
Botany 940 - Review 1

- Parsimony
- Likelihood Principle
- Ingredients of molecular evolution models:
 - Substitution rate matrix
 - Distribution of rates
- Bootstrap
- Divergence times

Next week: Bayesian methods; Testing monophyly

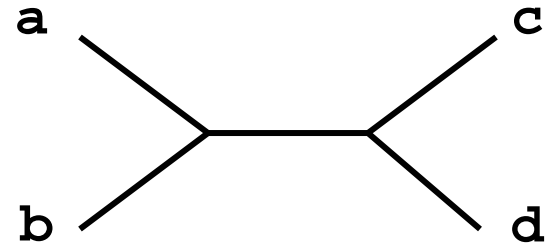
Parsimony

a: TAATAATTAAAA

b: TATTTTATATAA

c: TAAATAATTTTT

d: TAATAATTAAAT



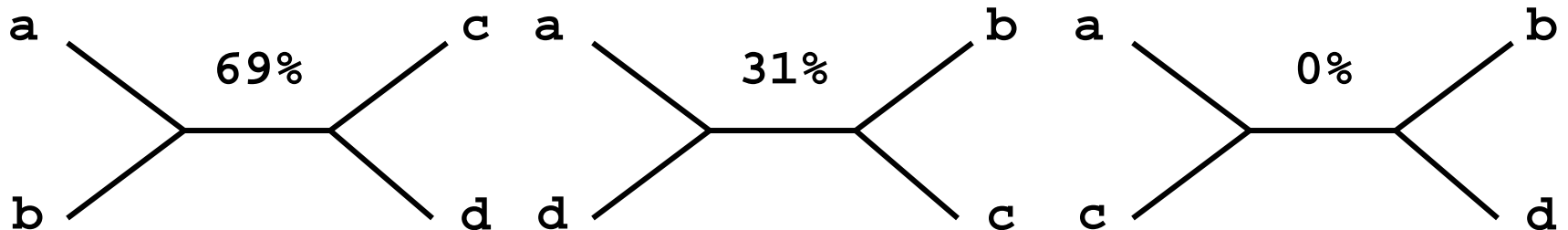
ab | cd L=11

ac | bd L=14

ad | bc L=12

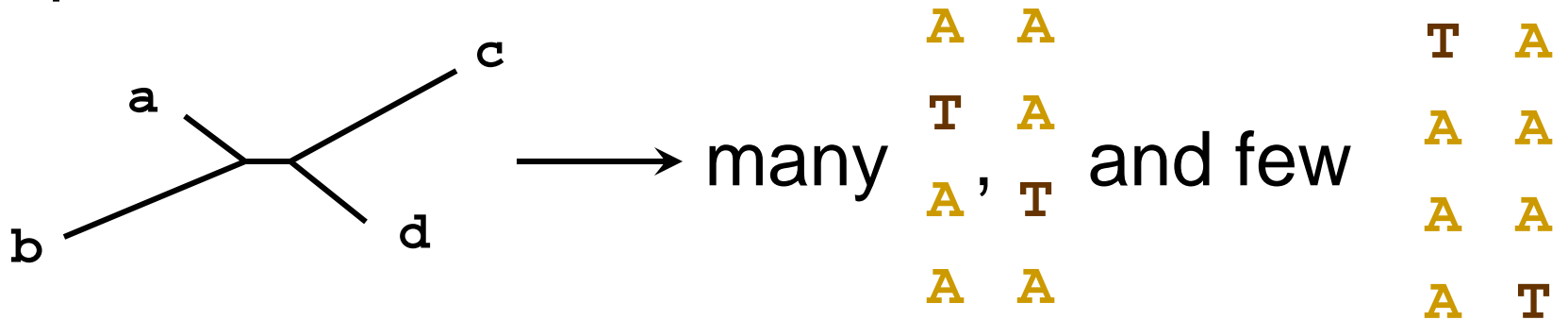
Bootstrap with Parsimony

- 12 volunteers:
 - choose a number between 1 and 12
- Get a new alignment and a new MP tree
- Repeat...



Likelihood principle

- Probability model: can be used to make predictions



- Data + probability model \longrightarrow likelihood
probability of the data under the model,
measures the fit of the model to the data.

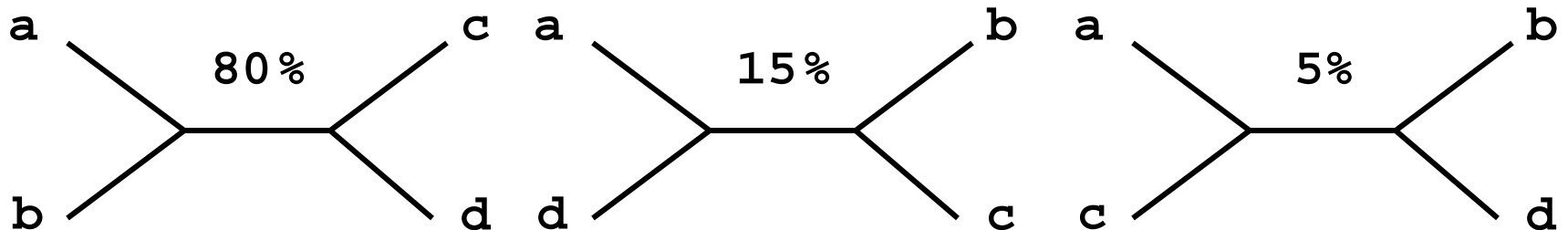
Maximum Likelihood

- ML tree: find the tree (and parameters) that maximize the likelihood

■ Tree	score:-log likelihood	parameters
ab cd	30.34	((a:0,b:.55):.35,(c:.55,d:0))
ac bd	31.02	((a:0,c:7.9):0,(b:.55,d:.35))
ad bc	31.02	((a:0,d:.35):.27,(b:.27,c:7.6))

Bootstrap with Maximum Likelihood

- 12 volunteers:
choose a number between 1 and 12
- Get a new alignment and a new ML tree
- Repeat...



Ingredients of molecular evolution models

- Topology
- Branch lengths: in number of substitutions per site. Usually unconstrained (no clock)
- Substitution rate matrix

A	-1	.33	.33	.33
C	.33	-1	.33	.33
G	.33	.33	-1	.33
T	.33	.33	.33	-1

JC

-1	.20	.60	.20
.20	-1	.20	.60
.60	.20	-1	.20
.20	.60	.20	-1

K80, k=3

-.85	.40	.35	.10
.15	-.60	.35	.10
.15	.40	-.65	.10
.15	.40	.35	-.90

F81, unequal base freq.

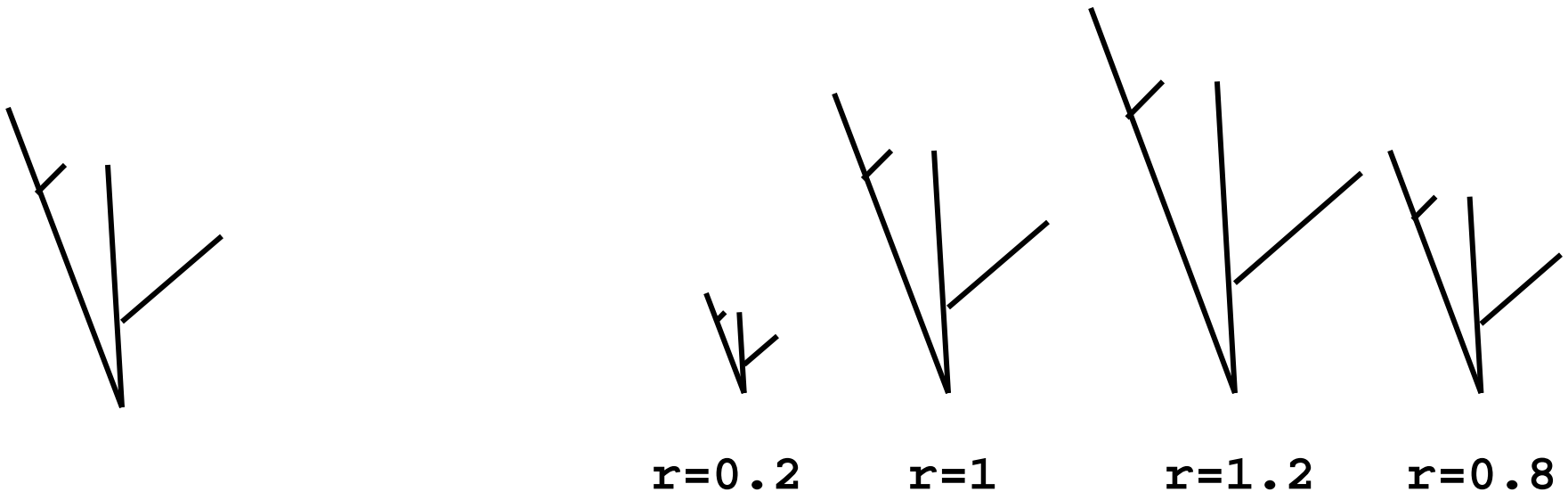
Ingredients of molecular evolution models

Common models: reversible. The root doesn't matter.

	Base freq.	ti/tv ratio	# param
JC	equal	1	0
K80	equal	free	1
F81	free	1	3
HKY85	free	free	4
F84	free	free	4
TN93	free	two ti/tv ratios are free	5
GTR	free	All rate ratios are free	8

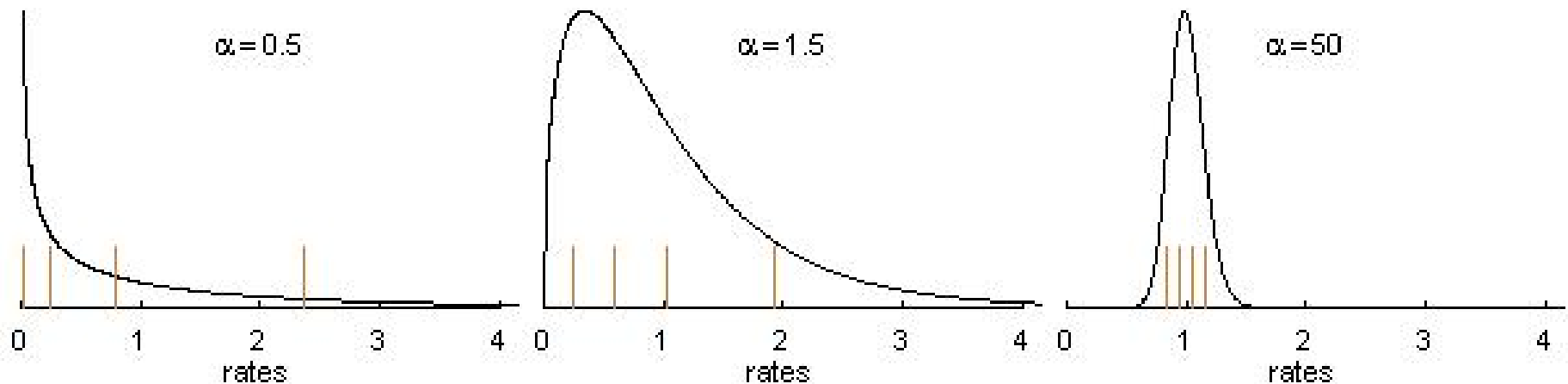
Ingredients of molecular evolution models

- Distribution of rates across sites
 - Some sites are slow, some are fast
 - Model: each site chooses a rate r at random, from some distribution



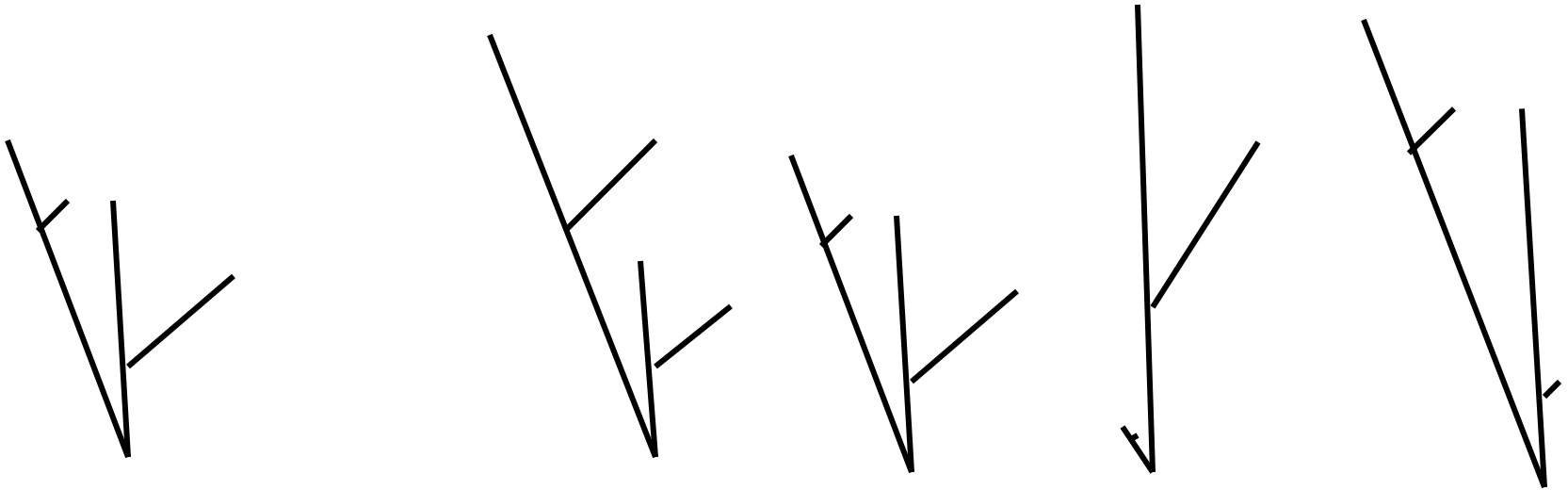
Ingredients of molecular evolution models

- Distribution of rates across sites
 - Model: Gamma distribution of rates, continuous and discretized for calculations



Ingredients of molecular evolution models

- Distribution of rates across lineages
 - Covarion models: hidden on/off process for absence/presence of constraint. 1-3 parameters.



Ingredients of molecular evolution models

- Partitions in the data
 - 1st, 2nd and 3rd codon positions
 - several genes are available,
 - or several exons and introns,
 - or nuclear + mt DNA.

 - Allow 'blocks' to have their own models and/or parameters
-

Estimation of Divergence Times

- Times and rates are confounded. Sites can
 - evolve for 1 million years at $5 \cdot 10^{-7}$ subst./year/site
 - or for 10,000 years at $5 \cdot 10^{-5}$ subst./year/site, the alignment will look the same: 0.5 subst./site on average.
 - Difficult problem:
 - fossil calibrations usually few and uncertain
 - rates vary a lot
-