

Supplementary protocols

Calculation of mutation expectation values

Single base mutations can be grouped into six different types of reactions (two types of transitions and four types of transversions). Transitional point mutations exchange one purine with another purine and a pyrimidine with another pyrimidine on the complementary DNA strand. In addition, four different types of transversional point mutations can be defined. All possible substitutions are shown in Table 1. Note that the orientation of a base pair is irrelevant.

Ti1	Ti2	Tv1	Tv2	Tv3	Tv4
$\begin{array}{c} \text{A} \\ \vdots \\ \text{T} \end{array} \rightarrow \begin{array}{c} \text{G} \\ \vdots \\ \text{C} \end{array}$	$\begin{array}{c} \text{C} \\ \vdots \\ \text{G} \end{array} \rightarrow \begin{array}{c} \text{T} \\ \vdots \\ \text{A} \end{array}$	$\begin{array}{c} \text{A} \\ \vdots \\ \text{T} \end{array} \rightarrow \begin{array}{c} \text{T} \\ \vdots \\ \text{A} \end{array}$	$\begin{array}{c} \text{A} \\ \vdots \\ \text{T} \end{array} \rightarrow \begin{array}{c} \text{C} \\ \vdots \\ \text{G} \end{array}$	$\begin{array}{c} \text{C} \\ \vdots \\ \text{G} \end{array} \rightarrow \begin{array}{c} \text{G} \\ \vdots \\ \text{C} \end{array}$	$\begin{array}{c} \text{C} \\ \vdots \\ \text{G} \end{array} \rightarrow \begin{array}{c} \text{A} \\ \vdots \\ \text{T} \end{array}$

Table 1: Types of single base pair substitutions. „Ti1“ and „Ti2“ correspond to transitional point mutations, whereas „Tv1“ – „Tv4“ correspond to different types of transversional point mutations.

To assign the probability for each of these six reactions, we counted the number of each of these substitutions in the whole sequence set after seven rounds of selection. These numbers were then normalized by dividing by the total number of A and T, or C and G, respectively.

The calculated probabilities were

$$p(\text{Ti1}) = p_{\text{AT} \rightarrow \text{GC}} = \frac{\sum \text{AT} \rightarrow \text{GC}}{\sum \text{A} + \sum \text{T}} = 2.96\%$$

$$p(\text{Ti2}) = p_{\text{CG} \rightarrow \text{TA}} = \frac{\sum \text{CG} \rightarrow \text{TA}}{\sum \text{C} + \sum \text{G}} = 0.96\%$$

$$p(\text{Tv1}) = p_{\text{AT} \rightarrow \text{TA}} = \frac{\sum \text{AT} \rightarrow \text{TA}}{\sum \text{A} + \sum \text{T}} = 0.08\%$$

$$p(\text{Tv2}) = p_{\text{AT} \rightarrow \text{CG}} = \frac{\sum \text{AT} \rightarrow \text{CG}}{\sum \text{A} + \sum \text{T}} = 0.16\%$$

$$p(\text{Tv3}) = p_{\text{CG} \rightarrow \text{GC}} = \frac{\sum \text{CG} \rightarrow \text{GC}}{\sum \text{C} + \sum \text{G}} = 0.04\%$$

$$p(\text{Tv4}) = p_{\text{CG} \rightarrow \text{AT}} = \frac{\sum \text{CG} \rightarrow \text{AT}}{\sum \text{C} + \sum \text{G}} = 0.07\%$$

Expectation values for an amino acid mutation ($p_{\text{mut,calc}}$) were calculated for each codon of the genetic code by adding up all substitution type probabilities that are possible for a given codon and – as a second requirement – will lead to a change of the encoded amino acid, exemplified in Table 2 for three different codons. The resulting expectation values for each codon are shown in Table 3.

		Ti1	Ti2	Tv1	Tv2	Tv3	Tv4	
Phe	T	X		X	X			$p_{\text{mut,calc,TTT}} = 2 \times p(\text{Ti1}) + 3 \times p(\text{Tv1}) + 3 \times p(\text{Tv2})$
	T	X		X	X			
	T	O		X	X			
Phe	T	X		X	X			$p_{\text{mut,calc,TTC}} = 2 \times p(\text{Ti1}) + 2 \times p(\text{Tv1}) + 2 \times p(\text{Tv2}) + p(\text{Tv3}) + p(\text{Tv4})$
	T	X		X	X			
	C		O			X	X	
Leu	T	O		X	X			$p_{\text{mut,calc,TTA}} = p(\text{Ti1}) + 3 \times p(\text{Tv1}) + 3 \times p(\text{Tv2})$
	T	X		X	X			
	A	O		X	X			

Table 2: Calculation of expectation values ($p_{\text{mut,calc}}$). Possible substitutions leading to a change of the encoding amino acid are marked by “X”, whereas substitutions that do not result in an amino acid change are marked by “O”. The observed probabilities for each substitution type are added to result in mutation expectation values for each codon.

	T	C	A	G	
T	6.64%	4.27%	6.64%	4.51%	T
	6.51%	4.27%	6.51%	4.38%	C
	3.69%	4.27%	stop	stop	A
	3.39%	4.27%	stop	5.34%	G
C	4.27%	2.14%	4.51%	2.14%	T
	4.27%	2.14%	4.38%	2.14%	C
	3.31%	2.14%	4.51%	2.07%	A
	3.31%	2.14%	4.38%	2.07%	G
A	6.48%	4.27%	6.64%	4.51%	T
	6.44%	4.27%	6.51%	4.51%	C
	9.36%	4.27%	6.64%	4.35%	A
	7.47%	4.27%	6.51%	4.22%	G
G	2.28%	2.14%	4.51%	2.14%	T
	2.28%	2.14%	4.38%	2.14%	C
	2.28%	2.14%	4.51%	2.14%	A
	2.28%	2.14%	4.38%	2.14%	G

Table 3: Expectation values for mutating each codon so that mutations result in an exchange of the encoded amino acid.