



Substitution model averaging

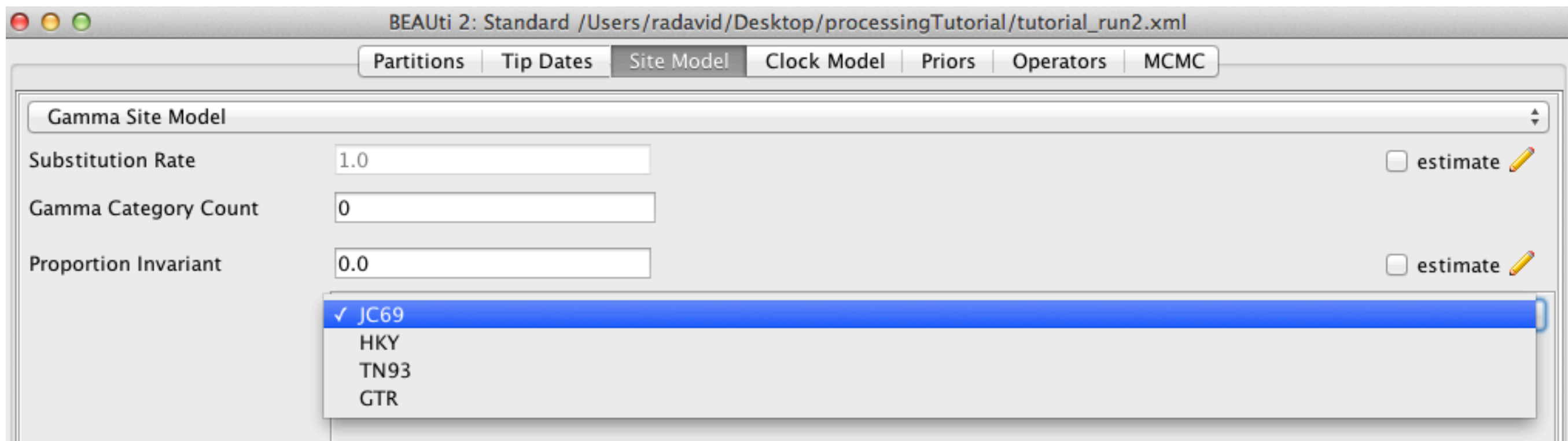
with bModelTest



Taming the BEAST - bModelTest tutorial

Big question when setting up a BEAST2 analysis:

Which nucleotide substitution model do I have to chose?



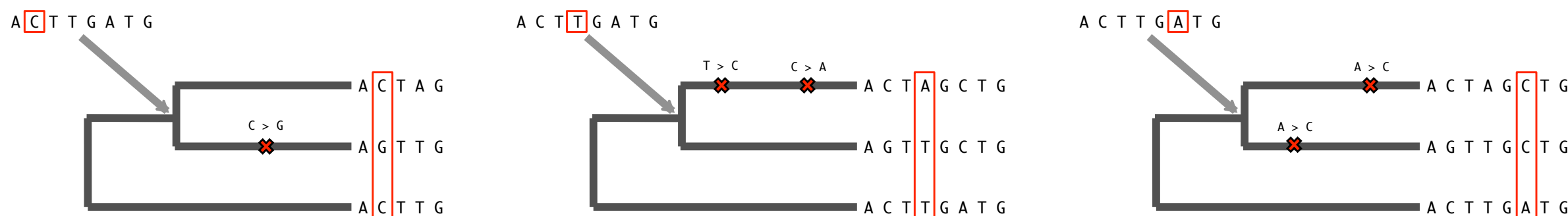
▶ Let the data tell us: bModel Test



Taming the BEAST - bModelTest tutorial

Recap: Nucleotide substitution models

I. Why do we need nucleotide models at all?



Problem of phylogenetics:

We observe sequences but not their evolutionary history. Thus we have to take all possible evolutionary trajectories into account (Markov chain model).

Transition probability matrix:

$$P(t) = \left(p_{XY}(t) \right)_{X,Y \in \{A,C,T,G\}} = e^{Qt}$$

matrix with
substitution
rates



Taming the BEAST - bModelTest tutorial

Bayes' formula for phylodynamics: (presentation of Louis du Plessis)

nucleotide substitution rate matrix

$$P(\text{tree} \mid \text{data}) = \frac{P(\text{data} \mid \text{tree}) P(\text{tree}) P(\text{model})}{P(\text{model})}$$

The equation is annotated with blue arrows pointing to the following components:

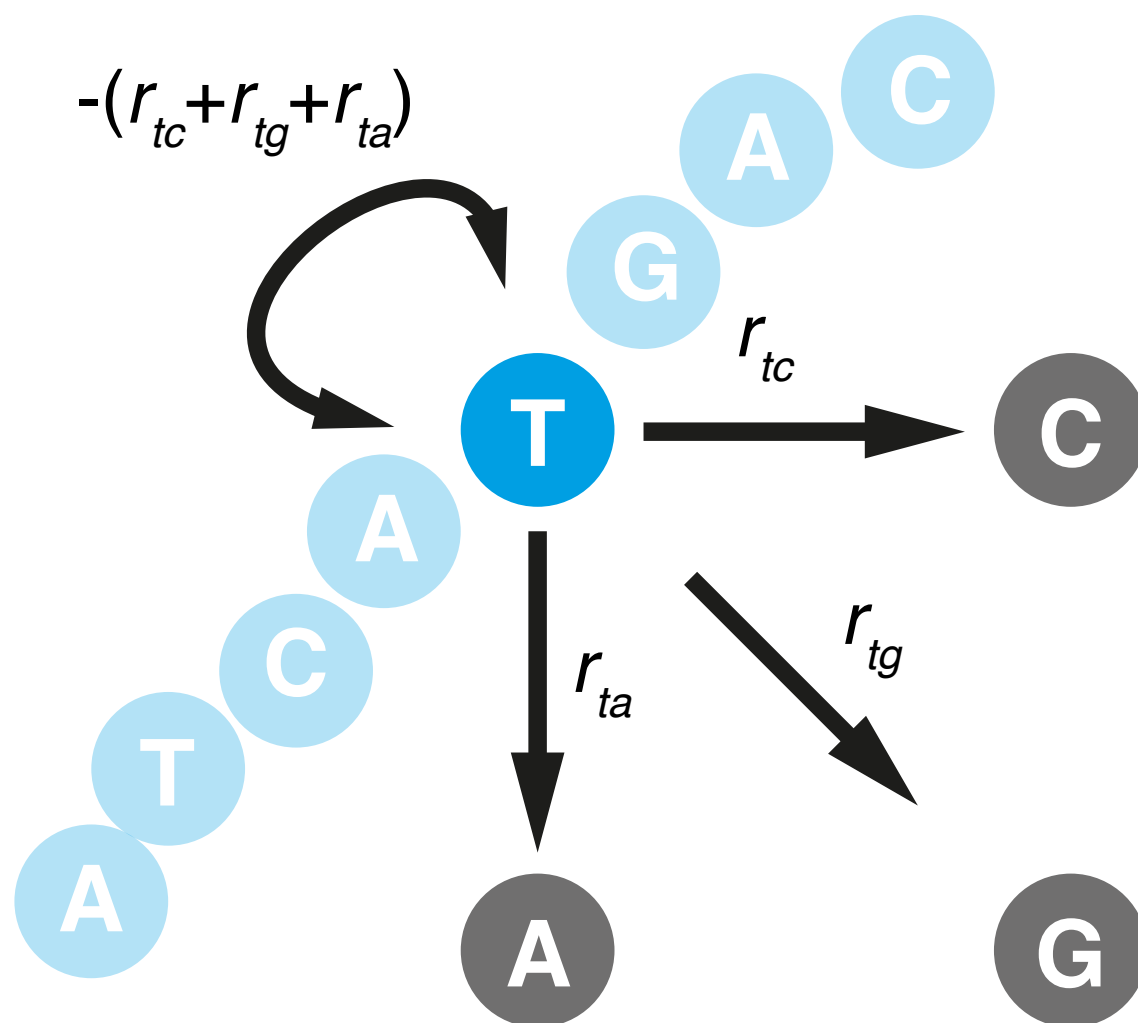
- Tree:
- Data:
- Model:



Taming the BEAST - bModelTest tutorial

Recap: Nucleotide substitution models

II. How are they parameterised?





Recap: Nucleotide substitution models

II. How are they parameterised?

In general: $Q =$

	A	C	G	T
A	$-(a + b + c)$	a	b	c
C	d	$-(d + e + f)$	e	f
G	g	h	$-(g + h + i)$	i
T	j	k	l	$-(j + k + l)$

bModelTest: only considers time reversible models:

$$Q = \begin{pmatrix} - & r_{ac}\pi_C & r_{ag}\pi_G & r_{at}\pi_T \\ r_{ac}\pi_A & - & r_{cg}\pi_G & r_{ct}\pi_T \\ r_{ag}\pi_A & r_{cg}\pi_C & - & r_{gt}\pi_T \\ r_{at}\pi_A & r_{ct}\pi_C & r_{gt}\pi_G & - \end{pmatrix} = \begin{pmatrix} - & r_{ac} & r_{ag} & r_{at} \\ r_{ac} & - & r_{cg} & r_{ct} \\ r_{ag} & r_{cg} & - & r_{gt} \\ r_{at} & r_{ct} & r_{gt} & - \end{pmatrix} \times \begin{pmatrix} \pi_A & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_G & 0 \\ 0 & 0 & 0 & \pi_T \end{pmatrix}$$



Taming the BEAST - bModelTest tutorial

bModelTest: only considers time reversible models:

$$Q = \begin{pmatrix} - & r_{ac}\pi_C & r_{ag}\pi_G & r_{at}\pi_T \\ r_{ac}\pi_A & - & r_{cg}\pi_G & r_{ct}\pi_T \\ r_{ag}\pi_A & r_{cg}\pi_C & - & r_{gt}\pi_T \\ r_{at}\pi_A & r_{ct}\pi_C & r_{gt}\pi_G & - \end{pmatrix} = \begin{pmatrix} - & r_{ac} & r_{ag} & r_{at} \\ r_{ac} & - & r_{cg} & r_{ct} \\ r_{ag} & r_{cg} & - & r_{gt} \\ r_{at} & r_{ct} & r_{gt} & - \end{pmatrix} \times \begin{pmatrix} \pi_A & 0 & 0 & 0 \\ 0 & \pi_C & 0 & 0 \\ 0 & 0 & \pi_G & 0 \\ 0 & 0 & 0 & \pi_T \end{pmatrix}$$

bModelTest: Parameterisation

rates are grouped into categories of same value:

for example:

r_{ac}	r_{ag}	r_{at}	r_{cg}	r_{ct}	r_{gt}	
1	1	1	1	1	1	← all rates the same
...						
1	2	3	4	5	6	← all rates different



Taming the BEAST - bModelTest tutorial

Named substitution models in BEAUti: Q=

JC69

$$\begin{pmatrix} - & \lambda & \lambda & \lambda \\ \lambda & - & \lambda & \lambda \\ \lambda & \lambda & - & \lambda \\ \lambda & \lambda & \lambda & - \end{pmatrix} \begin{matrix} r_{ac} & r_{ag} & r_{at} & r_{cg} & r_{ct} & r_{gt} \\ 1 & 1 & 1 & 1 & 1 & 1 \\ \pi_X = 0.25 \end{matrix}$$

HKY

$$\begin{pmatrix} - & \beta\pi_C & \alpha\pi_G & \beta\pi_T \\ \beta\pi_A & - & \beta\pi_G & \alpha\pi_T \\ \alpha\pi_A & \beta\pi_C & - & \beta\pi_T \\ \beta\pi_A & \alpha\pi_C & \beta\pi_G & - \end{pmatrix}$$

TN93

$$\begin{pmatrix} - & \beta\pi_C & \alpha_1\pi_G & \beta\pi_T \\ \beta\pi_A & - & \beta\pi_G & \alpha_2\pi_T \\ \alpha_1\pi_A & \beta\pi_C & - & \beta\pi_T \\ \beta\pi_A & \alpha_2\pi_C & \beta\pi_G & - \end{pmatrix}$$

GTR

$$\begin{pmatrix} - & r_{ac}\pi_C & r_{ag}\pi_G & r_{at}\pi_T \\ r_{ac}\pi_A & - & r_{cg}\pi_G & r_{ct}\pi_T \\ r_{ag}\pi_A & r_{cg}\pi_C & - & r_{gt}\pi_T \\ r_{at}\pi_A & r_{ct}\pi_C & r_{gt}\pi_G & - \end{pmatrix}$$

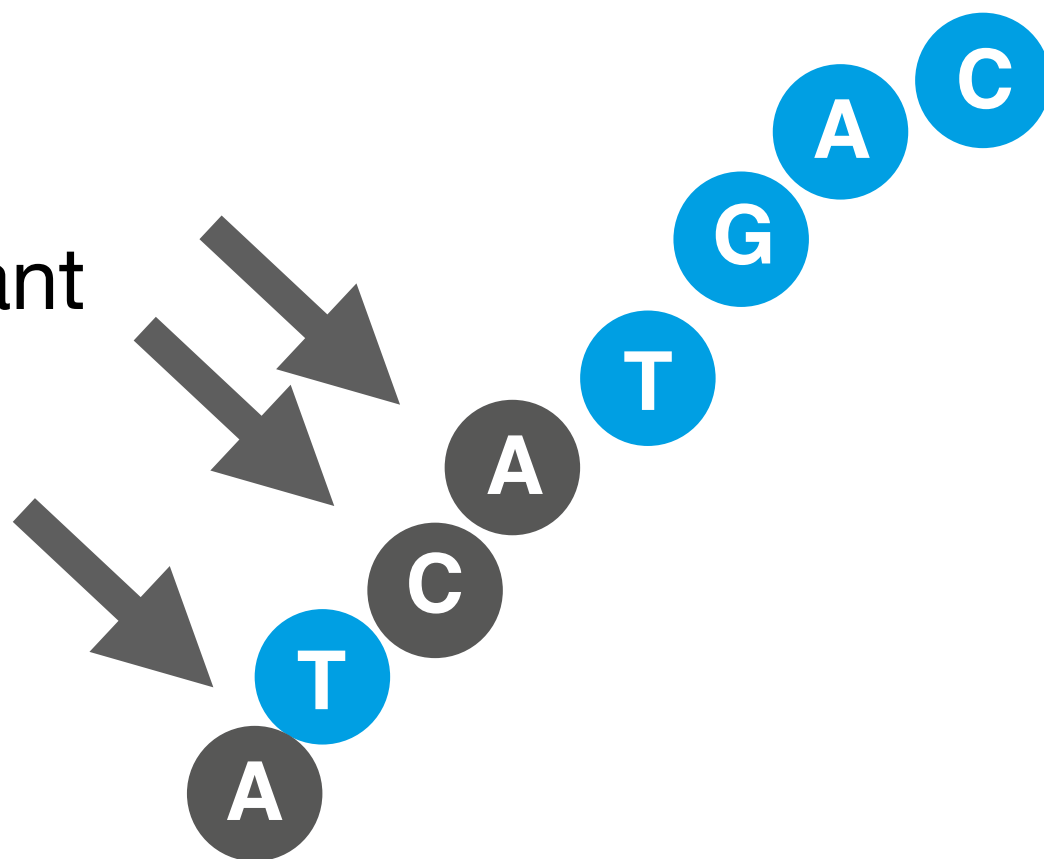
$$\begin{matrix} r_{ac} & r_{ag} & r_{at} & r_{cg} & r_{ct} & r_{gt} \\ 1 & 2 & 3 & 4 & 5 & 6 \\ \pi_X & \text{different} \end{matrix}$$



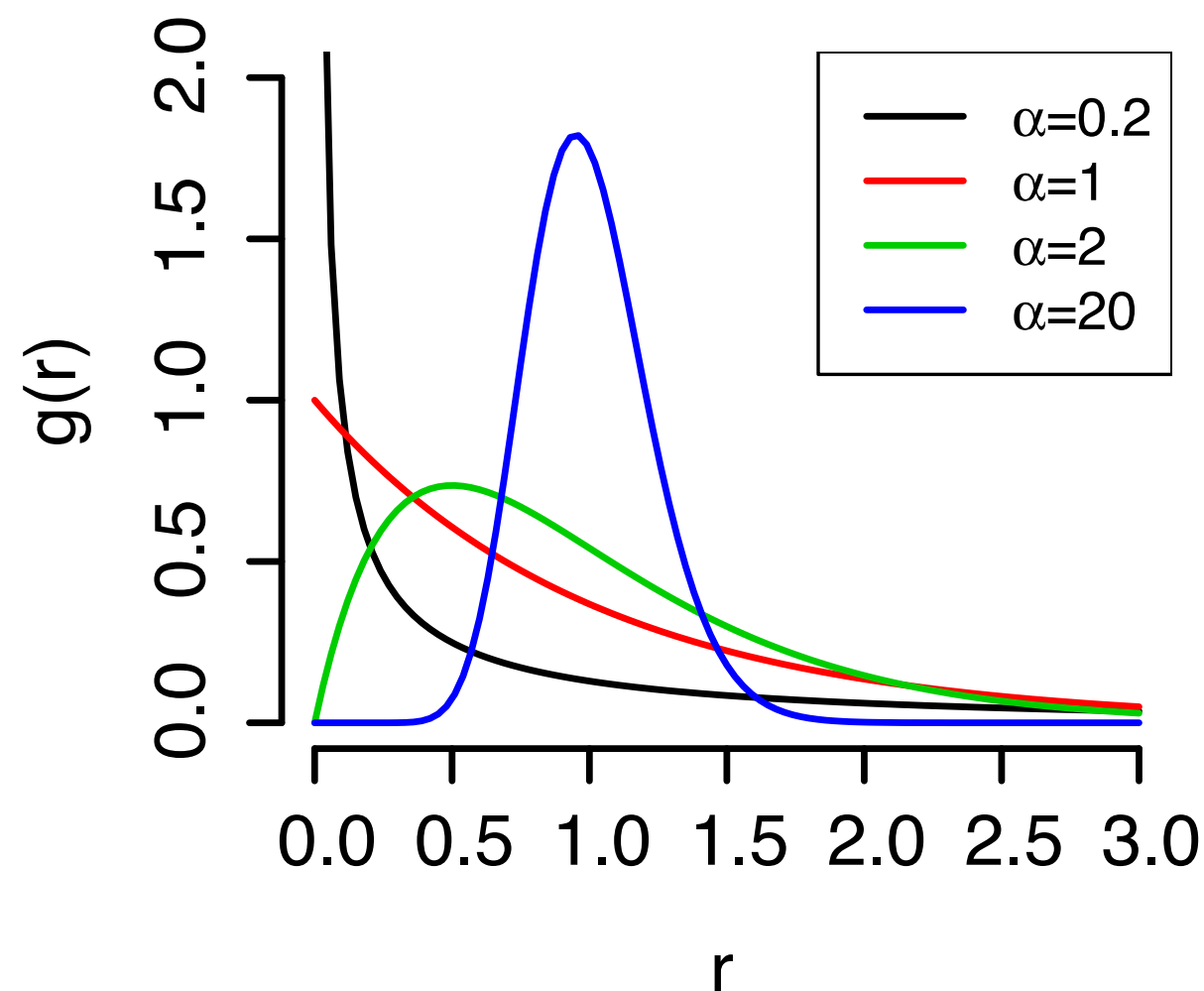
Taming the BEAST - bModelTest tutorial

III. Accounting for site variation

+I
invariant
sites



+ Γ
replace $r_{XY} \mapsto r_{XY}R$
where $R \sim \Gamma(\alpha)$



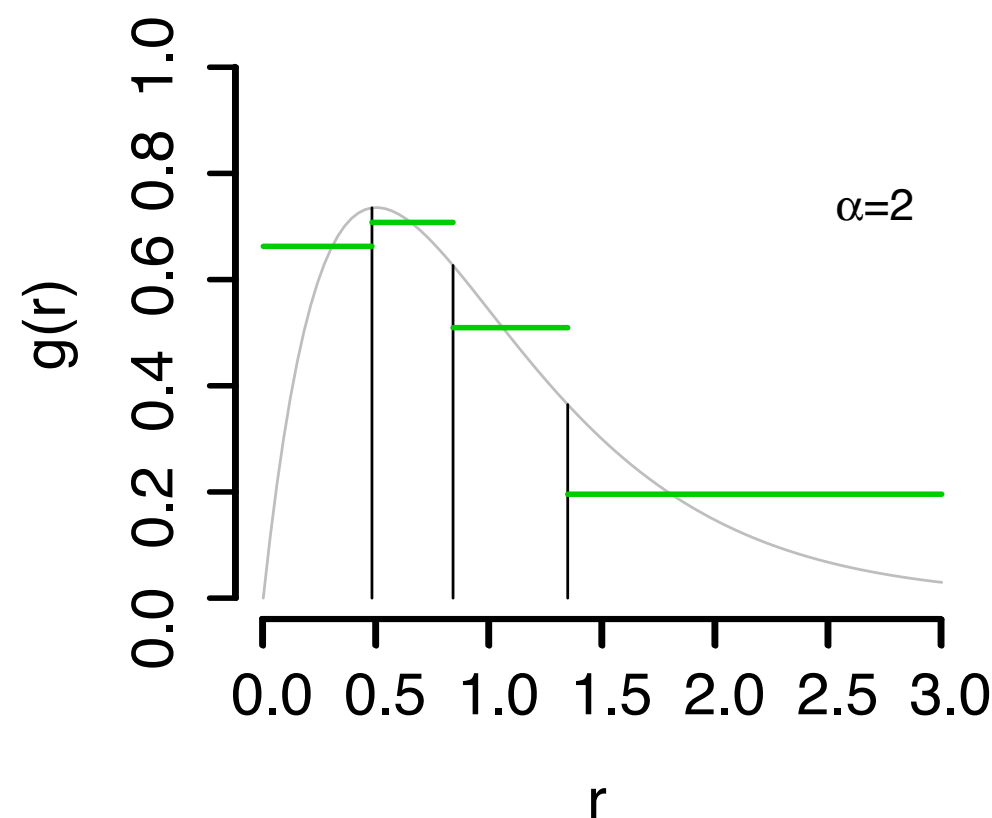


Taming the BEAST - bModelTest tutorial

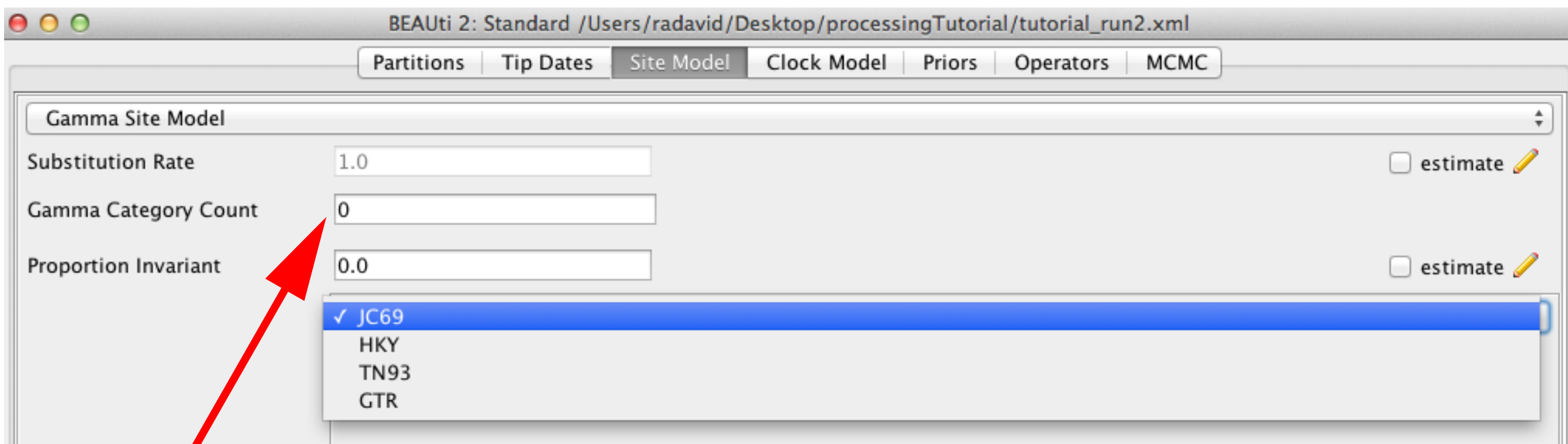
III. Accounting for site variation

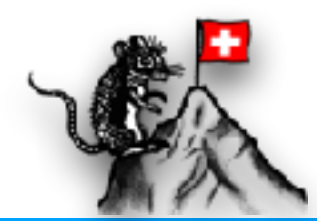
+ Γ

replace $r_{XY} \mapsto r_{XY}R$
where $R \sim \Gamma(\alpha)$



In BEAUti:





Taming the BEAST - bModelTest tutorial

<https://taming-the-beast.org/tutorials/Substitution-model-averaging/>