

BioE/MCB/PMB C146/246, Spring 2005

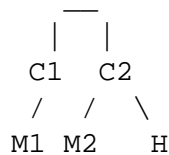
Problem Set 1

1. Answers varied but usually included wings of some sort, as homologs or as analogs. A common choice for homologous features was bat wings and human arms. Bird wings and insect wings were popular examples of analogous features. Bird wings and bat wings were used often as an example of analogy, but it is important to remember that they are *functionally* analogs and *structurally* homologs: both arose from a common ancestor limb, but that limb was not used for flight.

2. Since M1 and M2 are most closely related by a duplication event, these genes are paralogous, and are not orthologous. H and M1 are most closely related by a speciation event, and are therefore orthologous and not paralogous; the same holds for H and M2. Since orthology and paralogy are subsets of homology, and all three genes descended from a common ancestor, all three pairs of genes are homologous. None of these pairs are analogs or xenologs, because they come from a common ancestor without transfer.



Although the above tree gives the most parsimonious explanation for the relationship among the three genes (as it involves only a single duplication event), another possible scenario might be that the cenancestor had previously undergone a duplication in this gene, and the second copy had been lost at some point along the lineage leading to human:



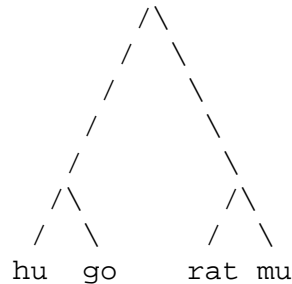
In this case, M1 and M2 will still be paralogs, since they are still most closely related by a duplication event, and M1 and H will still be orthologs, since they are most closely related by a speciation event. However, the relationship between H and M2 will now be paralogous, since they are now most closely related by the duplication event in the ancestral lineage rather than the human/mouse speciation event.

3. We could group the species into two closely related groups, human/gorilla and rat/mouse. Then we reconstruct the ancestral sequence of each group according to Fitch's method. If these ancestral sequences are more similar than the extant sequences, we would be inclined to think that all sequences share one common ancestor. If the ancestral sequences are less similar than the extant sequence, we would be inclined

to think that the similarity is due to convergent evolution rather than shared ancestry. Statistical tests are necessary to determine if the similarity is more or less than expected.

Note that Fitch's method relies on being able to group the sequences into subsets based on known homology. The method allows one to determine that the two subsets are also homologous.

Here the sequences in the human/gorilla group each differ by six residues from the rat/mouse group sequences. The parsimonious ancestral sequences only differ by as few as four residues. We are inclined to think that these sequences are indeed homologous.



ancestral hu/go sequence: A A TA(T/A)T(G/C)AA G CGCTACCTATA
 mutations x x x
 ancestral rat/mu sequence: A(T/A)TT A T C AA(A/G)CGGTGCCTACA

4. A. Homology is transitive. If A is homologous to B and C, that means that there is a common ancestor of A, B, and C. Thus, B and C have a common ancestor. By definition, they are homologous.

B. Orthology is not transitive. Consider a case similar to Question 2, where H is orthologous to M1 and M2. However, since M1 and M2 are paralogs within the same species (a gene duplication occurred after the speciation event that split H and M), they are not orthologous.

5. Answers vary slightly depending on which part of the PDB entry you used to define the helices.

```

atgtacaagaaa gatggtatcgac cacttcggaacc cagcgtgcagta gctaaggcttta
      HHHHHH HHHHHHHHHHHH HHHHHH HHHHHHHHHHHH HHHHHHHHHHHH

ggcattagcgat gcagcgtctct cagtggaaggaa gttatcccagag aaagacgcatac
      HHH HHHHHHHHHHHH HHH HHH HHHHHHHHHHHH

cgattagagatc gttacagctggc gccctgaagtac caagaaaacgct tategccaagcg
HHHHHHHHHHHH HHH

gcgtaa
  
```

6. Answers vary. One key thing to note is that you must check the constraints for B *after* the generation is done, and throw out *all* changes from that generation if any one of them does not pass the constraints. After all, in natural selection, the organism with the deleterious mutation doesn't survive to pass on any of its new mutations. Another important feature was that the constraints in B were meant to preserve the amino acid (residue) sequence, not the nucleotide (base) sequence. A nucleotide change that doesn't change the eventual protein sequence will not be selected against, at least in this overly-simplistic model.

The graphs for A and B1 should look very similar. Differences are due only to the random process of choosing which bases mutate. The graph for B2 should show fewer mutations overall, with many positions

unchanged. In an ideal example you would see that the selective pressure was higher in the helical regions, but the parameters for this problem actually do not make that particularly obvious from the graph.

The model of evolution is not particularly realistic. Improvements include accepting some amino acid changes more readily than others, such as changes between residues with similar size or hydrophobicity; incorporating other structural or functional constraints on the sequence than just helix regions; and making sure that the original start and stop codons do not change. Of course, it is hard to predict what changes might actually be beneficial to the organism. Models such as this are usually built more on negative selection than positive selection.

This simulation was intended to make you distinguish between the rate at which mutations occur in DNA (the mutation, transition, and transversion probabilities) and the rate at which mutations are accepted and passed along to future generations. The former is an almost purely random process. The latter is where the function of the protein becomes important.