

The Process of Molecular Phylogenetics

I. **Exercise #1 – Molecular phylogenetics using a pseudogene**

- Below are four gene sequences. These are taken from four animals that are believed to have “recent shared ancestry” (closely related).
- The gene sequences are from a so-called “broken gene” or pseudogene, the evolutionary remnant of a gene, which is now nonfunctional, in a given species or group of related species. In this case, the gene is called GULO (L-gulonolactone oxidase), which codes for the enzyme which catalyzes a key step in the synthesis of ascorbic acid (vitamin C). Along the way, some animals have lost the function of this gene (by random mutation) and must consume vitamin C in their diet. If there had been an abundance of citrus fruit in the diet of those animals (say, in Africa), there would have been no real disadvantage of a “broken allele” for the “vitamin C gene” and the DNA could persist in the population as a pseudogene.

Procedure:

1. Examine the four gene sequences below and mark any differences among the sequences that you can find.

```
—  
#1 GGAGCTGAAGGCCATGCTGGAGGCCACCCCGAGGTGGTGTCCCCTACCTGGTGGGGCTACGCTTCACCTGGAGG  
#2 GGAGCTGAAGGCCATGCTGGAGGCCACCCTGAGGTGGTGTCCCCTACCCGGTGGGGGTGCGCTTCACCCAGAGG  
#3 GGAGCTGAAGGCCGTGCTGGAGGCCACCCTGAGGTGGTGTCCCCTACCTGGTGGGGGTACGCTTCACCTGGAGG  
#4 GGAGATGAAGGCCATGCTGGAGGCCACCCTGAGGTGGTGTCCCCTAACCGGTGGGGGTGCGCTTCACCCAAGGG  
—
```

2. Discuss the following questions with your lab partner: Do you notice any specific pattern? What could this pattern mean regarding the ancestry/relatedness of the four species?
3. Together with your lab partner, make an hypothesis about the ancestry of these four species in the form of a phylogenetic tree. Draw this tree on a separate sheet of paper and make a few notes explaining why you drew it this way.

II. Exercise #2 – Molecular phylogenetics using a coding sequence (protein)

- While noncoding DNA sequences are extremely useful in analyzing the shared ancestry of different species, protein-coding DNA sequences are also useful.
- However, the mutation and evolution of protein-coding sequences of DNA is more “constrained” because most random changes to a protein sequence would be harmful to the function of the protein and thus, result in an allele that could reduce the health or reproductive success of the organism.
- Occasionally, however, random mutations in coding DNA could occur that result in a protein with no detectable loss of function. These changes are said to be “conservative,” because the protein function is conserved, despite an amino acid change. This type of allele would have little effect on the health and survival of those animals that carry it, and thus qualifies as “neutral variation.”
- (Even more rare are mutations that could actually *enhance* the function of the gene and the fitness of the individual – although rare, these events are the essence of adaptation!)
- 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 populations only because it happened to be present in the founders, not because they give any adaptive advantage to the population.
- Below is the amino acid sequence for a protein called SCML1, an enzyme necessary for male embryonic development and male fertility in mammals. It is encoded by a gene on the X-chromosome.
- The amino acid sequence is only slightly different amongst five mammals, as shown below: (“.../...” represents a stretch of identical amino acids, and is this omitted.)

```
—  
#1 MSNS.../...VIKT.../...DDNTI.../...EQLKTVDD.../...DALQN.../...RFHARSLWTNHNKRYG.../...KKHSYRLVL.../...YETF...  
#2 MSDS.../...VVKT.../...DDNTI.../...EQLRTVND.../...DALQN.../...RFYARSLWTNRKRSG.../...KKHSYRPVL.../...YETF...  
#3 MSNS.../...VVKT.../...DDDTI.../...EQLKTVND.../...DAMQN.../...RFHARFLWANRKRYG.../...KKHSYRLVL.../...YETF...  
#4 MSNS.../...VVKT.../...DDDTI.../...EQLKTVND.../...DAMQN.../...RFHARSLWTNRKRYG.../...KKYSYRLVA.../...YESF...  
#5 MSSS.../...VVKT.../...DDDTI.../...EQQKTVND.../...DAMQN.../...RFRARSLWTNRKRYG.../...KKYSYRLVA.../...YESF...  
—
```

Procedure:

1. Examine the five amino acid sequences above and mark any differences among the sequences that you can find.
2. As in exercise #1, use the differences in amino acid sequence to retrace the ancestry of these five mammals. Make an hypothesis in the form of a phylogenetic tree. Draw this tree on a separate piece of paper, along with your notes explaining it.

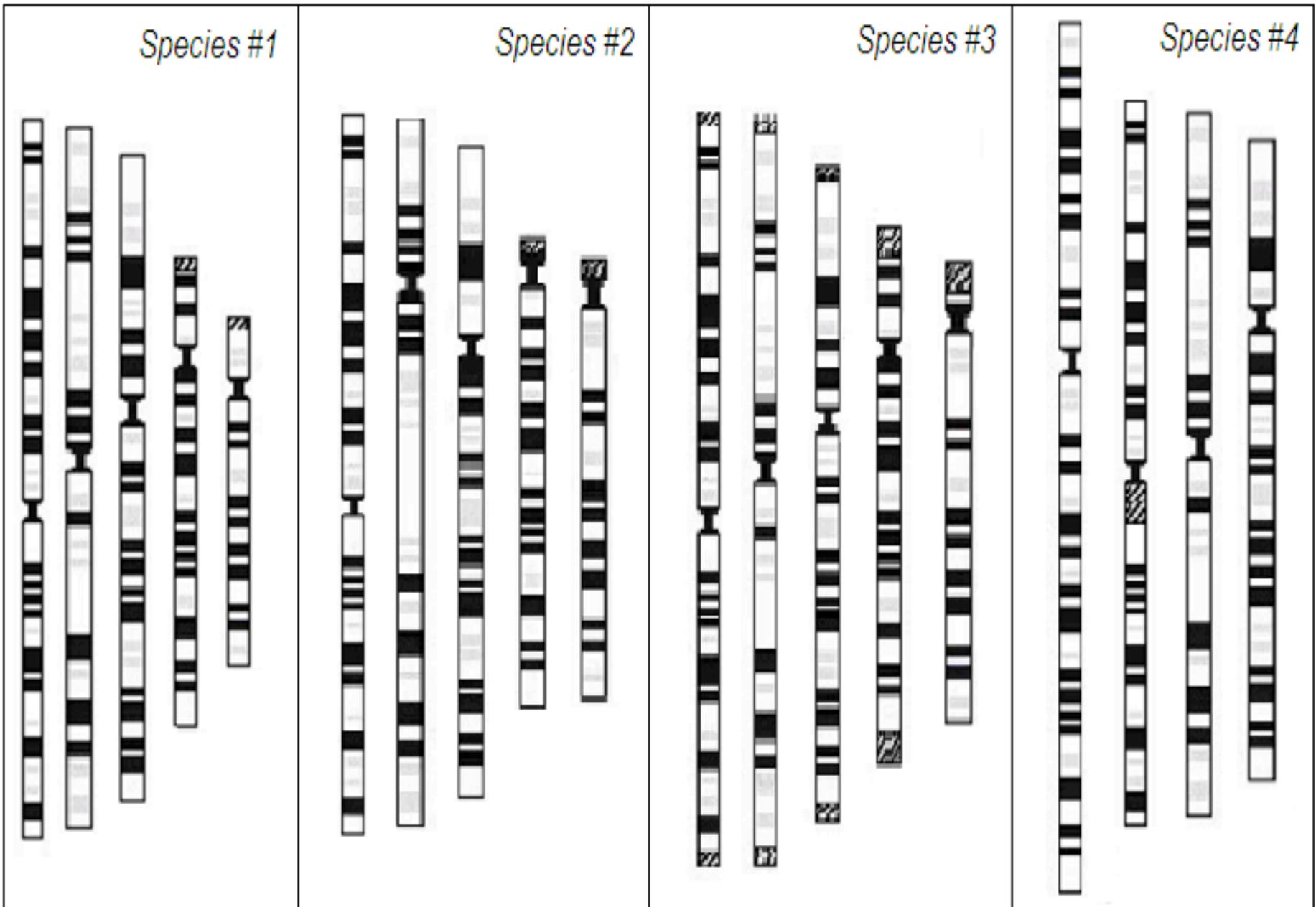
III. Exercise #3 – Evolution of Gross Chromosomal Structure

- Not all genetic changes are as subtle as single base-pair changes in DNA. Sometimes, large chromosomal rearrangements occur.
- These large rearrangements occur as errors in meiosis, often in the form of chromosomal breakage, followed by imperfect repair.
- Imperfect repair introduces a major structural change to the chromosome. Some examples include:
 - **Translocation** – when part of one chromosomes breaks loose, and is inadvertently glued back on to the end of a different chromosome.
 - **Inversion** – when part of a chromosome breaks loose and is glued back in place, but in the backwards orientation.
 - **Duplication** – When errors in DNA replication cause a certain region of a chromosome to be inadvertently repeated.
 - **Deletion** – When errors in DNA replication (or meiosis) cause the permanent loss of part of a chromosome.
- It may seem difficult to imagine how such a drastic change in chromosomal structure, resulting in large alterations to genes and genetic material, could possibly be tolerated. However, such changes would not always be lethal for the cell or individual in which this error occurs.
- Consider a simple hypothetical example:
 - Suppose a small region of chromosome #14 in a certain animal breaks off during the formation of gametes (meiosis).
 - This small region contains many crucial genes.
 - The exact point of chromosomal breakage is just upstream of a certain gene involved with cellular metabolism and energy consumption. Let's call this gene METAB1. The coding region of METAB1 is intact, but it is now separated from most of its "regulatory sequences" (promoters, enhancers, etc.).
 - The cell detects this error and acts to fix the error by gluing the small chromosome piece back in place. However, it commits an error and glues the piece back onto a different chromosome.
 - The chunk of DNA is rescued and intact, but now in a new location. With the exception of METAB1, all genes on the small "breakaway" piece of chromosome are still surrounded by their normal regulatory sequences and everything would work normally with these genes.
 - The coding region of METAB1, however, is now found in a new genomic "neighborhood" with different DNA sequence upstream, meaning a different promoter and regulatory sequences.
 - Perhaps the new upstream sequences of METAB1 cause this gene to be expressed at twice the level as before.

- In all cells of this organism, there would now be twice the amount of the product of the METAB1 gene.
- With twice the expression of METAB1, perhaps organisms with this chromosomal rearrangement are able to run faster for a longer length of time, generate more body heat during cold winters, or other beneficial effects of a faster metabolism. (This hypothetical examples assumes that the animals have plentiful access to food, which would be required by a faster metabolism.)
- In this case, natural selection would favor those individuals with the new chromosomal arrangement and over time, the species would evolve as this new chromosomal arrangement took over in the population.
- On the attached sheet are drawings of the gross chromosomal structure of some chromosomes of four related animals. The patterns of bands on the chromosome represent alternating degrees of electron density. Inactive regions of a chromosome (heterochromatin) are packed more densely than active areas (euchromatin), giving the banding pattern shown.

Procedure:

1. Examine the chromosomes on the attached handout. Keep in mind that you are looking only at a few of the larger chromosomes for the four species, and although all of these species are diploid, we are only looking at one chromosome of each pair.
2. Using scissors, cut out all of the chromosomes, but be sure to **label each one as to which species it comes from.**
3. Now, match up the chromosomes that very obviously are “homologous.” Once you have the homologue from all four of the species, place those homologous chromosomes in separate piles.
4. Some chromosomes may be leftover, meaning you cannot identify a clear homologue for all four species. Set these aside in a separate pile, and your instructor will discuss those later.
5. For those chromosomes that have an obvious homologue in all four species, as in the previous exercises, use the differences you see in the chromosomal structure to retrace the ancestry of these four animals.
6. The best strategy is to first consider each chromosome separately. Looking only at the first chromosome, which two or three species look the most similar, etc.? Then after you have completed each chromosome individually, see how each of your analyses compare, and then draw a larger conclusion, if you can.
7. As before, make your hypothesis about the relatedness of these animals in the form of a phylogenetic tree, along with your notes explaining how you made your decisions.
8. Your instructor will discuss the other chromosomes, those you could not identify a clear homologue for all species, at the end of the period. But if you have time, look closely at these and see if you can identify similarities between species.



Discussion/Homework questions:

- 1) Why are pseudogenes and other noncoding DNA sequences so commonly used by evolutionary biologists to determine relatedness between species?
- 2) While the GULO gene shows significant differences between these four species, the β -globin gene (used to make hemoglobin) shows very rare differences that are “silent” (the hemoglobin protein is identical) among these four species. Why do you suppose this is?
- 3) Discuss what is meant by “molecular clock” in terms of DNA sequences and how this “clock” is different amongst coding and noncoding DNA sequences.
- 4) Give one example of how a “neutral” genetic variant (allele) could become fixed in a population.
- 5) Give a purely hypothetical, but detailed example of how an error in meiosis could lead to a beneficial genetic change in a species.
- 6) Which kinds of scientific methodologies (as you learned about them in the recitation session) were present in the activities completed in this lab? Refer to the hand out on scientific practice if necessary.