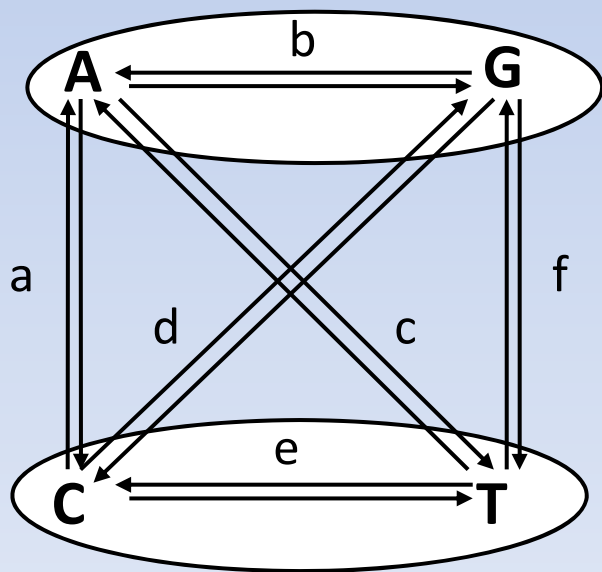
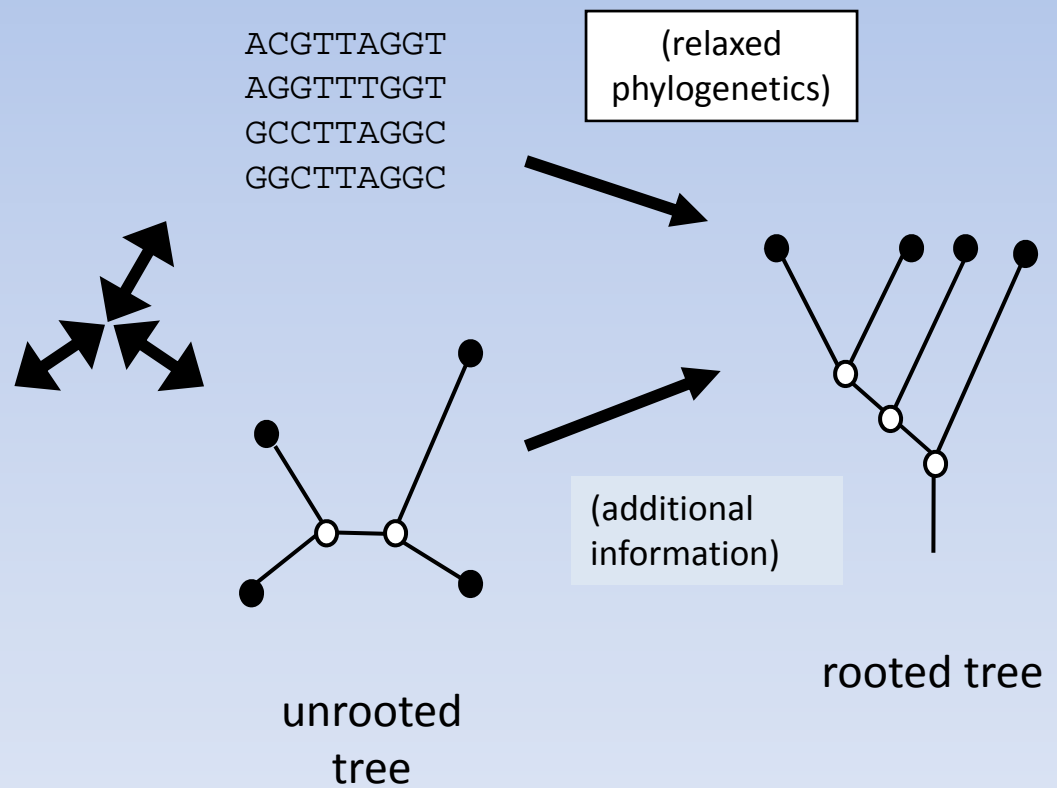


Q1. What is an evolutionary model?



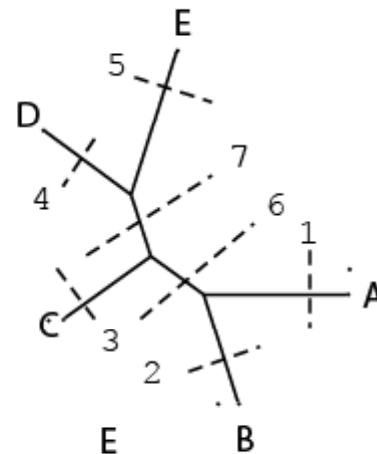
substitution model



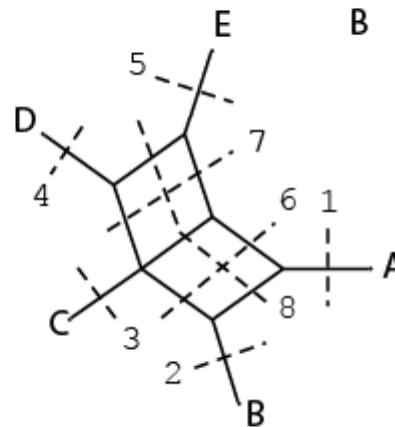
Q2. Is it okay to assume a bifurcating tree as part of the evolutionary model?

A	aaaaaaa	1234567
B	ggaaaaa	*****
C	gagaaga	..*****
D	gaagagg	.*.***.*
E	gaaaggg	**.***..
		***....

A	aaaaaaaa	12345678
B	ggaaaaag	*****
C	gagaagag	..*****.
D	gaagaggg	.*.***.*.
E	gaaagggg	**.***...*
		***....*

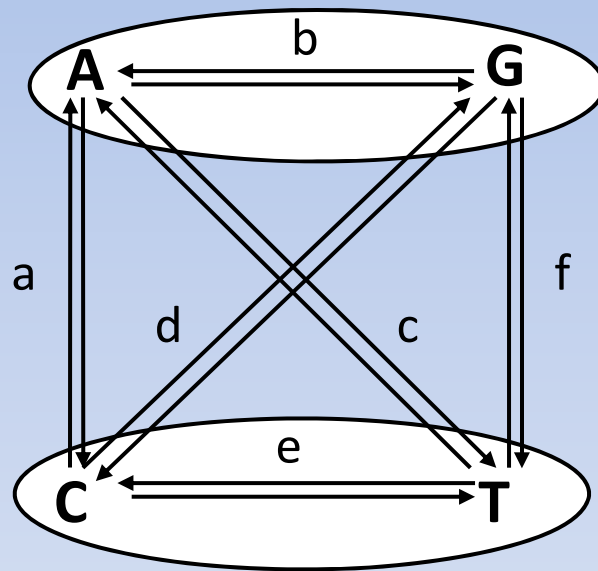


compatible splits



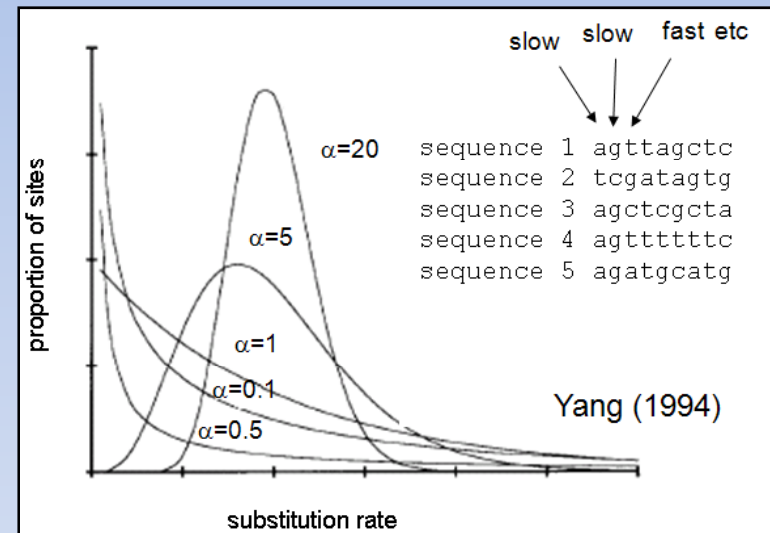
incompatible splits

Q3. What are substitution models?



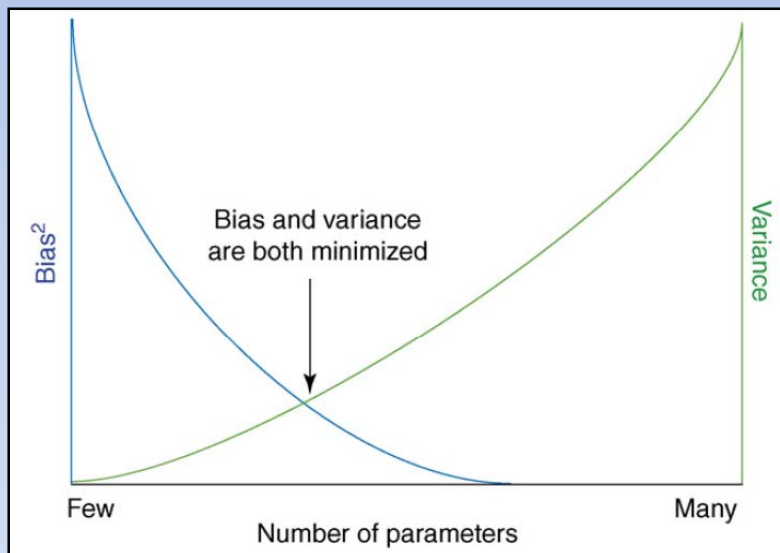
substitution model

+



Lio and Goldman (1998) Genome Research 8 1233-1244
Whelan et al. (2001) TIGS 17 262-272

Q4. What are important issues for picking the best one?



Kelchner and Thomas (2006)
TREE Vol 22 p87-94

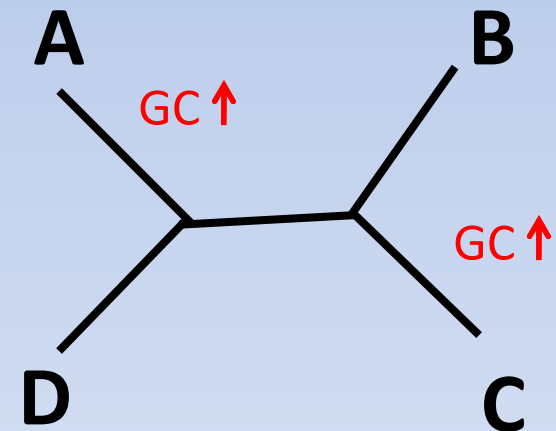
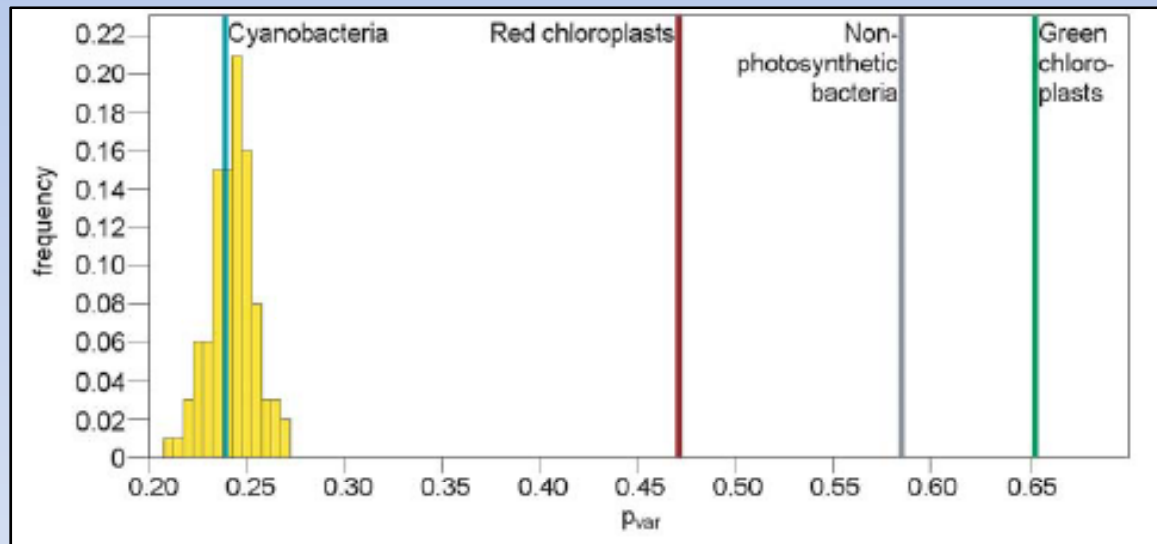
Are you trying to fit
an elephant?

Steel (2005) TIGS 21 307-309



Joyce and Sullivan (2005) Ann
Rev Ecol 36 445-66

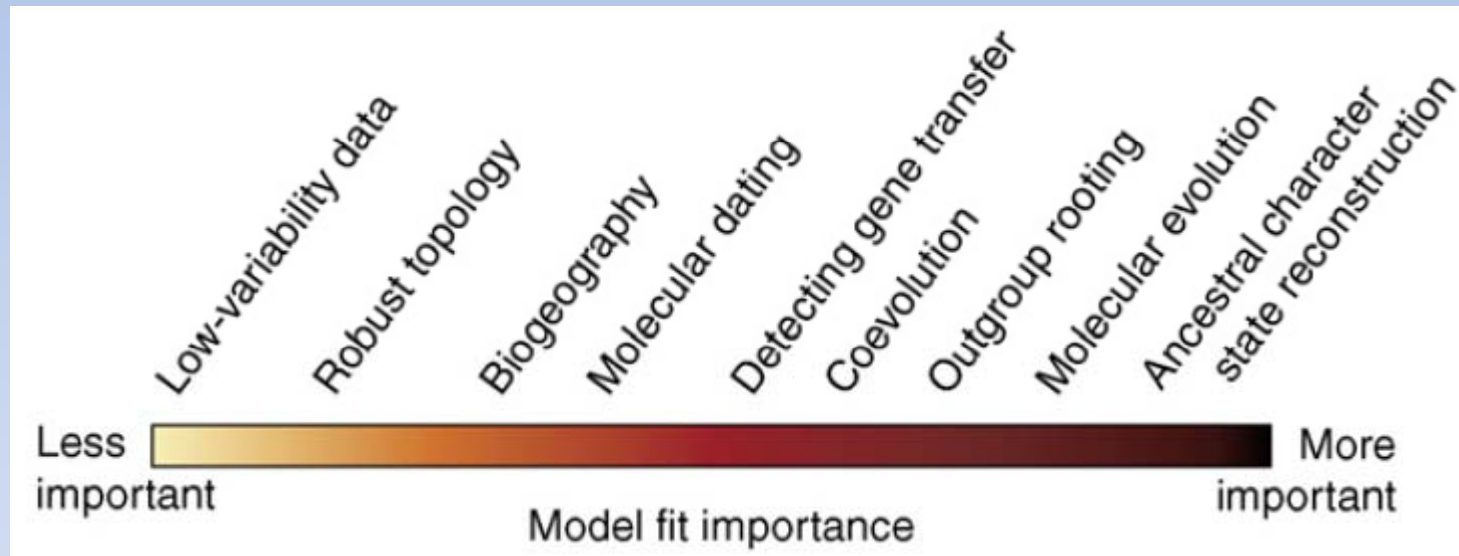
Q5. In what way might sequence evolution deviate from a time reversible stationary process?



Lineage specific changes in
a) evolutionary constraint – change in pvar
b) compositional heterogeneity

Foster (2004) Syst Biol
53 485-495
Lockhart et al. (2006)
Mol Biol Evol 23, 40-45

Q6. When does substitution model fit matter?



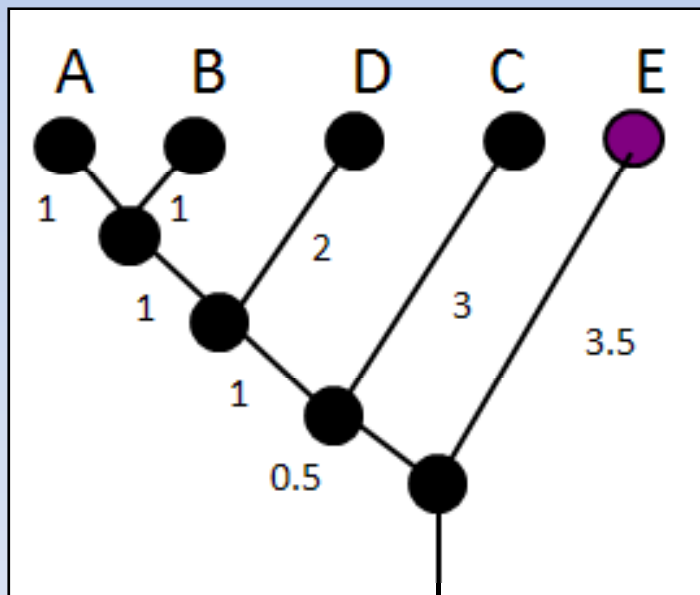
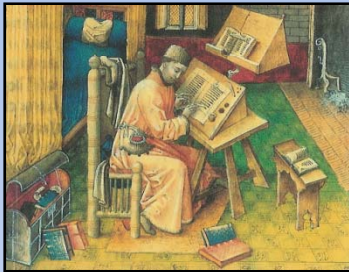
Kelchner and Thomas (2006) TREE Vol 22 p87-94

This is different from the question:

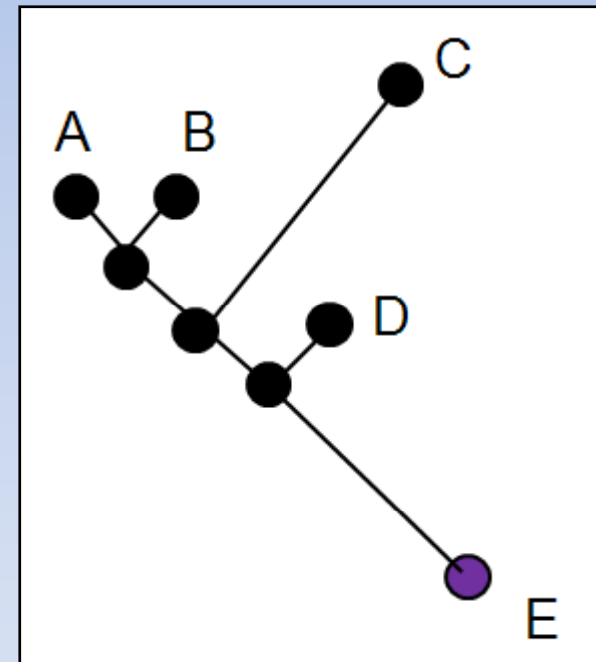
Q7. Does the evolutionary model have a sufficiently good fit to the data to be confident of phylogenetic inference?

e.g. Bollback (2002) MBE 19, 1171-1180

Q8. What are popular tree selection criteria?

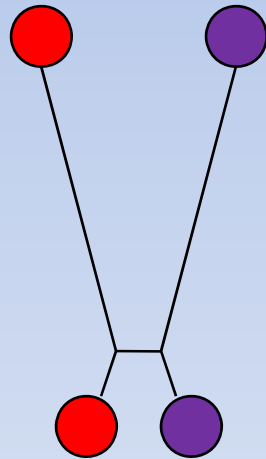


UPGMA

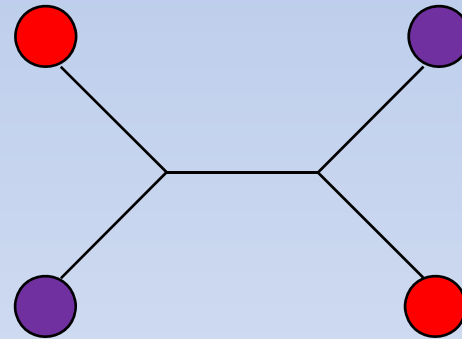


NJ

Maximum parsimony



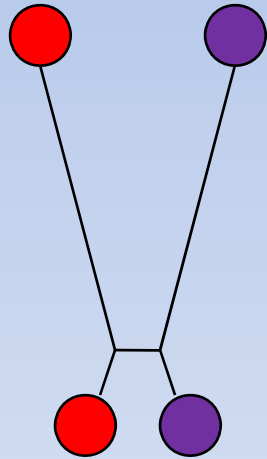
Parsimony



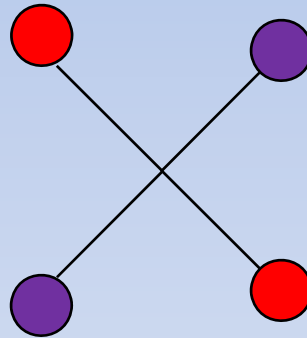
LBA

Felsenstein (1978) Syst
Zool 27 401-10
Hendy and Penny (1989)
38 297-309

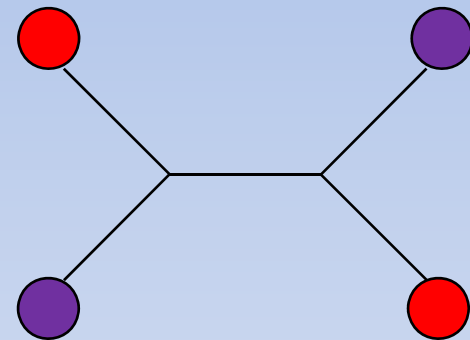
Maximum likelihood



ML



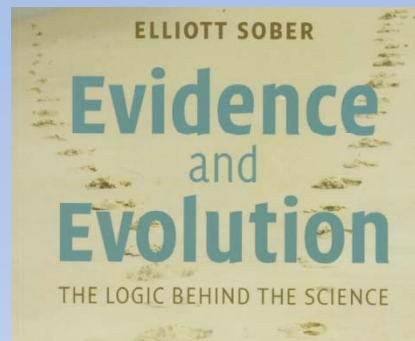
model fit



model misspecified
LBA

Felsenstein (1978) *Syst Zool*
27 401-10
Lockhart et al (1996) *PNAS USA*
93, 1930-1934

Bayesian inference and likelihood



O: you hear a noise coming from the attic...

H: Your hypothesis is that there are gremlins up there bowling



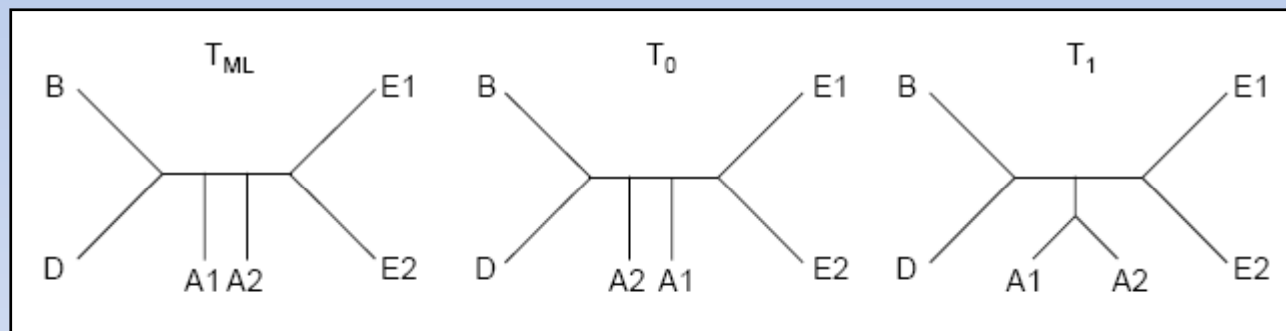
Posterior probability

Given some roof noise – it is not that probable that there are gremlins up there: $\Pr(H|O)$ is **low**

Likelihood

Given gremlins bowling on the roof, it is probable there will be noise up there : $\Pr(O|H)$ is **high**

Q9. Are there enough patterns in the data to be confident of your phylogenetic estimates?



Original InL scores	-5069.85	-5073.46	-5073.75
Non-parametric bootstrap	70%	16%	14%
SH test	1	0.3	0.26
SWOH	-	0.001	0.002