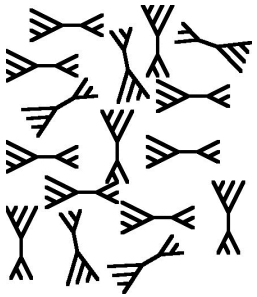# CMP 464/788 Lecture Notes

Phylogeny: More on Consensus Methods & Splits
Comparing Trees
Introduction to Computational Biology
28 October 2003

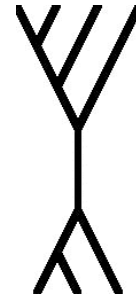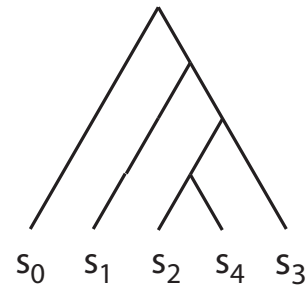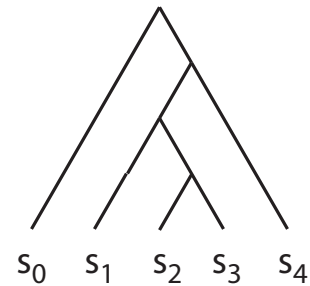# Summarizing Trees

Input
Trees

Consensus
Method

Output
Trees

$\longrightarrow$

Strict Consensus
Majority-rule

$\longrightarrow$

Strict Consensus Tree

Input trees
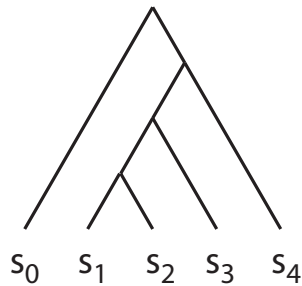
Strict Consensus

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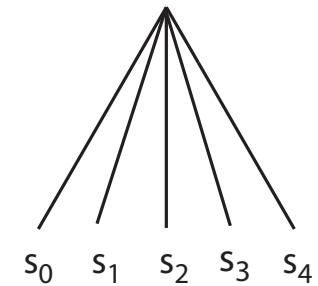
$s_1 s_2 \mid s_0 s_3 s_4$ $\quad$ $s_2 s_3 \mid s_0 s_1 s_4$ $\quad$ $s_2 s_4 \mid s_0 s_1 s_3$

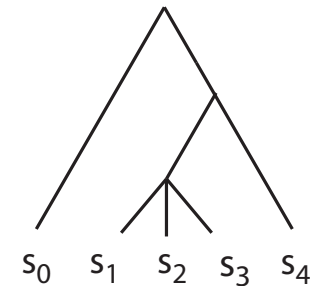$s_1 s_2 s_3 \mid s_0 s_4$ $\quad$ $s_1 s_2 s_3 \mid s_0 s_4$ $\quad$ $s_2 s_3 s_4 \mid s_0 s_1$

Majority-rule Tree

Input trees
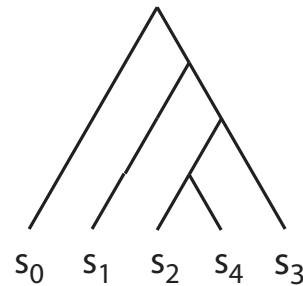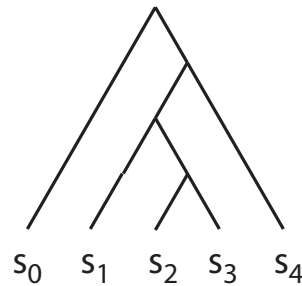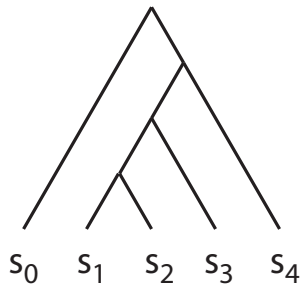
Majority-rule Tree

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

Compatible Splits

Let S be the set of all possible leaf labels, and
$\bar{A} = S - A$.
Two splits $A_1 \mid \bar{A}_1$ and $A_2 \mid \bar{A}_2$ are compatible if one of the follow sets is empty:

$$A_1 \cap A_2, \; A_1 \cap \bar{A}_2, \; \bar{A}_1 \cap A_2, \; \bar{A}_1 \cap \bar{A}_2$$

Example:
$$s_1 s_2 \mid s_0 s_3 s_4,$$
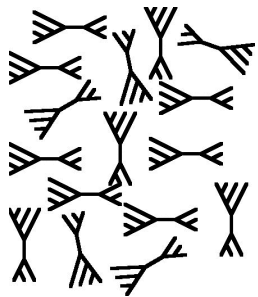$$s_2 s_3 \mid s_0 s_1 s_4$$

Trees & Compatible Splits

Facts:

- Every tree gives a compatible set of splits.

- Every compatible set of splits gives a tree.

Example:

$s_1 s_2 \mid s_3 s_4 s_5 s_6 s_7,$

$s_3 s_5 \mid s_1 s_2 s_4 s_6 s_7,$

$s_4 s_6 s_7 \mid s_1 s_2 s_3 s_5,$

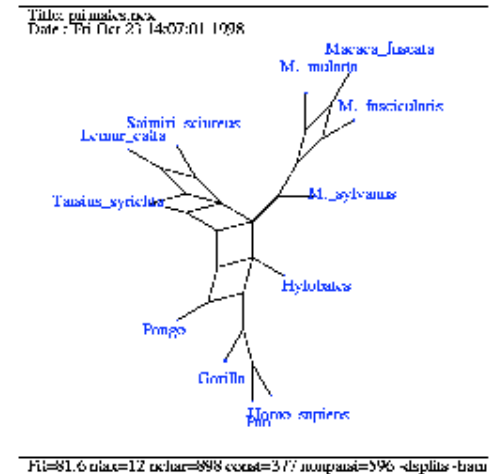$s_4 s_7 \mid s_1 s_2 s_3 s_5 s_7$

Summarizing Trees

Input Trees	Consensus Method	Output Network

$\longrightarrow$

Strict Consensus
Majority-rule

$\longrightarrow$

(Network from SplitsTree webpage, Prof. Daniel Huson, U. Tuebingen.)

Weak Compatibilty

- If the set of splits are compatible, then we get a tree.

- A relaxed version of compatibility:
  Three splits, $A_1 \mid \bar{A}_1$, $A_2 \mid \bar{A}_2$, and $A_2 \mid \bar{A}_2$, are weakly compatible if at least one of the intersection is empty:

$$A_1 \cap A_2 \cap A_3, A_1 \cap \bar{A}_2 \cap \bar{A}_3, \bar{A}_1 \cap A_2 \cap \bar{A}_3, \bar{A}_1 \cap \bar{A}_2 \cap A_3$$

  and at least one of the following intersections is empty:

$$\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3, \bar{A}_1 \cap A_2 \cap A_3, A_1 \cap \bar{A}_2 \cap A_3, A_1 \cap A_2 \cap \bar{A}_3$$
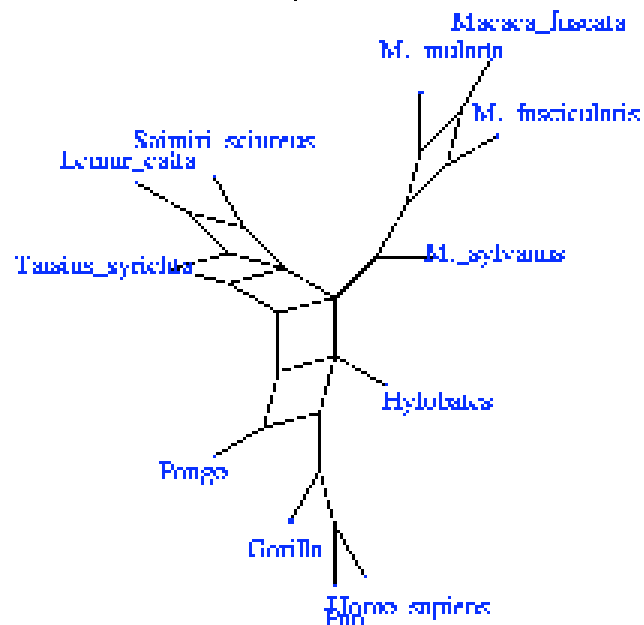
Splits Graph

A splits graph has the following properties:

- All leaves are multi-labelled by taxa so each taxon appears exactly once,

- edges are labeled by splits, and each split appears at least once,

- deleting all edges labeled by a split gives a two component graph corresponding to the split, and

- the graph is minimal with these properties.

Splits Graph
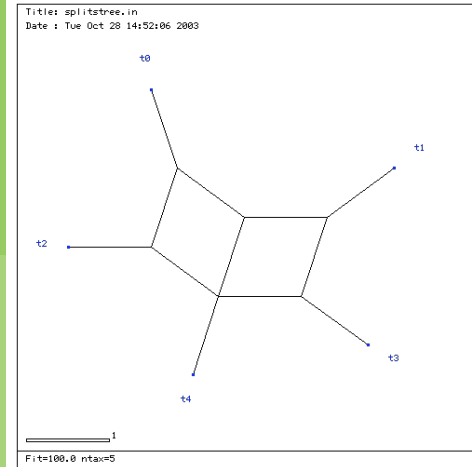
Distances Between Trees

$T_1 =$ [tree diagram with Turtles, Lizards, Snakes, Birds, Crocodiles]

$T_2 =$ [tree diagram with Turtles, Lizards, Snakes, Birds, Crocodiles]
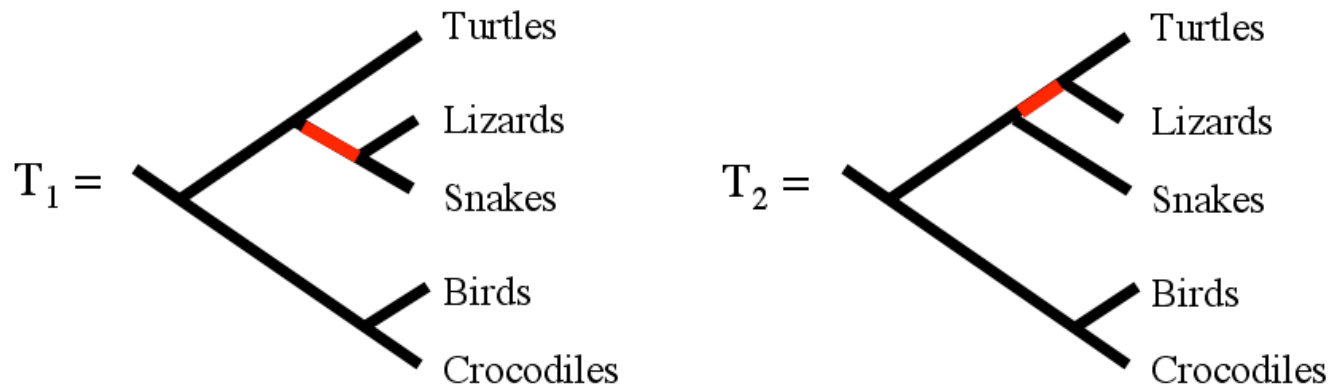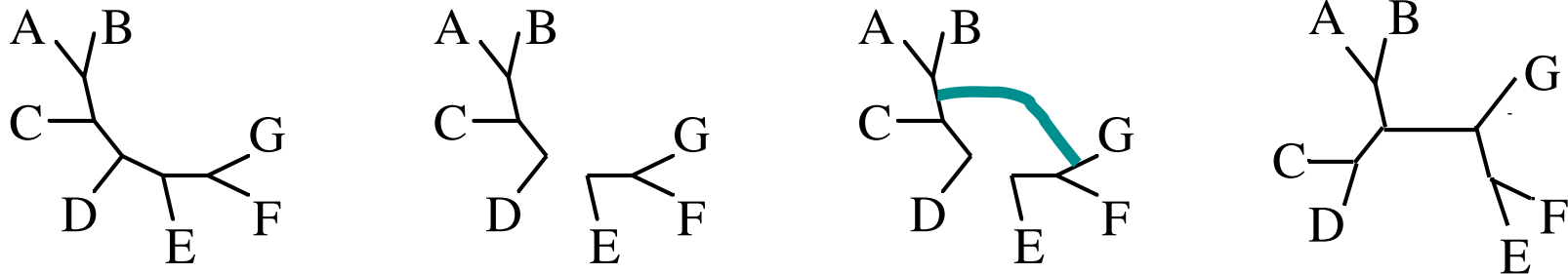
- Robinson-Foulds distance: # of edges that occur in only one tree.

- Calculate in $O(n)$ time using Day's Algorithm (1985).

- Extends naturally to weighted trees.
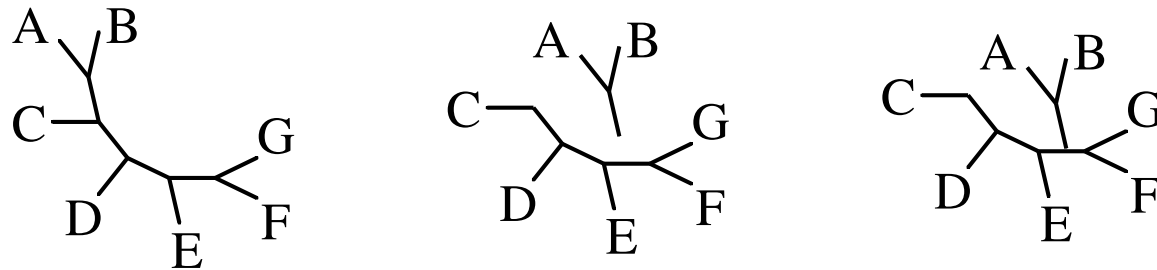
Other Natural Metrics

- Tree-bisection-reconnect (TBR):

- TBR is NP-hard. (Allen & Steel '01)

Other Natural Metrics

- Subtree-prune-regraft (SPR):

- Conjectured to be NP-hard.