Homework 10

BCB 548

Nov 28, 2006

1  Fitch’s algorithm, 20 points total

Prove the correctness of Fitch’s algorithm.

2  Bifurcating trees, 10 points

Show that for every multifurcating tree that is most parsimonious, there exists an equivalent bifurcating tree that is equally parsimonious.

3  Phylogenetic trees, 20 points

A rooted binary parsimony tree for a set of characters is called perfect if for every character each state evolves at most once. Let \( C \) be a set of binary characters with states 0 and 1 over a same species-set, where the state 0 can only change into the state 1. Find an EXACT criterion for the characters \( C \) that must be satisfied such that a perfect parsimony tree for \( C \) exists. Show the correctness of your criterion.

4  Maximum parsimony, 20 points

Is it the case that, for any input tree \( T \), there is an optimal labeled tree in which every node along some path from the root to a leaf has the same character? Why or why not?

5  Informative sites, 10 points

Informative sites in an alignment are defined as columns that favor one tree topology over another. Find the informative sites in the following alignment:

ATGTA
TAGTA
CGCTG
GCCTG
Which tree will you find using the maximum parsimony algorithm for this alignment and set of sequences?

6 Sequence assembly, 20 points total

Your task for this problem is to apply the assembly method discussed in class to the small data set that is attached. The data set includes a list of read sequences and a list of constraints. There is no sequence of quality values. You need to compute overlaps, contigs and scaffolds, as specified below. Find all perfect overlaps of at least 15 bases between reads. You may use an alignment program to find a potential overlap for each pair of reads in the same or opposite orientation. Each overlap is specified by showing the names and orientation of the two reads in the overlap. Assemble reads into contigs by using the overlaps. Each contig has to contain at least two reads. Each contig is specified by showing a list of overlapping reads in proper orientation in the contig. Link contigs into scaffolds by using constraints, where each link between two contigs has to be supported by at least two constraints. Each scaffold is specified by showing a list of contigs that are ordered and oriented. You can work on this problem manually or with programs written by you.