they draw should reflect that.
- Another way to simply for them – tell them to ignore all mutations that exist in just one species. This is because this one obviously occurred after it diverged from the others and is thus not helpful for retracing shared ancestry with other species. However, this is an opportunity to discuss the “molecular clock;” the number of mutations being useful for determining the time that has passed since the divergence of two species.

(I couldn't find the gorilla GULO gene (must not be sequenced))

- Thus, the phylogenetic tree should look like this:



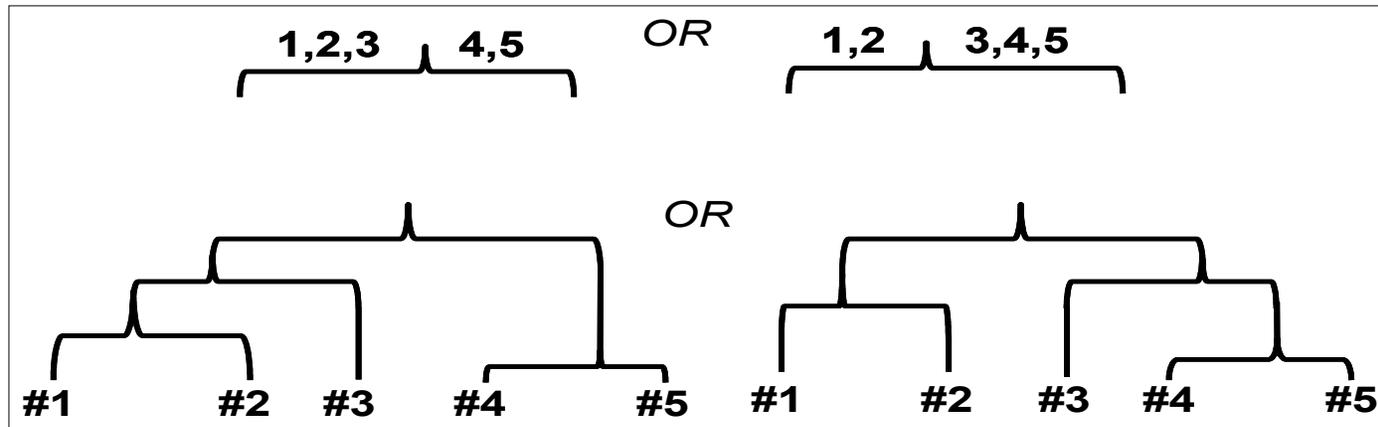
- **BEFORE MOVING ON TO EXERCISE #2** - have them hand in the tree that they draw, with whatever essential math or explanation is necessary to defend why they drew it that way.
- The major point to discuss here is that this is just a short stretch of DNA, less than 100bp, and still, the relatedness of the four species can be deduced. Molecular evolutionary biologists generally analyze *millions* of basepairs of DNA sequence to help them retrace ancestry.
- It might even be worth telling them (and this is no lie or exaggeration) that this particular stretch of sequence was picked pretty much at random. The gene was chosen on purpose (a pseudogene), but the person who designed this activity (N. Lents) did not try out a bunch of sequences and then choose the stretch that “worked best” for this activity. This stretch was the first one he selected and although perhaps others might work even better and show the ancestry in an even more obvious way, he decided to keep this “first attempt,” as a demonstration that the molecular evidence for shared ancestry is not hard to find!

- Code: But do not break this until the very end!!!!
#1 = *Pan troglodytes* (chimpanzee) #2 = *Pongo pygmaeus* (orangutan)
#3 = *Homo sapiens* (humans) #4 = *Macaca mulatta* (macaque or Rhesus monkey)
- One idea for the final wrap-up discussion, if there is time, is to do a BLAST search, or genome browser search for the human sequence and show how the human genome sequence is now fully available and searchable online. Just a short demo of bioinformatics can really “wow” them.

Exercise 2:

- Before setting the students loose on the activity, lead a small discussion of how using a protein sequence is different from using a DNA sequence. Protein is the expressed product of a gene. There could be lots of silent mutations lurking in this gene, which do not result in a change to the protein. Here are some possible discussion questions.
 - Q:** Why is a protein sequence more evolutionarily “constrained?”
 - A:** All mutations are random and thus, most mutations in a functioning gene would diminish or disrupt the function of the gene product. In most cases, a gene product with diminished function would reduce the viability or health of the organism. Thus, most of these mutations would be quickly removed from the population through the death of the organism. However, silent or conservative mutations would be tolerated.
 - Q:** Can mutations ever actually *improve* the function of the gene and help the organism survive?
 - A:** Yes! While this is rare, occasional advantageous mutations are the basis of adaptation! Such are often called “gain of function” mutations because they enhance the function of a gene, or even give a new function or property to a gene.
 - Q:** So wait, if many mutations are silent or neutral, how do they result in a permanent change in all members of a certain species? How do neutral alleles take over an entire population?
 - A:** Remember that, during speciation, the “founders” of a new species are generally a very small group of individuals. In a phenomenon called “genetic drift,” these individuals might have combinations of “neutral” alleles that are not necessarily reflective of the larger population. Thus, some neutral alleles, even rare ones, could become fixed in the new population only because it happened to be present in the founders, not because they give any adaptive advantage to the population.

- Putting all these together, the students will get one of the phylogenies below:

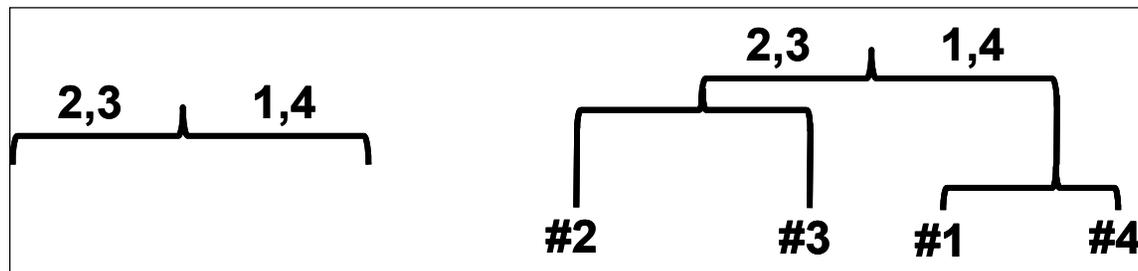


- Once again, remind them that this is just a short stretch of protein sequence, but biologists would analyze the sequence of hundreds of proteins, as more data becomes available.
- Once again, you can tell them that this protein was picked pretty much at random, as one whose protein sequence was easy to find online for all these species. We did not try out a bunch of proteins and then choose the one that “worked best” or gave us the result we wanted for this activity.
- Code: But do not break this until the very end – when you are doing the final wrap-up!
#1 = *Homo sapiens* (humans) #2 = *Pan troglodytes* (chimpanzee) #3 = *Gorilla gorilla* (Gorilla)
#4 = *Pongo pygmaeus* (orangutan) #5 = *Macaca mulatta* (macaque or Rhesus monkey)
- Again, you could do a protein BLAST search with one of the stretches that is long enough, to show them the databases. You could even use one of the nonhuman sequences and show how it found the “best hit” (human) but marked the differences.

Exercise 3:

- In this activity, the comparisons of chromosome patterns are much more subjective, and it’s more likely that they will have a hard time making the big separations and instead see the changes as more of a gradual transition from one to the other.

- When searching for the homologous chromosomes (shared by all four), there will be a common tendency to match up, then stick with, regions of only partial homology. You may have to constantly remind students that for chromosomes to be considered true homologues, they must share significant homology throughout the length of the chromosome, and for this exercise, we are only interested in homologues shared by all four species.
- Students will also notice that one species has one fewer chromosome, to which we respond that a) this is not important for building the phylograms, and b) it is true that species #4 indeed has one fewer chromosome than the other three species (although all four have many more chromosomes than are shown here).
- Once the students have found the three homologous chromosomes shared by all four species, they can proceed with the attempt to dissect the ancestry of the four species. This will not be a trivial task for them. But if they take each chromosome set one at a time, they point to the same relationship: #1 and #4 are considerably closer to each other than either is to #2 and #3.
- However, before simply branching the two sides into symmetrical final branches, give a new challenge:
 - Which of the two pairs is more similar, which are more different? How could this be communicated in the phylogram?
 - The conclusion of this is that #2 and #3 have been evolving separately from each other for a longer period of time than #1 and #4. Thus, they should get something like this for their phylogenetic tree:



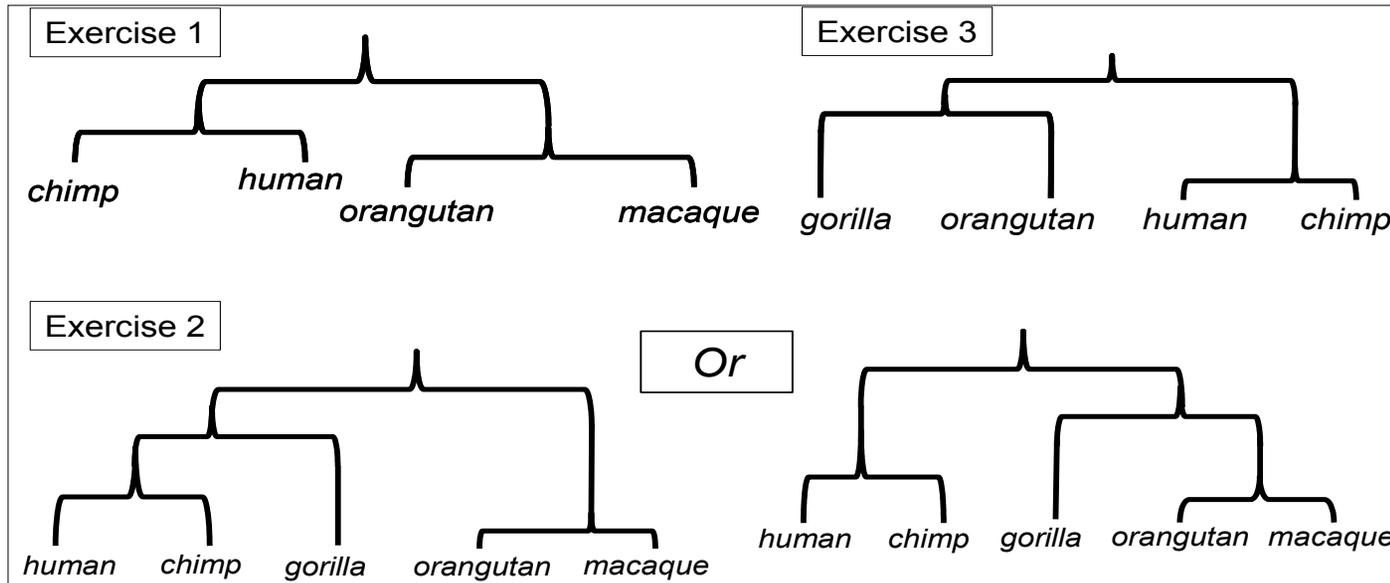
- There will almost certainly be some in the class that will notice what is going with the “leftover chromosomes.” The second-to-largest chromosome in species #4 (humans!) has no true homologue in the other species.
- However, if you take the smallest two chromosomes of the other three species (this works best with #1, chimps, but you can see it with any of them), and fuse them together (one upside down compared to the other), the resulting fused chromosome... viola!! It looks almost identical to that lone second chromosome #2 of species #4.

- I would only reveal the following points if some student has already heard about this and begins to speak openly about it... (otherwise, do it at the end)
- The reality here is that humans have 23 pairs of chromosomes, while all other great apes have 24. Where did this missing chromosome go in the *Homo sapiens* lineage? Nowhere! Our chromosome #2 is the result of the fusion between two other chromosomes in a human ancestor some time after the divergence from our most recent common ancestor with chimpanzees, our closest relatives.
 - Explain that this theory is strongly supported by extensive DNA evidence, such as the presence of two telomere-like stretches arranged end-to-end within chromosome #2 and the remnants of an additional centromere.
 - Interestingly, some anthropologists propose that this chromosomal fusion event contributed to the reproductive isolation of the human ancestor population from the ancestor population of the modern chimpanzee, solidifying the unique genetic divergence of our ancestors
- If this discussion happens now, rather than after the end, go ahead and break the code for activity #3:
 - #1 = Chimpanzees; #2 = Orangutan; #3 = Gorilla; #4 = Humans;

Final Discussion

- You should have collected the students' phylograms and notes for each exercise as they go. Have the CLT make photocopies of these and keep a copy of these. (Prof. Lents is taking data on this, so he may want these copies.)
- At this point, hand the papers back to the students, but tell them NOT to correct or change anything (they will not be graded on getting them "correct.") Now you can "break the code" and have the students re-write the phylograms on these sheets, but instead of the unnamed species numbers, have them write the actual species names on the three phylograms from the three exercises. Now they can compare the results from the three exercises to see if they agree.

- This mimics the multi-disciplinary nature of biology. Some scientists work with DNA sequences. Others study anatomical structures. Scientists all over the world study the biochemistry, nutrition, lifestyle, behavior, and physiology of all the biodiversity on the planet, past and present. These “multiple lines of evidence” all cooperate to help scientists retrace the ancestry of species. Students’ phylograms will look something like this:



- One neat thing to point out to them – the results of exercise #3 help to clarify the dispute in exercise #2. Even though there was, perhaps, slightly more data supporting gorillas being closer to humans and chimps, the chromosomal maps argue the opposite. This is scientific controversy! Conflicting results tell a slightly different story, but the more data that is gathered, a more clear picture emerges and the scientific community incremental advances toward “the truth.”

More Discussion Points:

- The exercises above involve REAL data, real DNA sequence, etc. Nothing was fabricated. These examples were chosen largely at random, NOT because they happen to uphold current scientific dogma.
- You can use BLAST searching or one of the genome browser to demonstrate a) the validity of the data, and b) the bioinformatics tools available to scientists.

- The sequence comparisons of this exercise involved only a short stretch of data. Evolutionary scientists use tens of thousands or hundreds of thousands of DNA base-pairs, or even whole genomes to conduct their research, which can take years. In general, the more data one considers, the more likely they are to “get it right” and capture the true ancestry.
- There is no one complete and correct phylogenetic tree. They are made to convey relationships between a few species, even many species.
- As more fossilized remains and DNA sequences become available, our knowledge about the evolutionary history of life on earth becomes more complete. Although phylogenies are considered “hypotheses,” some are supported by so much data that scientists have a great deal of confidence that they are correct. Occasionally, existing phylogenetic trees have to be refined in light of new data. Even more rarely, new data challenges a currently dominant hypothetical phylogeny. In this way, the collaborative, skeptical, and self-correcting nature of scientific research continually advances our knowledge.
- Shown here is the currently accepted phylogenetic tree that shows the relationships among the five genera considered today. It is worthwhile to put this on the board so that they can compare their results to this.
- Importantly, most students would have reached the same conclusion as shown here! Further, the diagram here supports the conclusion drawn in activity #3 that Gorillas are more closely related to humans and chimps than they are to Orangutans, but only slightly.

