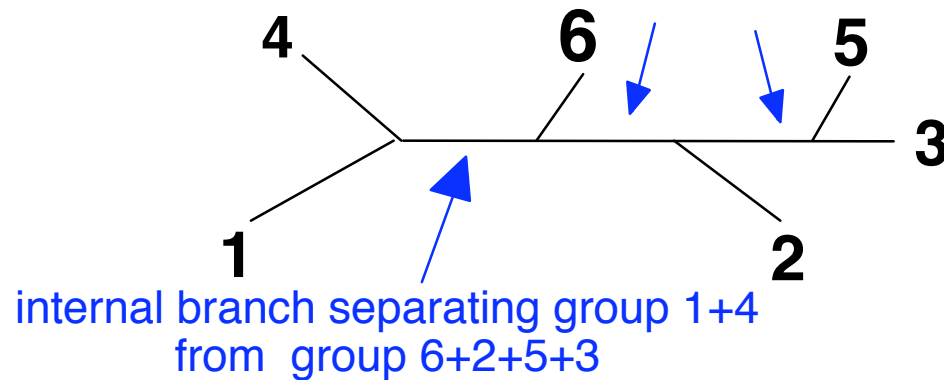Wellcome trust advanced course
"Molecular Evolution"

Statistical support of trees:
the (non-parametric) bootstrap

Manolo Gouy
CNRS - Université de Lyon

- The phylogenetic information contained in an unrooted tree is entirely located in its internal branches.

internal branch separating group 1+4 from group 6+2+5+3

- The tree shape is entirely determined by the list of its internal branches (or splits).

- Evaluation of the statistical support of a tree = evaluation of the support of each internal branch.
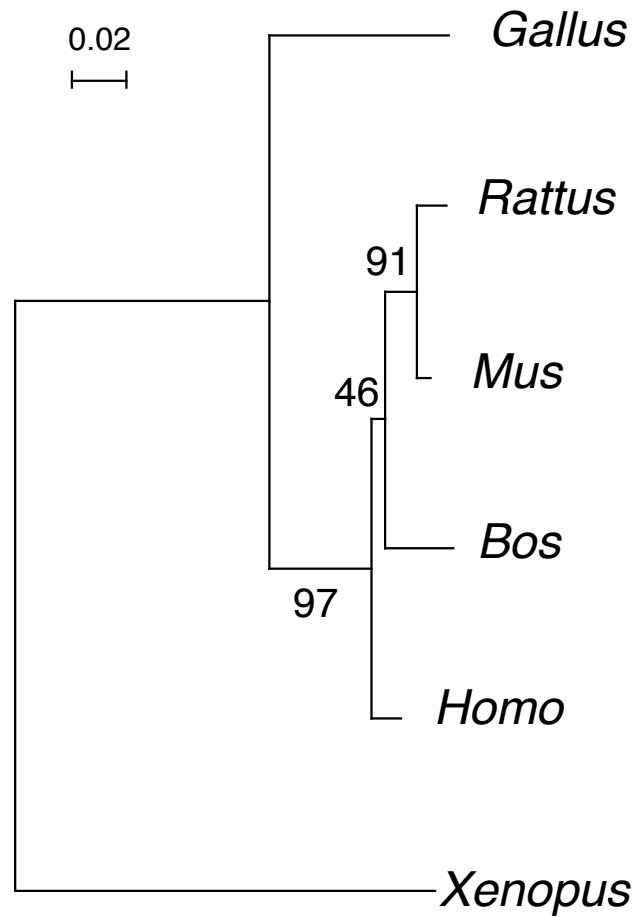
Bootstrap Procedure
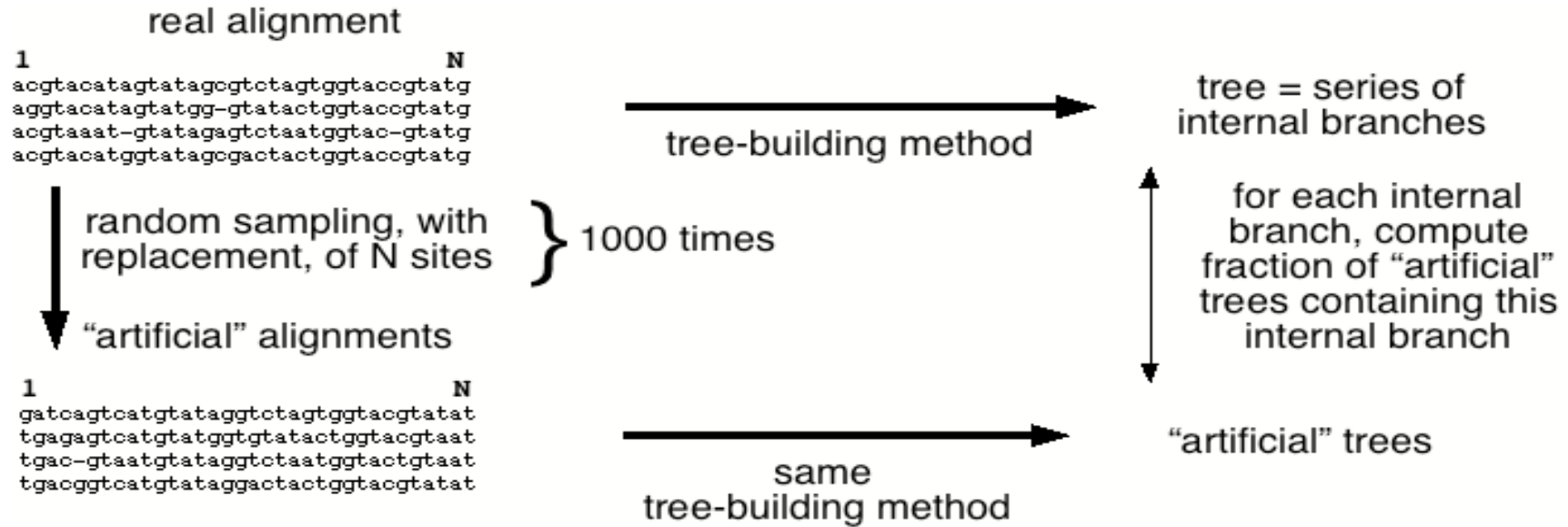
The support of each internal branch is expressed as percent of replications containing this branch.

Alternative procedure:

- Compute the consensus of all bootstrap (or "artificial") trees;
- Compute the support of branches of the consensus tree.

Felsenstein (1985) Evolution 39:783

tree with bootstrap support

Bootstrap Procedure : properties

- Internal branches supported by ≥ 90% of replicates are statistically significant.

- The bootstrap procedure detects whether sequences are long enough to support a given node.

- The bootstrap procedure does not help determine if the tree-building method is appropriate. A wrong tree can have 100% bootstrap support for all of its branches.

- Extreme example: a tree-building method that would always return the same tree, whatever the sequence data, would have 100% bootstrap support.

- If there is a systematic bias in the pair sequence data + tree-building method (*e.g.*, long branch attraction), the bias will also operate on bootstrapped sequences, so the bootstrap procedure will support the LBA artefact.