

CMP 464/788 Lecture Notes

Phylogeny: Consensus Methods & Splits
Introduction to Computational Biology
23 October 2003

Computing Phylogenies

Input
Data

A GTTAGAAGGC...
B CATTTGTCCT...
C CAAGAGGCCA...
D CCGACTTCCA...
E ATGGGGCACG...
F TACAAATACG...

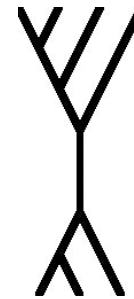


Reconstruction
Algorithms

Maximum Parsimony
Maximum Likelihood
...



Output
Tree



Computing Phylogenies

Input Data

A GTTAGAAGGC...
B CATTTGTCCT...
C CAAGAGGCCA...
D CCGACTTCCA...
E ATGGGGCACG...
F TACAAATACG...

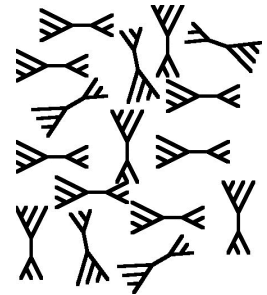


Reconstruction Algorithms

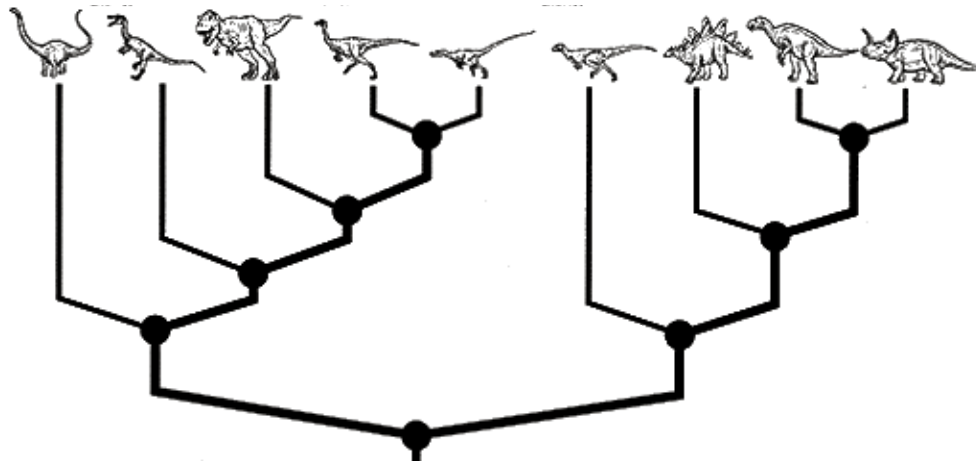
Maximum Parsimony
Maximum Likelihood
...



Output Trees

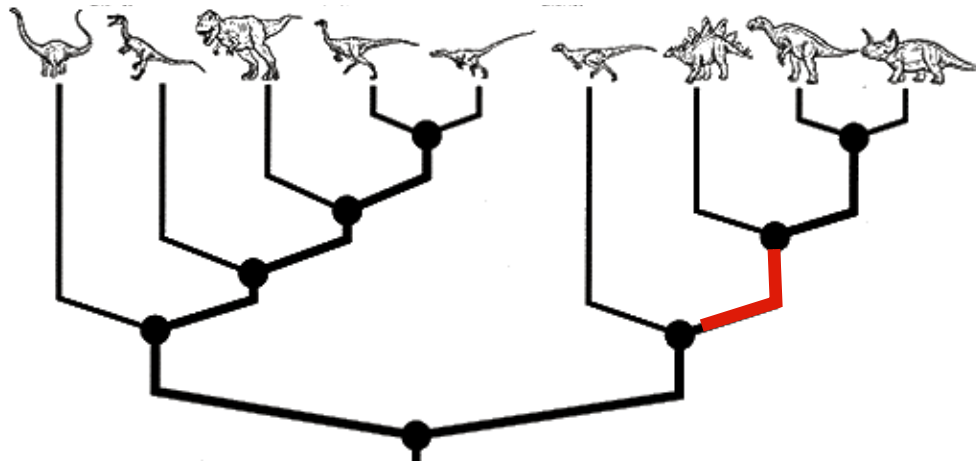


Phylogenies



(www.amnh.org/education/teacherguides/dinosaurs)

Phylogenies

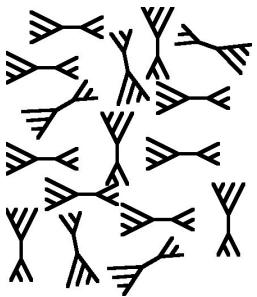


(www.amnh.org/education/teacherguides/dinosaurs)

Logically tree is defined by its set of **splits**.

Summarizing Trees

Input
Trees

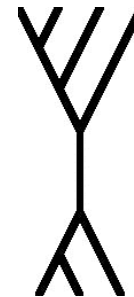


Consensus
Method

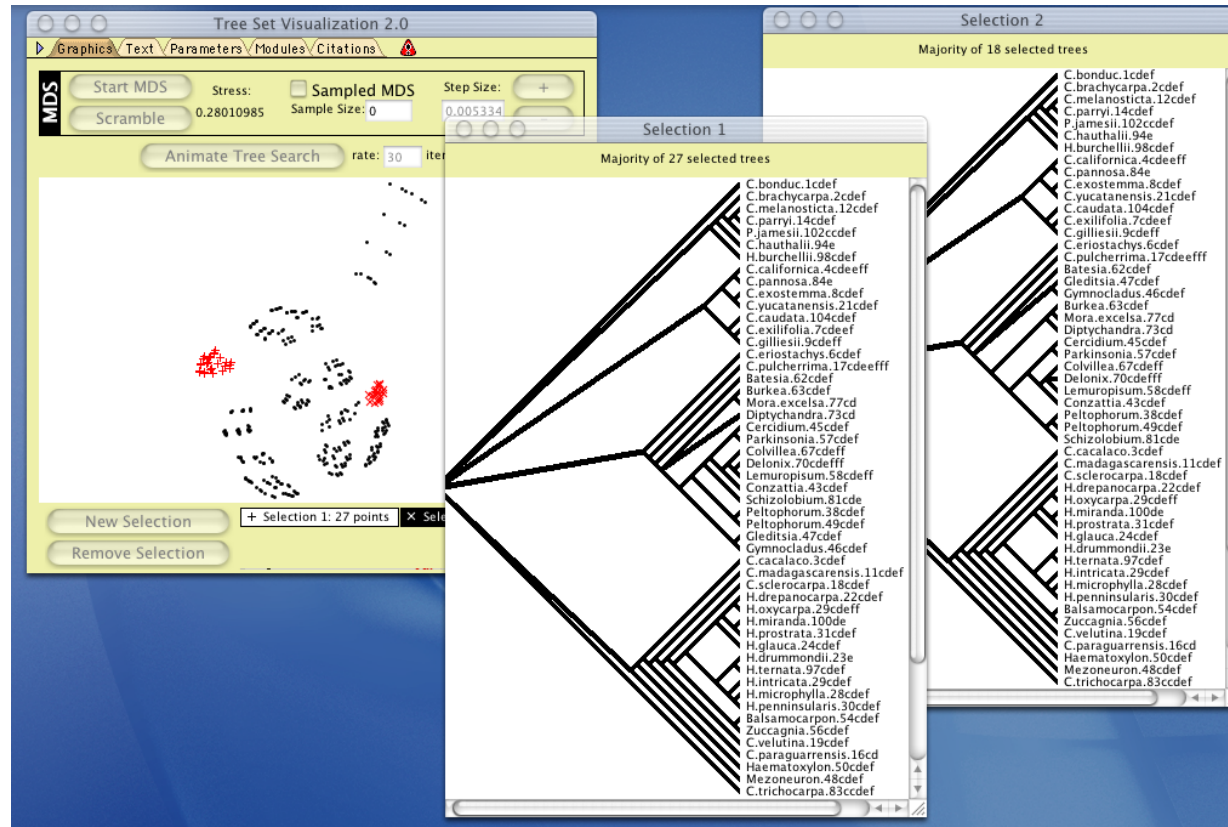
Strict Consensus
Majority-rule



Output
Trees



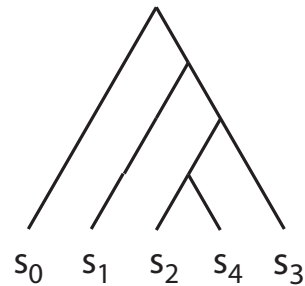
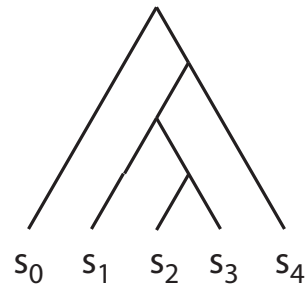
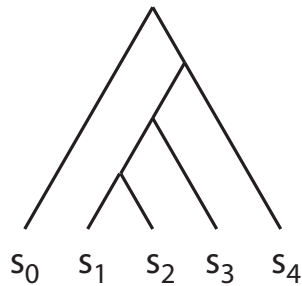
Visualizing Sets of Trees



Efficiency is important for real-time visualization.

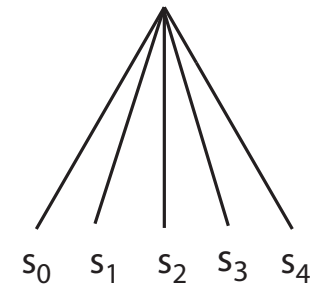
Strict Consensus Tree

Input trees



→

Strict Consensus



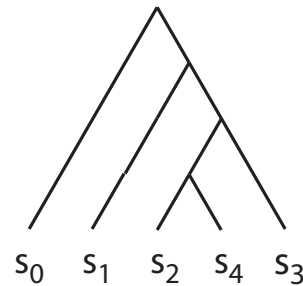
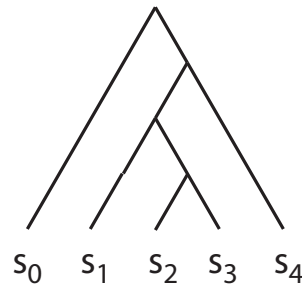
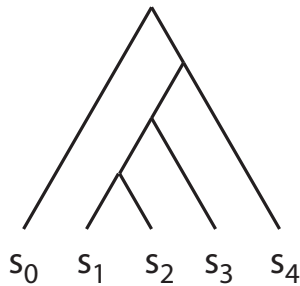
$s_1 s_2 \mid s_0 s_3 s_4$
 $s_1 s_2 s_3 \mid s_0 s_4$

$s_2 s_3 \mid s_0 s_1 s_4$
 $s_1 s_2 s_3 \mid s_0 s_4$

$s_2 s_4 \mid s_0 s_1 s_3$
 $s_2 s_3 s_4 \mid s_0 s_1$

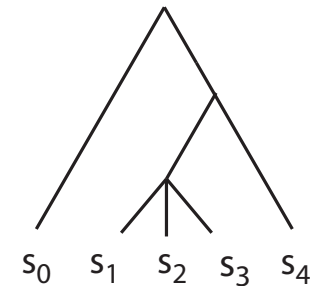
Majority-rule Tree

Input trees



→

Majority-rule Tree



Includes splits found in a majority of trees
Can be 2/3 majority, etc.

Past Work

- Strict Consensus:
 - $O(nt)$: Day, 1985 ($n = \#$ of leaves, $t = \#$ of trees, and assuming the size of the machine word $w = O(\lg n)$)
- Majority-rule Consensus:
 - $O((n/w)(nt + \lg x + n^2))$: Margush & McMorris, 1981. ($x =$ total number of splits, $\leq nt$)
 - $O(n^2 + nt^2)$: Wareham, 1985.
 - Randomized algorithm with expected running time $O(nt)$: Amenta, Clarke, & St. John, 2003.