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

Input Data

A  GTTAGAAGGC...
B  CATTGTCTCT...
C  CAAGAGGCCA...
D  CCGACTTCCA...
E  ATGGGGCAGC...
F  TACAAATACG...

Reconstruction Algorithms

→

Maximum Parsimony
Maximum Likelihood
...

Output Tree
Computing Phylogenies

Input Data

A  GTTAGAAGGC...
B  CATTTCATCCT...
C  CAAGAGGCCA...
D  CCGACTTCCA...
E  ATGGGGCAGC...
F  TACAAATACG...

Reconstruction Algorithms

Maximum Parsimony
Maximum Likelihood
...

Output Trees
Phylogenies

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

Logically tree is defined by its set of splits.

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

Input Trees $\rightarrow$ Consensus Method $\rightarrow$ Output Trees

- Strict Consensus
- Majority-rule
Visualizing Sets of Trees

Efficiency is important for real-time visualization.
Strict Consensus Tree

Input trees

Strict Consensus

\[ s_0 s_1 s_2 s_3 s_4 \]

\[ s_0 s_1 s_2 s_3 s_4 \]

\[ s_0 s_1 s_2 s_3 s_4 \]
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 ($n = \# \text{ of leaves}, t = \# \text{ of trees, and assuming the size of the machine word } w = O(lg n)$)

• Majority-rule Consensus:
  – $O((n/w)(nt + 1g x + n^2))$: Margush & McMorris, 1981. ($x = \text{total number of splits, } \leq nt$)
  – $O(n^2 + nt^2)$: Wareham, 1985.