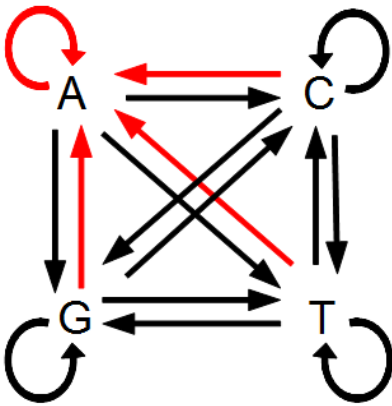


Setting DNA substitution models in BEAST

5 min read

2016/10/11



JC model of sequence evolution from [TreeThinkers](#).

BEAST (Bayesian Evolutionary Analysis Sampling Trees) software provides a growing set of *clear* options for specifying evolutionary models of DNA substitution for alignments loaded into the program. However, the set of models that are readily available and “spelled out” in drop-down menus in the BEAUti (Bayesian Evolutionary Analysis Utility) GUI is much smaller compared to the standard set of 11 substitution schemes considered by other software programs for model selection (giving rise to a total of 88 models in [jModelTest](#), or 56 models in [PartitionFinder](#)). This poses a problem that I have heard several BEAST users bring up in online discussion groups, and that several people have asked me about personally. In this post, I shed some light on this problem and clear up some of the confusion by

giving details on how users can easily specify different substitution models for their BEAST runs.

Background: Substitution models in BEAST

I will assume that readers are familiar with [DNA substitution models](#). Also, there are a number of details about how BEAST handles DNA substitution models, which you can learn from the [BEAST2 book](#) or the manual, but that I will not cover. For example, by default, BEAST normalizes the rate matrix so that the mean rate is 1.0, and so on. Moving on to the issues...

It seems that many people are having issues with the fact that it's not a simple "one-click" procedure to specify the diverse variety of substitution models that are frequently selected during model selection inference conducted prior to setting up BEAST runs. This is because [BEAST1](#) only gives the *obvious* options to specify three of the most popular models of DNA substitution: TrN, HKY, and GTR, while BEAST2 versions give users *obvious* options to specify one of four models: JC69, TrN, HKY, and GTR. Thus other models such as TIM, K80, and F81 can seem cryptic to some users, who will often exclaim in dismay, "Where is the option for model X???" or "I don't see how to specify model Y!!!" and then post a question about it (that may languish, never to even be directly answered, who knows).

Solving the problem: Learning BEAST functionalities "under your nose"

The real solution to the above problems is to realize the options

available to you, especially that BEAUti actually gives several important options in the Site Model panel that each user must learn to use to tailor the models of evolution for each alignment/partition. These options are located right “under your nose,” but I’ll show you where they are and how to use them in **Tables 1 and 2 below**, as follows:

Table 1. Details for setting substitution models in BEAST1 (e.g. v1.8.3).

Best-fit substitution model	(Base) Model to select in BEAUti	Additional specifications in BEAUti
JC69	JC69	None
TrN	TN93	None
TrNef	TN93	base Frequencies set to "All Equal"
K80 (K2P)	HKY	base Frequencies set to "All Equal"
F81	GTR	turn off operators for all rate parameters
HKY	HKY	None
SYM	GTR	base Frequencies set to "All Equal"
TIM	GTR	edit XML file by hand so that all other rates are equal to the AG rate
TVM	GTR	unchecking AG rate operator
TVMef	GTR	unchecking AG rate operator and setting base Frequencies to "All Equal"

GTR	GTR	None
-----	-----	------

Table 2. Details for setting substitution models in BEAST2 (e.g. 2.4.0+).

Best-fit substitution model	(Base) Model to select in BEAUti 2	Additional specifications in BEAUti 2
JC69	JC69	None
TrN	TN93	None
TrNef	TN93	base Frequencies set to "All Equal"
K80 (K2P)	HKY	base Frequencies set to "All Equal"
F81	GTR	fix all rate parameters to 1.0 (uncheck the "estimate" box)
HKY	HKY	None
SYM	GTR	base Frequencies set to "All Equal"
TIM	GTR	fix CT and AG rate parameters to 1.0 (uncheck the "estimate" box)
TVM	GTR	fix the AG rate parameter to 1.0 (uncheck the "estimate" box)
TVMef	GTR	fix the AG rate parameter to 1.0 (uncheck the "estimate" box), and also set base Frequencies to "All Equal"
GTR	GTR	None

I have quickly constructed these tables based on different options

available in BEAST versions, my knowledge of the substitution models, and the list of model code for different substitution models on the [BEAST1 website](#) (information for XML format for BEAST1 / earlier versions than present). How can you use the above tables to set a desired site model? Easy. You simply check **Tables 1 and 2** and for a given model you might want to set (**first column**), the tables tell you which model to select in BEAUti (**second column**) and the modifications, if any, you will need to make to that model or the XML file itself to get the model you want (**third column**).

Another way...

If you would like more detail, you can also draw on other resources available online to do this properly. For example, if you wanted to set the F81 model for a partition of your dataset, you could also search the [substitution model code page](#) mentioned above for "F81". If you did this you would find the following block for BEAST1:

```
<!-- *** SUBSTITUTION MODEL FOR PARTITION Gene2 *** -->
<!-- The F81 substitution model (Felsenstein, 1981) -->
<gtrModel id="F81_Gene2">
  <frequencies>
    <frequencyModel dataType="nucleotide">
      <frequencies>
        <parameter id="F81_Gene2.fre
      </frequencies>
    </frequencyModel>
```

```
</frequencies>
<rateAC>
    <parameter id="F81_Gene2.ac" value="1.0"/>
</rateAC>
<rateAG>
    <parameter id="F81_Gene2.ag" value="1.0"/>
</rateAG>
<rateAT>
    <parameter id="F81_Gene2.at" value="1.0"/>
</rateAT>
<rateCG>
    <parameter id="F81_Gene2.cg" value="1.0"/>
</rateCG>
<rateGT>
    <parameter id="F81_Gene2.gt" value="1.0"/>
</rateGT>
</gtrModel>
```

We can infer from this that the F81 model can be specified in any BEAST version by first specifying the GTR model, and then setting all rates to 1.0 and also setting all base frequencies to be equal “All Equal” (i.e. each equal to 0.25). And you would specify this information in the Site Model panel in BEAUti. You would also need to make sure that the priors were not included for these parameters, since they will be fixed; so you do that by commenting out the rate priors within the XML file.

Accordingly, you would also need to delete the rate parameters from the

log, because you won't need to log those parameters since you won't be estimating them.

~J

Additional Reading

****Substitution model basics from Wikipedia: [**substitution model, models of DNA evolution](#)**

Beast users google group threads:

- [BEAUti can set 40 models?](#)
- [F81 mutation model?](#)

BEAST2 website FAQ: [link](#)

PartitionFinder website: [PartitionFinder](#)