

Real change, observed change, and expected change

In this exercise you will perform an analog simulation (!) of the evolution of a 12-bp long piece of DNA. It is the aim to show you that (1) not all substitution events are detected when directly comparing sequences (the problem of multiple or superimposed substitutions), (2) we can compute how much observable change we would expect if a certain amount of actual change takes place, (3) we can also use the formula from (2) in reverse, and calculate how much actual change there probably was based on an observed difference.

Simulation

First: get a 6-sided die (real or simulated).

You will now simulate the evolution of a short stretch of DNA (12 bp) that is evolving along a branch in a phylogenetic tree according to the Jukes and Cantor model (all substitutions are equally likely). Let us assume that the branch corresponds to, say, 100 million years of evolution, and that during this timespan 8 substitution events will occur in our 12-bp piece of DNA. This is the same as saying that the branch length is 8 substitutions per 12 sites, or 0.67 substitutions per site. Use the following equation to compute how much observable change you would expect given that the actual distance is 0.67 substitutions per site:

$$d_{\text{obs}} = 0.75 * (1 - \exp(-(4/3) * d_{\text{real}}))$$

Expected observable distance: _____ substitutions per site

On the second page of this handout you will find the DNA sequence (which initially consists of 12 As) along with a table showing how to interpret various die rolls. You should now perform the simulation according to the steps below, and we will then examine the pooled results from the entire class.

(1) Determine site of substitution:

Roll the die two times. The first roll selects one particular 6-bp stretch of the sequence (1-3 selects the first half, 4-6 selects the second half). The second roll selects one specific site within the chosen stretch of 6 bp. This selection scheme is also indicated schematically above the sequence. Make sure to note which site you selected.

(2) Determine new nucleotide at site:

Roll the die again and use the table below to select the new nucleotide. In each case the interpretation of the roll depends on the nucleotide that is present at the site initially (e.g., if the original nucleotide is A and you roll a 3, then the new nucleotide should be G). Write the new nucleotide in the proper place below the sequence.

(3) Repeat 8 times:

You should repeat the above steps 8 times, corresponding to 8 substitution events. Occasionally a site will experience multiple substitutions.

(4) Report results

When you are done, you should count the number of observable substitutions (the number of non-A sites in your sequence) and use that to compute the observed distance per site (divide by 12):

Number of observable substitutions: _____
Observed distance: _____ substitutions per site

Report this number to the teacher.

Real change, observed change, and expected change

First die roll: 1-3 | 4-6
 Second die roll: 1 2 3 4 5 6 | 1 2 3 4 5 6
 Sequence: **A A A A A A A A A A A A**

Mutation table:

Initial nucleotide	die roll		
	1-2	3-4	5-6
A	C	G	T
C	A	G	T
G	A	C	T
T	A	C	G