Computational Biology HS19 – Exercise 2 answers to theory questions

21th October 2019

1. When the overall rate of change is extremely low compared to the time scale of the tree, we would expect substitution events to be rare. The distribution of nucleotides for each of the 10 species would be very similar to that of the ancestral sequence. In other words, the sequence would be quite conserved between the 10 species, with more Ts and Cs than As and Gs, and the ratio between Ts and Cs would not deviate far from 1:1.

2. When the overall rate of change is extremely high compared to the time scale of the tree, we would expect multiple substitution events to be quite likely, i.e. the ancestral sequence will be prone to gaining sequence changes, and the evolution will happen on a shorter timescale. The distribution of nucleotides for each of the 10 species would be divergent from the ancestral sequence.

3. The minimal amount of time for the probability transition matrix P to be numerically indistinguishable from the stationary distribution $\pi = (\pi_T, \pi_C, \pi_A, \pi_G) = (0.22, 0.26, 0.33, 0.19)$ is 600 mya.

4. The overall rate of change for each nucleotide, i.e. the number of any substitution events happening at each instantaneous moment, is $-q_{ii}$. Hence the time of the next substitution event would be the inverse of the overall rate of change, given by $\frac{1}{q_{ii}}$.

5. For each row in the substitution matrix Q, we could normalize the rates of each possible substitution event against the sum of that for all substitution events. This would give us the probability of how likely each substitution event happens. We can use this to sample the nucleotide that the original one is substituted by.

(Cont’d)

For example if the first row of Q_{TN93} looks like this:

	T	C	A	G
T	-0.95	0.6	0.25	0.1

Probability of T substituted by C $= \frac{0.6}{0.6+0.25+0.1} = \frac{0.6}{0.95}$

Probability of T substituted by A $= \frac{0.25}{0.95}$

Probability of T substituted by G $= \frac{0.1}{0.95}$

This gives us a probability mass function we can sample the next nucleotide with.