

# Models

Will DeWitt, Sarah Hilton, Andy Magee

## Goal and Housekeeping

Andy re-implemented the model from [Nasrallah and Huelsenbeck \(2013\)](#), which models pairwise epistasis due to RNA secondary structure. Below is the model as Andy implemented it in RevBayes. [\[Comments from Sarah look like this.\]](#) [\[Comments from Will look like this.\]](#) [\[Comments from Andy look like this.\]](#)

## RNA epistasis model

Here is the model from [Nasrallah and Huelsenbeck \(2013\)](#). [\[I am just copying over what is found in the paper. We should change this around in an order we like and with notation we like.\]](#)

$Q$  is the instantaneous rate matrix describing changes from doublet  $x$  to doublet  $y$ . For  $x = (x_1, x_2)$ ,  $x_1$  is the 5' nucleotide and  $x_2$  is the 3' nucleotide.

$$Q = \begin{cases} \xi \pi_y S_{x_1, y_1} & \text{if single substitution at the 5' site,} \\ \xi \pi_y S_{x_2, y_2} & \text{if single substitution at the 3' site,} \\ \xi \pi_y S_{x_1, y_1} S_{x_2, y_2} d & \text{if double substitution where } x \text{ and } y \in W, \\ 0 & \text{if any other double substitution,} \\ -\sum_{y \neq x} q_{xy} & \text{if } x = y \end{cases} \quad (1)$$

$S$  is the GTR exchangeability matrix ([Tavaré, 1986](#)),  $W = AT, CG, GC, TA$  is the set of Watson-Crick pairs,  $\pi = (\pi_{AA}, \pi_{AC}, \dots, \pi_{TT})$  are the stationary state frequencies of the 16 possible doublet states,  $d$  is the relative rate of double to single mutations between doublets, and  $\xi$  is the rate-scaling factor.

## Points to clarify

### How do we interpret $d$ ?

$d$  is the relative rate of double to single mutations between doublets or the "strength" of epistatic interactions.

Here is a copy of table 1 from [Nasrallah and Huelsenbeck \(2013\)](#):

	$\pi_y = \pi_{y_1}\pi_{y_2}$	$\pi_y \neq \pi_{y_1}\pi_{y_2}$
$d = 0$	Independent and nonepistatic	Model inadequacy
$d > 0$	Dependent but nonepistatic	Dependent and epistatic

[We talked about this last time but I want to make sure I totally understand. When we are simulating with  $d = 0$ , we are actually in the "model inadequacy" quadrant?]

### How do we normalize $Q(\xi)$ ?

[From slack:  $0.5 * \text{Pr}(\text{single}) + \text{Pr}(\text{double})$  ]

### Given the fit parameter values for a GTR model, how do we simulate under this model?

[From github, " This is the tree inferred under GTR+GAMMA using RAxML version 8.2.12. Inferred model parameters

alpha shape parameter = 0.440894 relative exchange rates (ac ag at cg ct gt) = 1.882161 7.009179 0.914813 0.495852 7.666181 1.000000 base frequencies = 0.340152 0.190828 0.225045 0.243974"]

## References

- Nasrallah CA, Huelsenbeck JP. 2013. A phylogenetic model for the detection of epistatic interactions. *Molecular Biology and Evolution*. 30:2197–2208.
- Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences*. 17:57–86.