CISC 889 Bioinformatics (Spring 2004) Phylogenetic Trees (II) Character-based methods

CISC889, S04, Lec13, Liao

1









• Both algorithms run in O(nm), where n is number of sequences and m is the sequence length in terms of number of residues.
 Weighted parsimony, when using S(a,a) = 0 for all a and S(a,b) = 1 for all a ≠ b, gives the same cost as that for the traditional parsimony.
• Traceback in weighted parsimony can find assignments that are missed in the traditional parsimony.
• The cost from the traditional parsimony is independent of the position for the root node. Therefore, the cost can be computed using unrooted trees.
• Still the number trees to search using parsimony grows huge as the number of leaves increases. It is proved that finding the parsimonious tree is an NP-hard problem.
 Branch-and-bound Guarantee to find the optimal tree Worse-case complexity is the same as exhaustive search.
CISC889, S04, Lec13, Liao 6



- "Plug-in" sampling with replacement
 - Given an alignment of, say, one hundred bps.
 - Randomly select one column from the original alignment as the first column, and repeat this process until one hundred columns are selected forming a new alignment.
 - Use this artificially created alignment for parsimony analysis, a new tree is found.
 - Repeat this whole process many times (say 1000).
 - The frequency with which a chosen phylogenetic feature appears is used as a measure of the confidence we have in this feature.

CISC889, S04, Lec13, Liao

7





CISC889, S04, Lec13, Liao