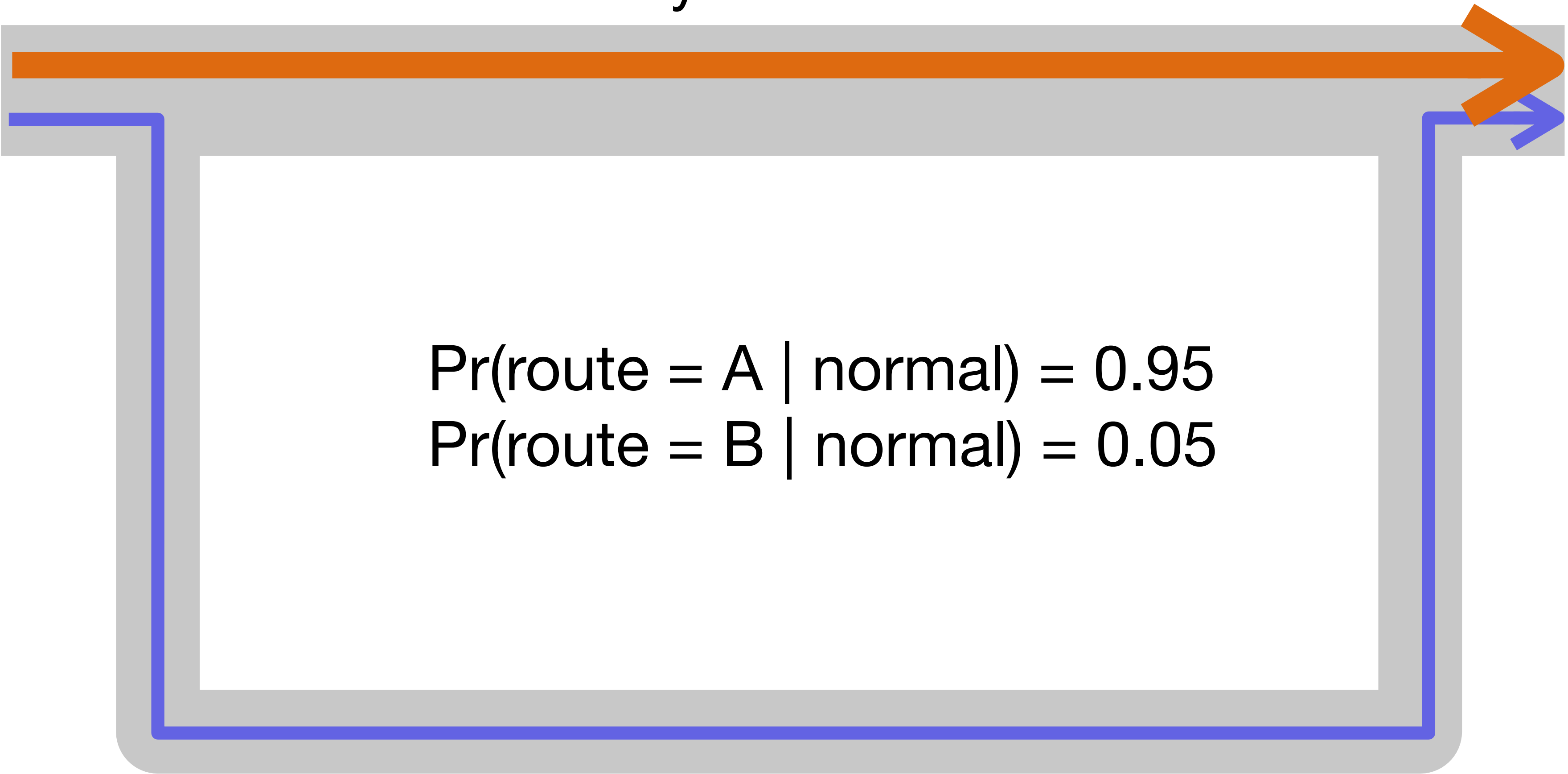
Dependence Example

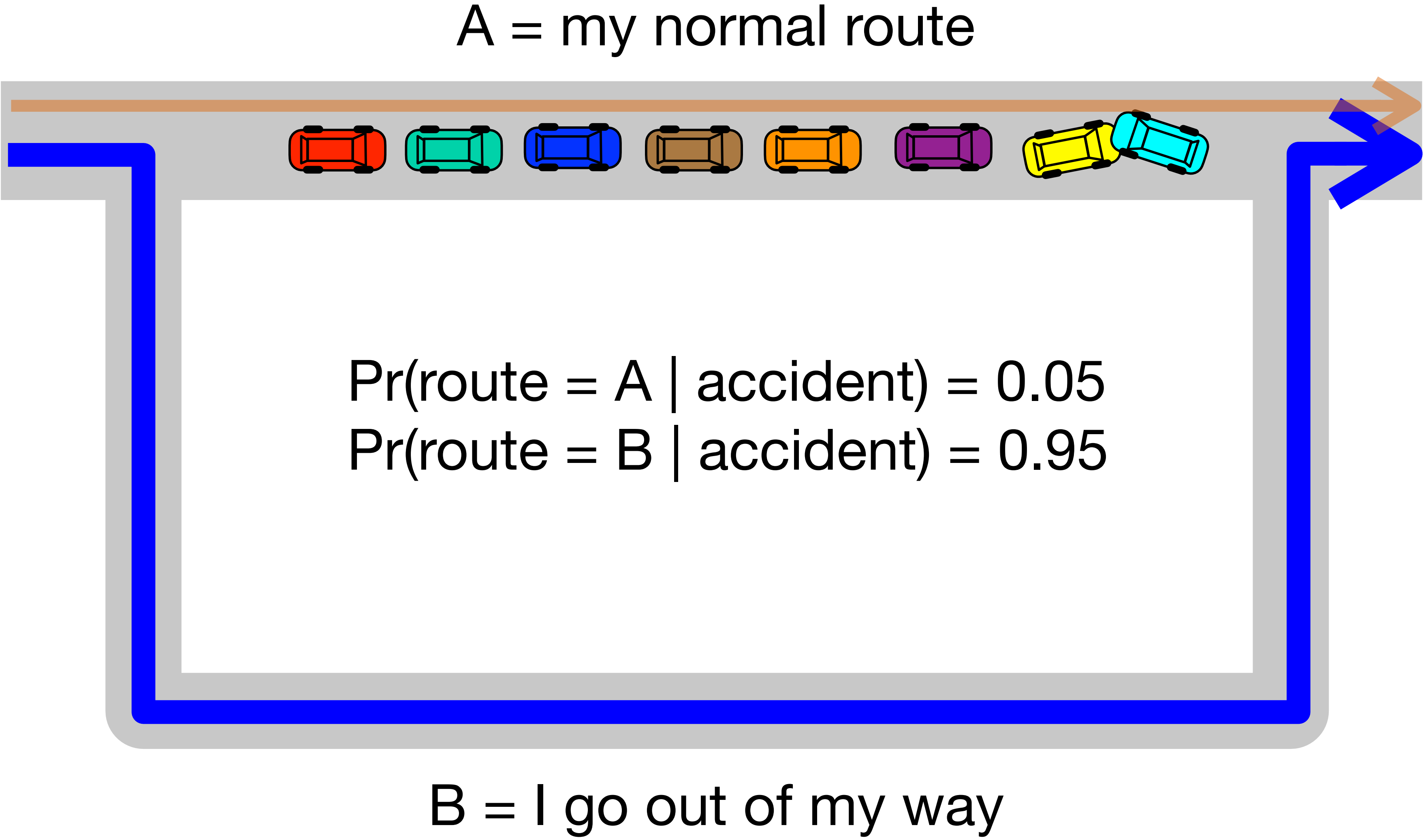


A = my normal route

Pr(route = A | normal) = 0.95
Pr(route = B | normal) = 0.05

B = I go out of my way

Dependence Example



A = my normal route

Pr(route = A | accident) = 0.05
Pr(route = B | accident) = 0.95

B = I go out of my way

How frequently do I take route B?

(assume $\Pr(\text{accident}) = 0.1$)

Likelihood of the simplest tree

First 32 nucleotides of the ψη-globin gene of gorilla and orangutan:

```
gorilla   GAAGTCCTTGAGAAATAAACTGCACACACTGG
orangutan GGACTCCTTGAGAAATAAACTGCACACACTGG
```

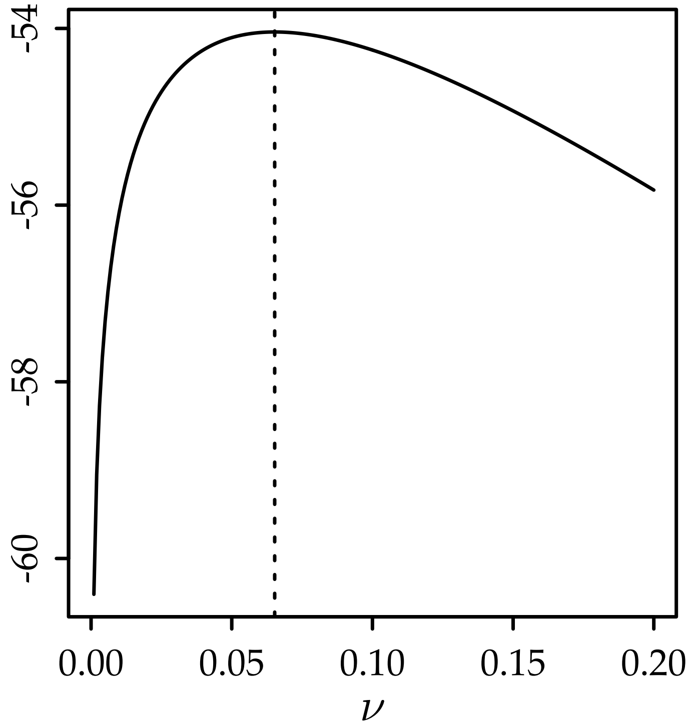
first site: G <--> G

second site: A <--> G

Maximum likelihood estimation

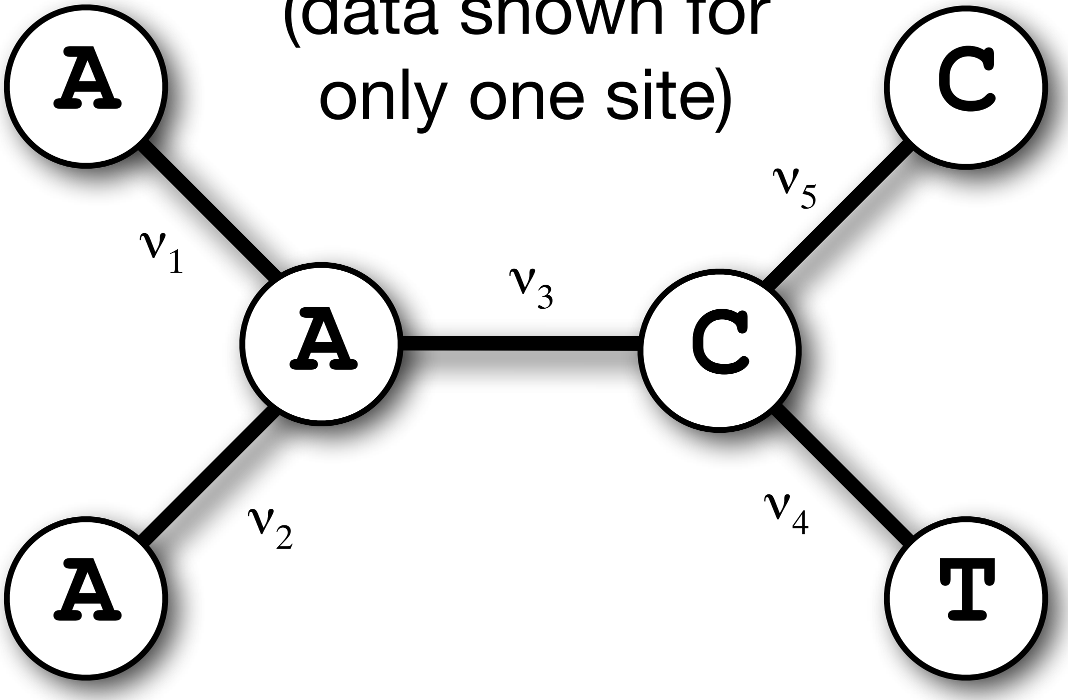
First 32 nucleotides of the ψη-globin gene of gorilla and orangutan:

```
gorilla   GAAGTCCTTGAGAAATAAACTGCACACACTGG
orangutan GGACTCCTTGAGAAATAAACTGCACACACTGG
```



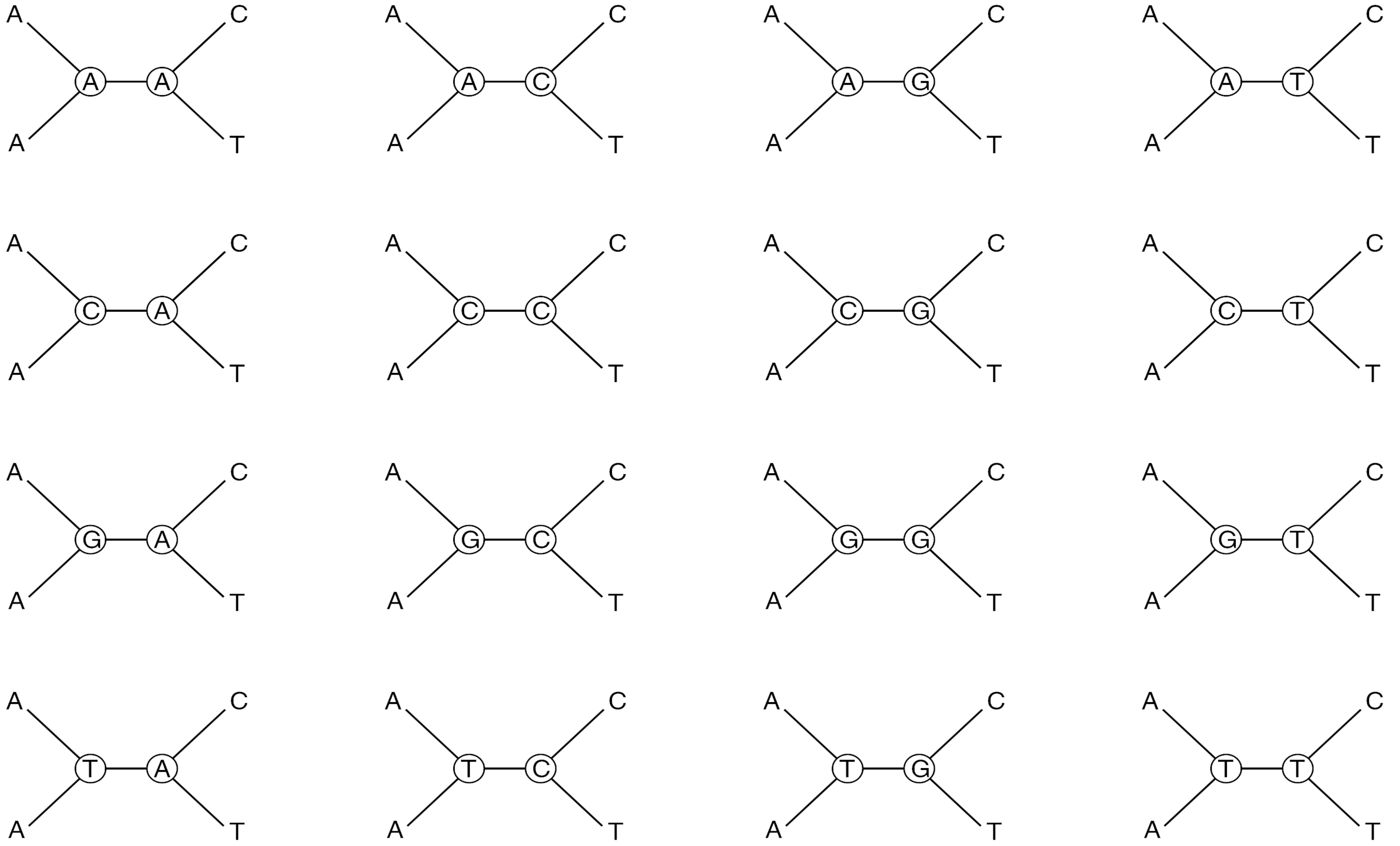
0.06525853

Likelihood of an unrooted tree



(data shown for
only one site)

Brute force vs pruning algorithm



Felsenstein (1981)

Putting it all together

- Site likelihoods involve products of transition probabilities, summed over ancestral states
- Overall log-likelihood (log*L*) for a tree is sum of site log-likelihoods
- Overall log-likelihood for each tree must be *maximized*!
  - must find MLEs for all edge lengths and all model parameters
  - this involves computing the overall log-likelihood many times as different combinations of parameter values are tried
- The *maximized* log*L* for a given tree topology is then compared to the *maximized* log*L* of other tree topologies in the context of a search

site-likelihoods likelihood maximize likelihood

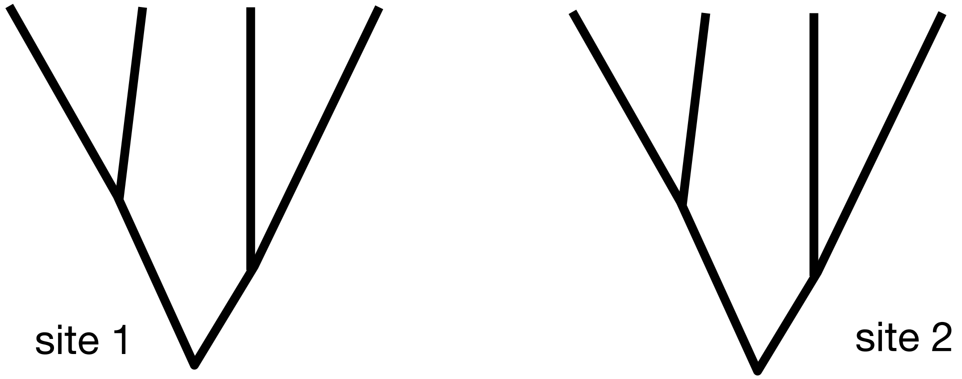
Green Plant rbc*L*

First 88 amino acids (translation is for *Zea mays*)

```
M--S--P--Q--T--E--T--K--A--S--V--G--F--K--A--G--V--K--D--Y--K--L--T--Y--Y--T--P--E--Y--E--T--K--D--T--D--I--L--A--A--F--R--V--T--P--
Chara            (green alga; land plant lineage)  AAAGATTACAGATTAACTTACTATACTCCTGAGTATAAAACTAAAGATACTGACATTTTAGCTGCATTTCGTGTAACTCCA
Chlorella        (green alga)                      .....C...C.T....................T..CC..C.A.....C.....T...C.T..A..G..C...A.G.....T
Volvox           (green alga)                      ........TC.T.....A.....C..A.....C...GT.GTA.....C........C.....A.........A.G......
Conocephalum     (liverwort)                       ........TC..........T........G..T...G.............G..T........A......A.AA.G.....T
Bazzania         (moss)                            ........T........C..T.....G.....A...G.G..C.....G..A..T.....G..A.........A.G.....C
Anthoceros       (hornwort)                        ........T........CC.T.....C.....T..CG.G..C..G........T.....G..A..G.C.T.AA.G.....T
Osmunda          (fern)                            ........TC....G...C..........C..T...G.G..C..G........T.....G..A.....C..AA.G.....C
Lycopodium       (club "moss")                     .GG...............C.T..C........T.....G..C.....A..C..T...C.G..A........AA.G.....T
Ginkgo           (gymnosperm; Ginkgo biloba)       ..............G.....T...........A...C....C...........T..C..G..A.....C..A........T
Picea            (gymnosperm; spruce)              ....................T...........A...C.G..C........G..T.....G..A.....C..A........T
Iris             (flowering plant)                 ..............G.....T...........T..CG....C...........T..C..G..A.....C..A........T
Asplenium        (fern; spleenwort)                ........TC..C.G.....T..C..C..C..A..C..G..C........C..T..C..G..A..T..C..GA.G..C...
Nicotiana        (flowering plant; tobacco)        .....G....A...G.....T..............CC....C..G........T..A..G..A.....C..A........T

Q--L--G--V--P--P--E--E--A--G--A--A--V--A--A--E--S--S--T--G--T--W--T--T--V--W--T--D--G--L--T--S--L--D--R--Y--K--G--R--C--Y--H--I--E--
CAACCTGGCGTTCCACCTGAAGAAGCAGGGGCTGCAGTAGCTGCAGAATCTTCTACTGGTACATGGACTACTGTTTGGACTGACGGATTAACTAGTTTGGACCGATACAAAGGAAGATGCTACGATATTGAA
.....A..T........A........G..T..G........A........A..A........T.....G.....A........T..T...........A.....T........TC.T..T..T..C..C..G
.....A..T...............TGT..T.....T..T.....T.....A..A..A.....T.....A.....A........T..T.....A...C.T.....T........TC.T..T..T..C..C..G
..G.....G..A...G.A...........A..A.....T.....T..........................A...........T..TC.T....ACC.T..T..T..T.....TC.......T.G......C
.....G..A..A.................A..G...........T........A..C.....G.....C..G........C..T..GC.T..A...C.C..T..T........TC.......T..C..C...
T....A..G..G.................A..C...........T........A..........................C..T...C.T..C..CC.T.....T........TC..........C......
.....C..A..A..GG....G.....T..A..............G...........A.....G.....C.....A.....G..T...C.T..C...C.T..T..T..T..G..TC.................
....T...A..A.....C..G.....G..A..C...........T........C..........................C..T...C.T..C...C.C..T..C........TC.G.....T..A......
........A..G........G.....G..A..............C........C..............C...........C..T...C.T..C...C.T..T..T.....G...........T..C..C..G
.....A..G..G..G..C..G.....G..A..A...........T........C..C...........C...........C..T...C.T......C.T..T..T.....G..GC.......T..C..C..G
.....C..A.....TG..........G.....C..G........C.......................A..A..G........T...C.T..C...C.T..T..T.........C........C.C..C..G
.....C..A..A...G..........C..A.................G..C.....A...........C.....G.....A.....G..G..C..CC.T.....T.....G..CC.............C..G
........A.......................C..G........C.......................A.....A.....C..T...C.T..C..CC.T..T..T........GC........CGC..C..G
```

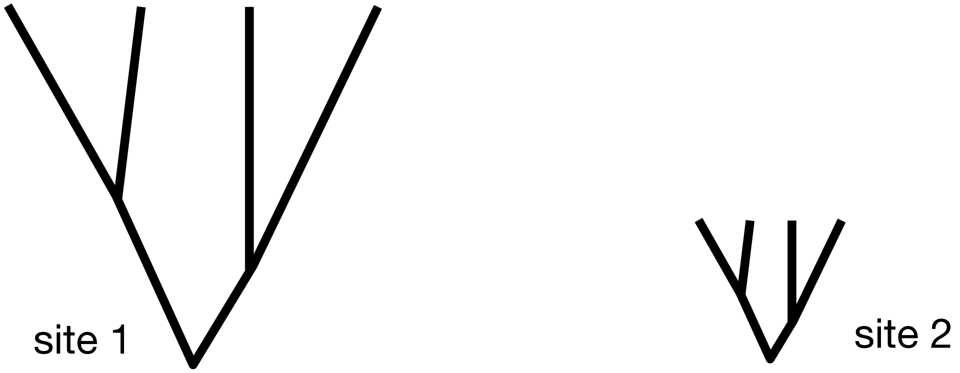
Rate homogeneity



All edges same for every site

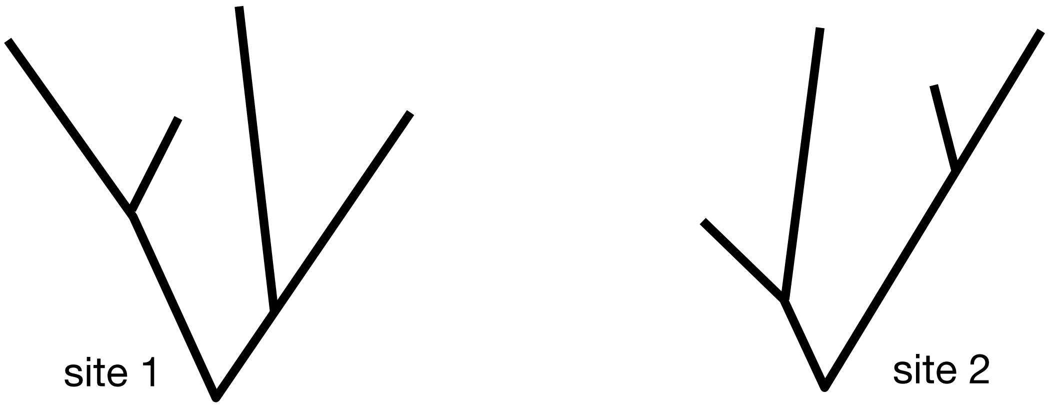
Among-site rate variation (ASRV)

(this is modeled by site-specific rates, discrete gamma (G), or proportion of invariable sites (I) models)



Rate varies across sites but edge lengths proportional (tree scales up or down across sites)

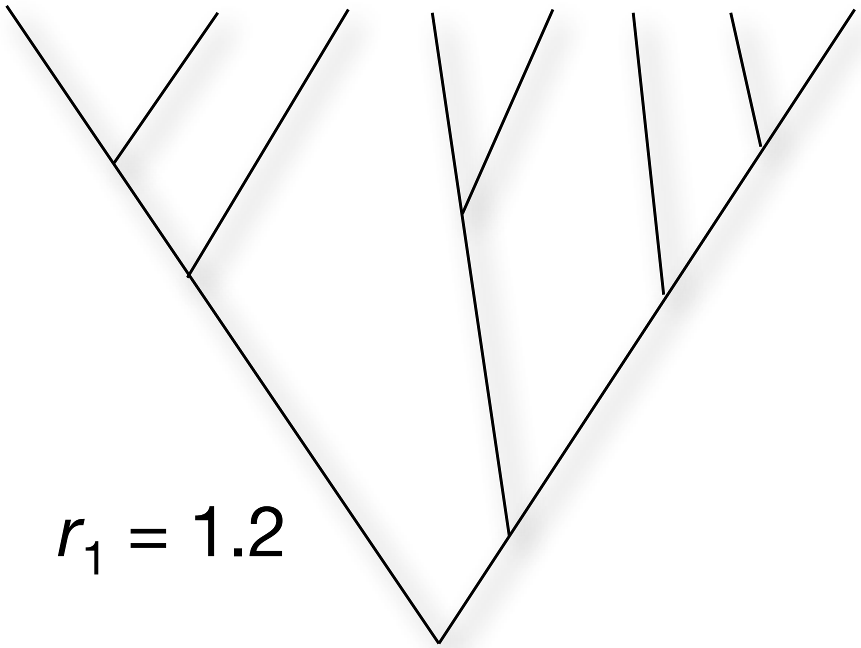
Heterotachy



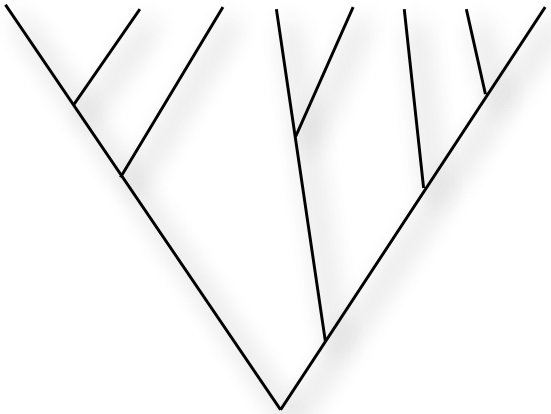
Rates vary but not simple scaling
(rate for a particular edge slow for
some sites, fast for others)

Site-specific rates

$$L = \Pr(D_1|r_1)\cdots\Pr(D_{1000}|r_1)\ \Pr(D_{1001}|r_2)\cdots\Pr(D_{2000}|r_2)$$



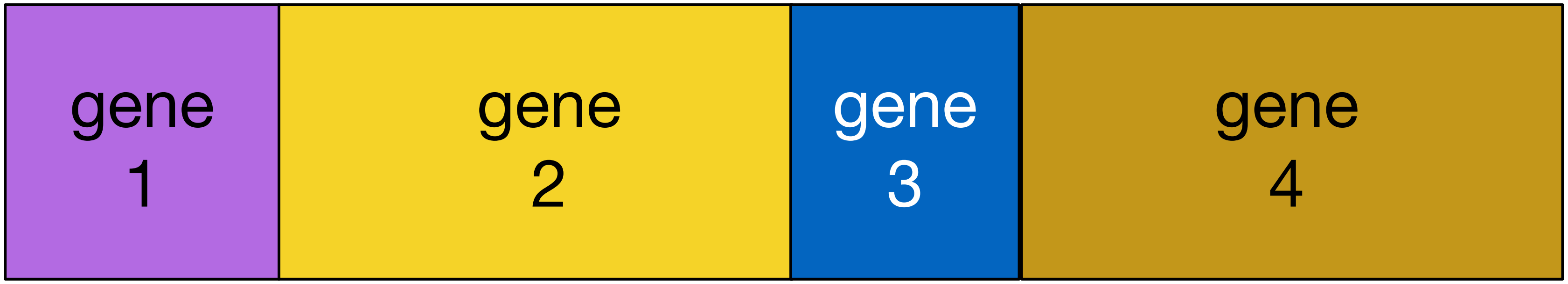
$r_1 = 1.2$



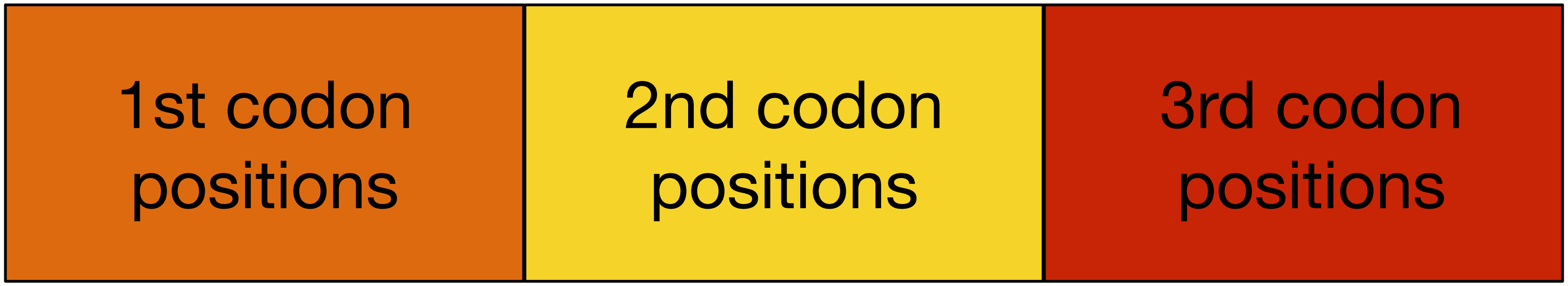
$r_2 = 0.8$

Site-specific rates

Partitioned models



partitioned by gene, 4 subsets



partitioned by codon position, 3 subsets

Mixture Models

free rates/probs

Invariable sites (I) model

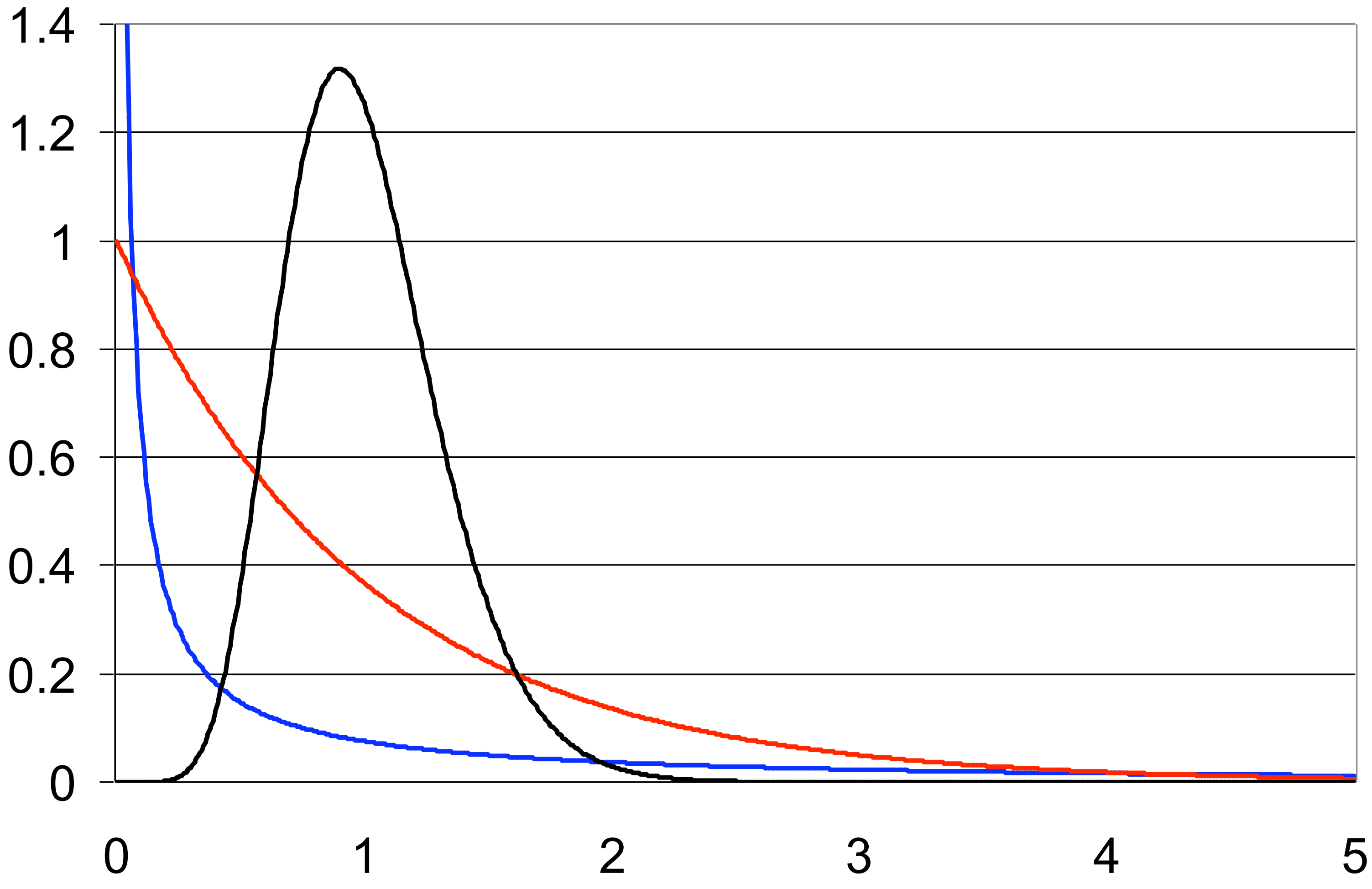
pinvar   constant vs. variable sites   Reeves (1992)

Discrete Gamma (G) model

mixture model formula

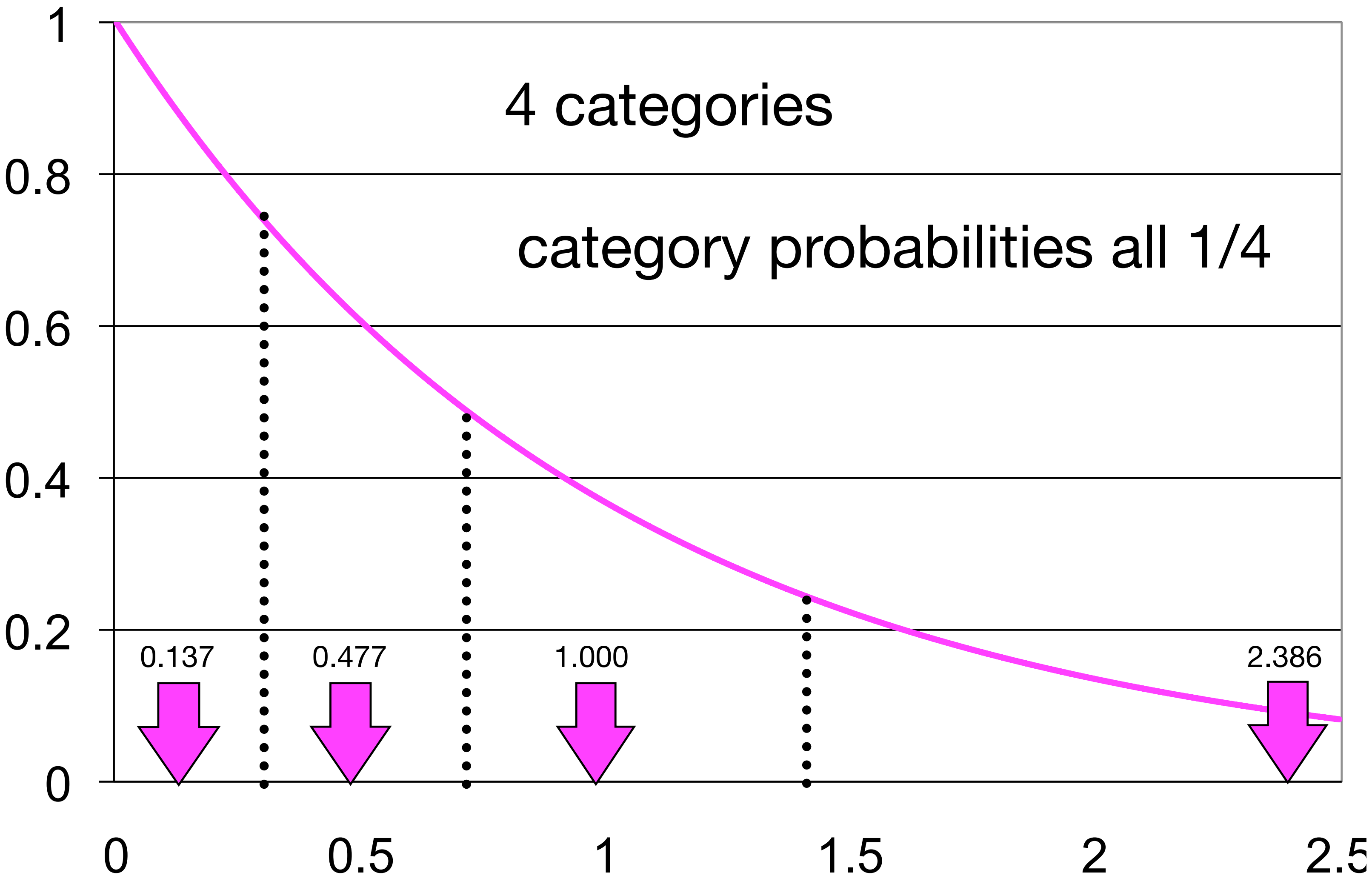
Yang (1993, 1994)

Gamma distributions



mean = 1 axis labels shape = 1/variance

Discrete Gamma (+G)



Free rates (+R)

category	relative rates	category probabilities
1	r_1	p_1
2	r_2	p_2
3	r_3	p_3
4	r_4	p_4

$$p_1 + p_2 + p_3 + p_4 = 1$$

$$p_1 r_1 + p_2 r_2 + p_3 r_3 + p_4 r_4 = 1$$