

The data

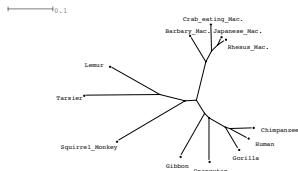
Phylogenetics is the branch of biology concerned with inferring evolutionary relationships between *species* or *taxa*.

For an analysis from molecular data, aligned DNA (or protein, etc.) sequences are used:

Primate mitochondrial DNA sequences, HindIII

from Hayasaka, K., T. Gojobori, and S. Horai. MBE (1988) 5:626-644.

Gorilla	AAGCTTCACGGCGCAGTTGTTCTTATAATTGCCACGGACTTACATCAT...
Orangutan	AAGCTTCACGGCGCAACCACCCTCATGATTGCCCATGGACTCACATCCT...
Human	AAGCTTCACGGCGCAGTCATTCTCATAATCGCCACGGGCTTACATCCT...
Chimpanzee	AAGCTTCACGGCGCAATTATCCTCATAATCGCCACGGACTTACATCCT...
Gibbon	AAGCTTTACAGGTGCAACCGTCTCATAATCGCCACGGACTAACCTCTT...
Crab-eat_Mac	AAGCTTCTCGGGCGCAACCACCCTTATAATCGCCACGGGCTCACCTCTT...
Lemur	AAGCTTCATAGGAGCAACCATTTCTAATAATCGCACATGGCCTTACATCAT...
Barbary_Mac	AAGCTTCTCGGTGCAACTATCCTTATAGTTGCCATGGACTCACCTCTT...
Japanese_Mac	AAGCTTTTCGGCGCAACCATCCTTATGATCGCTCAGGACTCACCTCTT...
Squirrel_Mon	AAGCTTCACGGCGCAATGATCCTAATAATCGCTCAGGGTTTACTTCGT...
Rhesus_Mac	AAGCTTTTCTGGCGCAACCATCCTCATGATTGCTCAGGACTCACCTCTT...
Tarsier	AAGTTTCATTGGAGCCACCCTCTTATAATTGCCCATGGCTCACCTCCT...



of length 898 sites. Tree reconstruction (NJ) leads to tree shown.

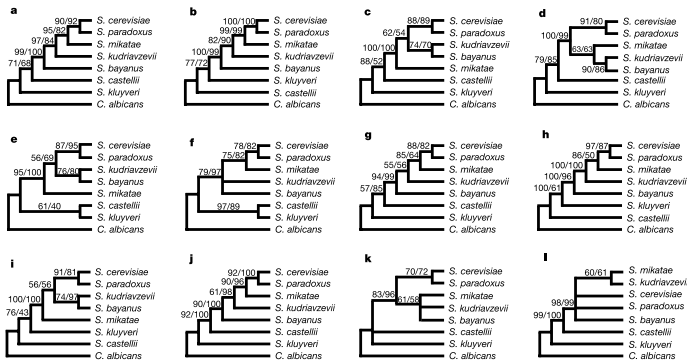


Figure 1 Single-gene data sets generate multiple, robustly supported alternative topologies. Representative alternative trees recovered from analyses of nucleotide data of 106 selected single genes and six commonly used genes are shown. The trees are the 50% majority-rule consensus trees from the genes YBL091C (a), YGL001C (b),

YER005W (c), YGL001C (d), YNL155W (e) and YOL097C (f), as well as those from the commonly used genes actin (g), hsp70 (h), β -tubulin (i), RNA polymerase II (j) elongation factor 1- α (k) and 18S rDNA (l). Numbers above branches indicate bootstrap values (ML on nucleotides/MP on nucleotides).

Gene trees vs. species trees

- ▶ **Gene tree:** tree that represents the evolutionary history for a particular gene
 - ▶ Estimated using genetic data (e.g. DNA sequence alignments)
- ▶ **Species tree:** tree that represents the sequence of speciation events that gave rise to the observed collection of species
 - ▶ This is usually the parameter of interest, but data are only indirectly informative about the species tree.

**Gene trees and species trees are NOT the same,
and often disagree.**

Sources of conflict

There are many reasons gene trees may differ from species trees

- ▶ lateral gene transfer

(e.g. viral insertion of genetic material into host genome)

- ▶ hybridization

(interbreeding between distinct species to produce hybrid population that shares genetic contributions from both parental organisms)

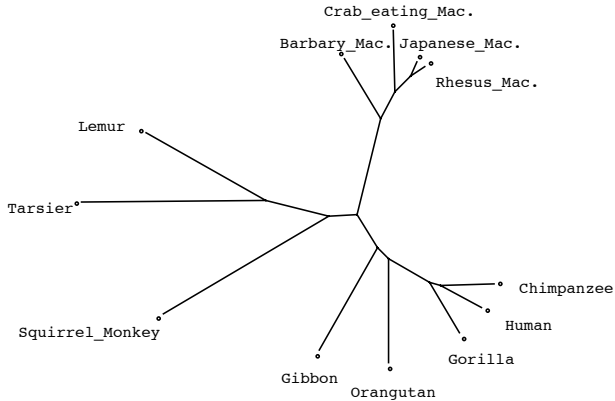
- ▶ effects from population genetics

(e.g. incomplete lineage sorting)

Gene tree discord gives information about the species tree.

... use sample of gene trees to infer species tree

0.1



Adjectives ???

Trees

- ▶ Leaf-labelled graphs
- ▶ binary = trivalent vs. multifurcating
- ▶ Rooted or Unrooted
- ▶ Metric vs. topological