

Algorithms in Bioinformatics I, WS2002/3

Assignment sheet # 10

Daniel Huson

December 16, 2002

To solve the problems below, please download the following file:

www-ab.informatik.uni-tuebingen.de/teaching/ws02/abi1/programs/program12.zip. This contains all necessary files and a number of input sequences and trees. Please run your programs on all supplied data sets.

1 The parsimony score of a phylogenetic tree (5 points)

Write a program `ParsScoreTree.java` that takes as input a phylogenetic tree T and a corresponding multiple alignment A and returns as output the parsimony score $PS(T, A)$. To do this, implement the Fitch algorithm in a method `getParsimonyScore(T)` in the java class `Characters.java`.

2 Stepwise addition heuristic (3 points)

Write a program `ParsStepAdd.java` that takes as input a multiple alignment of sequences $A = \{a_1, \dots, a_n\}$ and produces as output the score and the tree T obtained by the stepwise addition heuristic applied in the order in which the taxa are listed in the characters matrix.

3 Branch and bound search (5 points)

Write a program `ParsBandB.java` that takes as input a multiple alignment of sequences $A = \{a_1, \dots, a_n\}$ (with n small) and produces as output the optimal parsimony score and the tree T obtained by the branch-and-bound method.

4 Application of methods (2 points)

Use the `JCSimulator.java` sequence simulator to produce 20 different data sets from the model tree in file `primates.new`, across an interesting range of u and L values, e.g. vary u between 0.01 and 50 and L between 100 and 5000.

On each data set, run neighbor-joining, the step-wise addition heuristic and the branch and bound program. Use the `ParsScoreTree.java` program to evaluate and compare the results.

Due by 10am, Wednesday, January 8, 2003.