

Green Plant rbcL

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



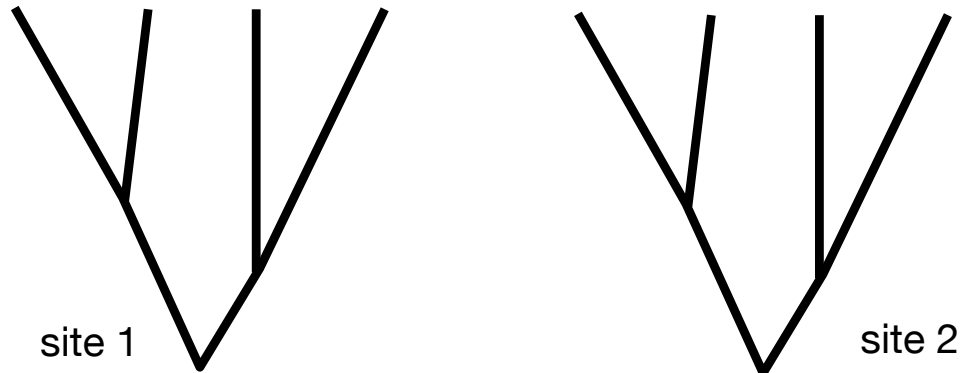
```

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) AAAGATTACAGATTAACCTACTATACTCCTGAGTATAAAACTAAAGATACTGACATTTTAGCTGCATTTCGTGTAACCTCCA
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.....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..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--
CAACCTGGCGTTCCACCTGAAGAAGCAGGGGCTGCAGTAGCTGCAGAATCTTCTACTGGTACATGGACTACTGTTTGGACTGACGGATTAAGTAGTTTGGACCGATAACAAGGAAGATGCTACGATATTGAA
.....A..T.....A.....G..T..G.....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.....G..T....C..T..C..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..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
  
```

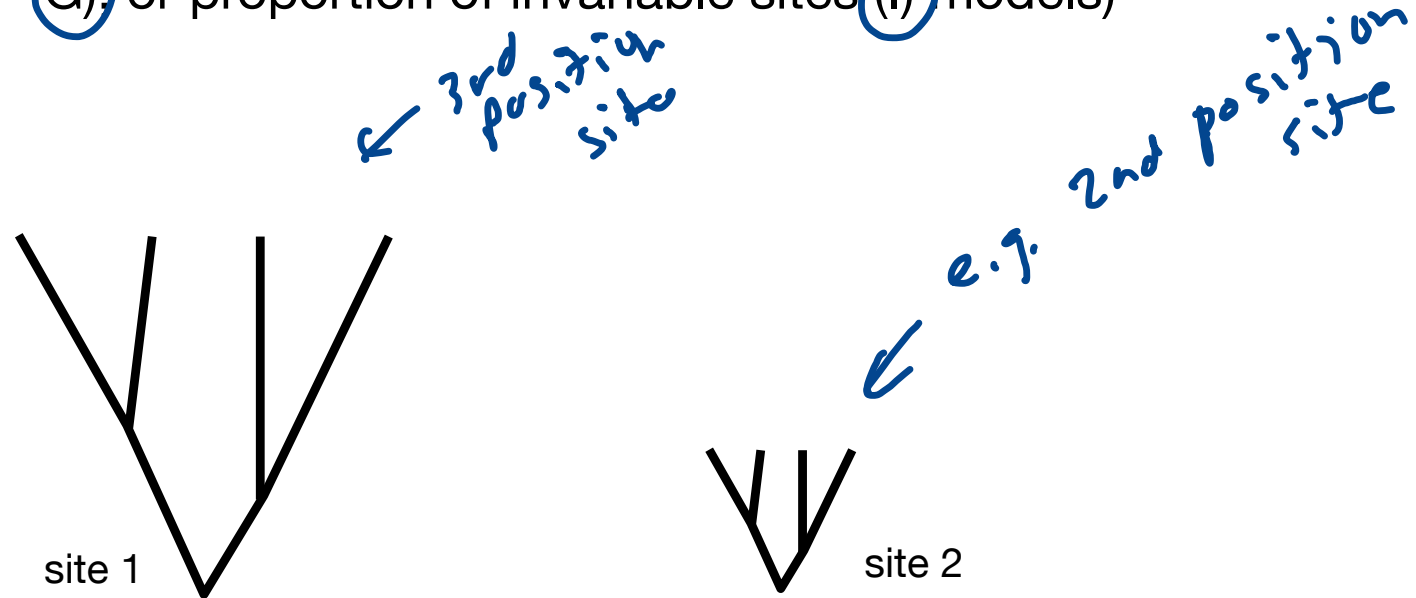
Rate homogeneity



All edges same for every site

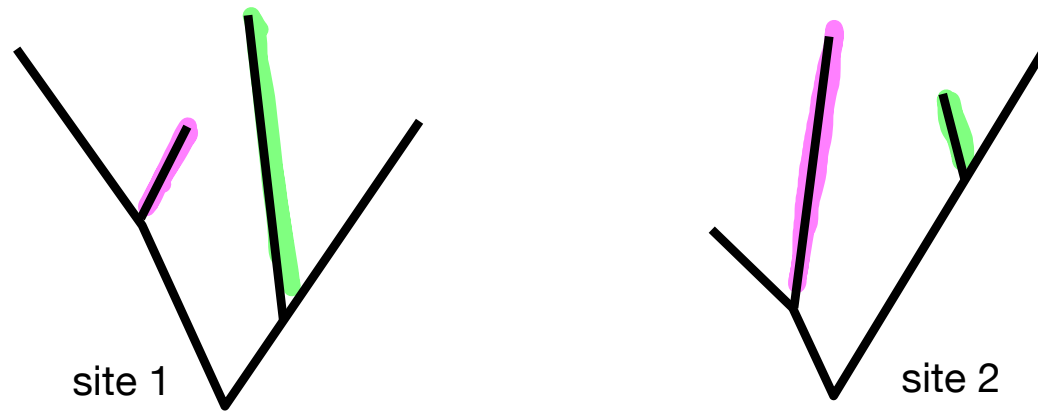
Among-site rate variation (ASRV)

(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)

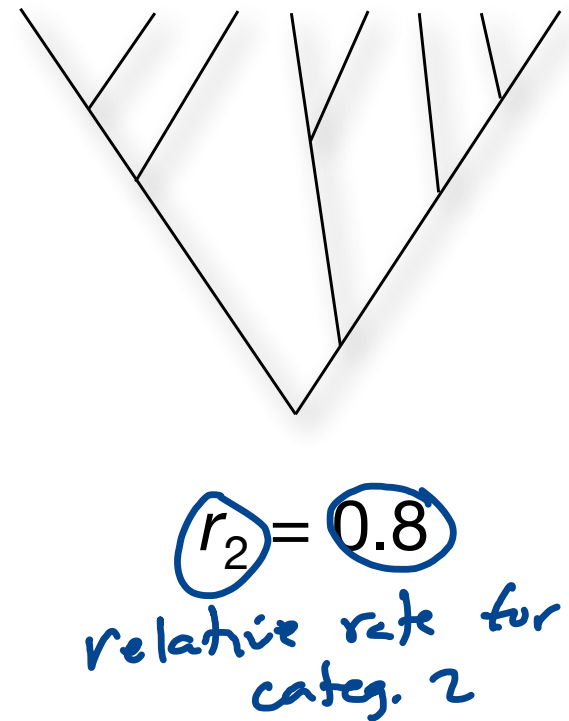
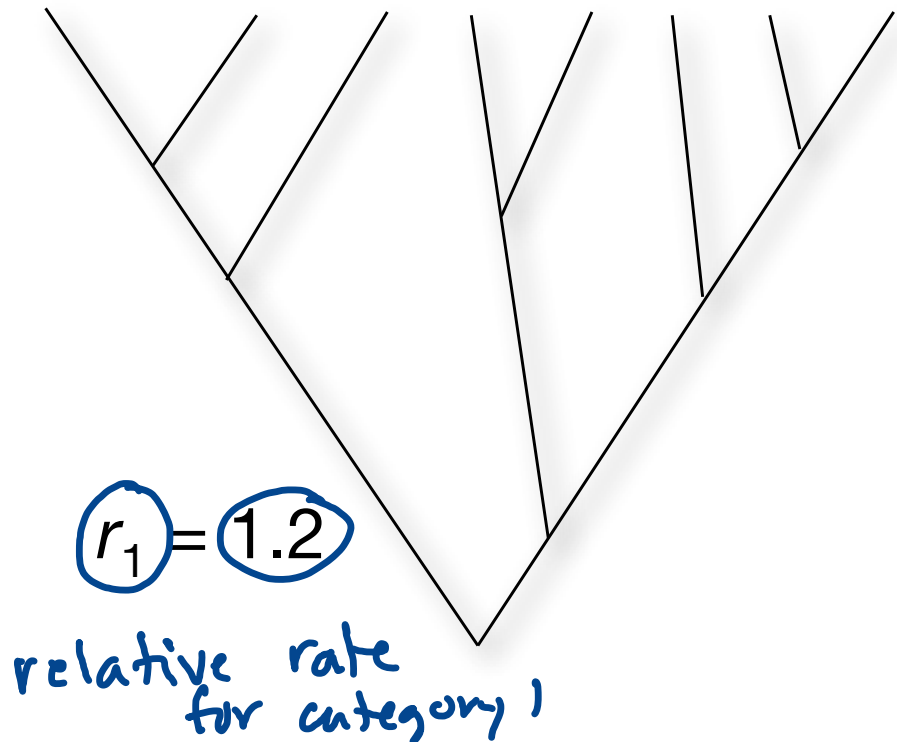
Heterotachy



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

Site-specific rates

$$L = \left(\overset{\text{1000 sites } r_1}{\Pr(D_1|r_1) \cdots \Pr(D_{1000}|r_1)} \right) \left(\overset{\text{1000 sites } r_2}{\Pr(D_{1001}|r_2) \cdots \Pr(D_{2000}|r_2)} \right)$$



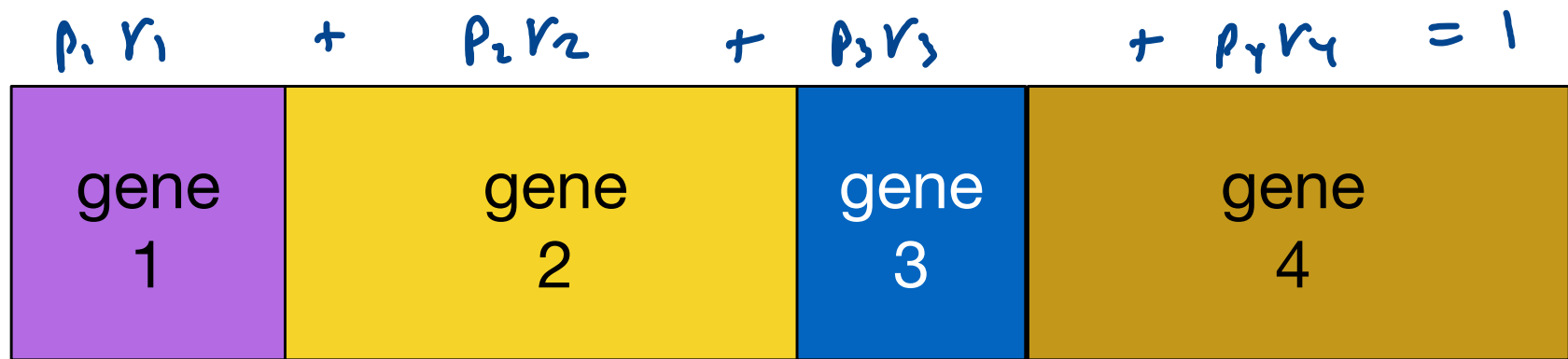
Site-specific rates

$$P_{ij} = \frac{\frac{1}{4} - \frac{1}{4} e^{-\frac{4\nu r_i}{3}}}{\quad}$$

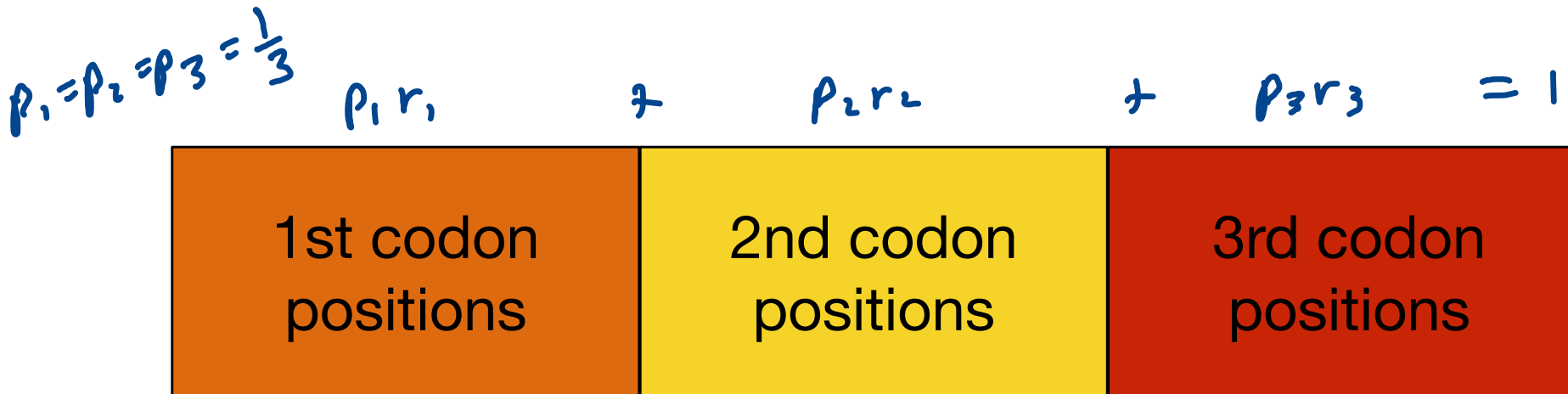
use relative rate r_i

$$P_{ii} = \frac{1}{4} + \frac{3}{4} e^{-\frac{4\nu r_i}{3}}$$

Partitioned models



partitioned by gene, 4 subsets



partitioned by codon position, 3 subsets

G, I models Mixture Models

$$L^{(k)} = P_1 L_{(1)}^{(k)} + P_2 L_{(2)}^{(k)} + \dots + P_m L_{(m)}^{(k)}$$

↑
↑
↑
↑
↑

using r_1
using r_2
using r_m

Invariable sites (I) model

$$L^{(k)} = p_{\text{invar}} L_1^{(k)} + (1 - p_{\text{invar}}) L_2^{(k)}$$

↑
proportions of
invariable sites

$$r_1 = 0$$

$$r_2 =$$

$$\underline{p_{\text{invar}}} r_1 + \underline{(1 - p_{\text{invar}})} r_2 = 1$$

$$\cancel{(p_{\text{invar}})}(0) + (1 - p_{\text{invar}}) r_2 = 1$$

$$r_2 = \frac{1}{1 - p_{\text{invar}}} \leftarrow$$

Discrete Gamma (G) model

4 categories

r_1, r_2, r_3, r_4

$p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$

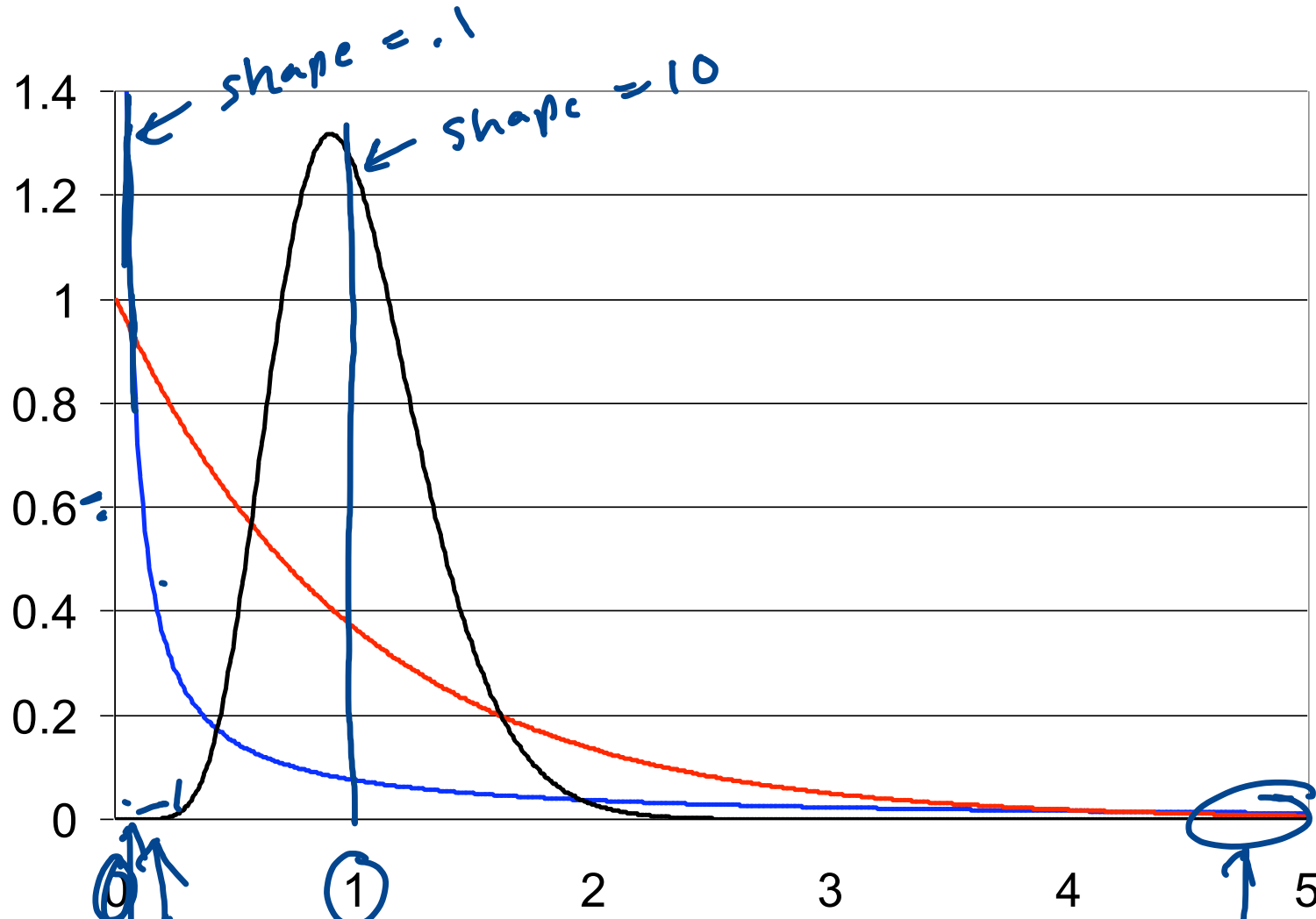
$$L^G = \frac{1}{4} L_{\textcircled{1}}^{(k)} + \frac{1}{4} L_2^{(k)} + \frac{1}{4} L_3^{(k)} + \frac{1}{4} L_{\textcircled{4}}^{(k)}$$

↑
using r_1

$$+ \frac{1}{4} L_{\textcircled{4}}^{(k)}$$

↑
using r_4

Gamma distributions



0.2

most sites evolving slowly

Some sites evolving very fast

STOPPED HERE 2024-02-08

mean = 1

axis labels

shape = 1/variance

Relative rates in 4-category case

