Bioinformatics – Lecture Notes

Announcements

This Thursday (3/28) I will be out of town. Isabel Darcy will teach the class. She will talk about DNA structure and how it relates to its function. She will demonstrate how the mathematical field of topology can aid in studying these topics.

Class 19 – March 26, 2002 -

A. Phylogenetic Trees – Maximum Likelihood Approaches continues

1. To find the optimal branch lengths for a fixed topology is a hard combinatorial problem

$$L(t(d_1,\ldots,d_k)) = Pr[a_1,\ldots,a_n \mid t(d_1,\ldots,d_k)]$$

An easier method employs Felsensetin's *Pulley Principle*.

Let $t(d_1,\ldots,d_k)$ be the tree for a given taxa $(a_1,\ldots,a_n)$ with edge distances $(d_1,\ldots,d_n)$. Assuming the reversible Markov chain condition stated at the beginning of this section, the likelihood of a tree $t(d_1,\ldots,d_k)$ is unaffected by the placement of the root.

What this means is that we can start the likelihood calculation at an arbitrary edge of a tree since the root can be placed arbitrarily.

Proof of Pulley Principle

Example – Figure 4.11 can be written at Figure 4.12

We will use a Lemma of the Pulley Principle that will allow us to calculate the Likelihood of an arbitrary edge t. Then we will maximize this likelihood by finding the local extrema for the likelihood function.

2. To determine the topology we choose the tree (model) that has the maximum likelihood out of all possible trees.

Procedure to for a unrooted tree with n taxa.

i) Start with the first two taxa. Then perform the following iteration

ii)   After k species have been added, the tree has k external
        nodes (and hence k-2 internal nodes) so that the number
        of edges is one less that the total number of nodes.
              k + (k-2) – 1 = 2k-3
iii)   Apply the previous described maximum likelihood
        method to determine the optimal placement of the next
        node and continue

Note: There is no guarantee that this is the best topology among
the large (exponentially many) number of possibilities.
Furthermore, the final topology might depend on the order the
taxa are added.

Question: Can you think of a way that might be useful to test
how good the topology you find is?

Next time: Quartet Puzzling
               Hidden Markov Models