K_a/K_s Ratios: What do they tell us?

I. Definitions

K_a is the number of non-synonymous substitutions per non-synonymous site per time period.

K_s is the number ofsynonymous substitutions per synonymous site per time period.

II. The maximum K_a/K_s ratio in the absence of positive selection is 1.0

Let α be the proportion of non-synonymous mutations that are neutral, and let β be the proportion of non-synonymous mutations that are advantageous. In the absence of positive selection, $\beta = 0$ by definition.

Let μ be the overall mutation rate for disadvantageous, neutral, and advantageous mutations combined. Then the rate of neutral non-synonomous mutations is

$$\mu_{ns} = \alpha\mu$$

and the rate of neutral synonymous mutations (all synonymous mutations are assumed neutral) is

$$\mu_s = \mu$$

But the rate of fixation of neutral mutations is just equal to the mutation rate, so

$$\frac{K_a}{K_s} = \frac{\mu_{ns}}{\mu_s} = \alpha \ .$$

Since 0 $\leq \alpha \leq 1$, it follows that $\frac{K_a}{K_s} \leq 1$. In other words, if all non-synonymous mutations are either neutral or deleterious, this condition must hold.

II. With positive selection, K_a/K_s can be > 1

The occurrance of positive selection means $\beta > 0$. To calculate the rate of fixation of non-synonymous mutations, we must first calculate the number of neutral and the number of advantageous non-synonymous mutations that arise per generation:

neutral mutations/gen $= 2N\mu\alpha$

advantageous mutations/gen $= 2N\mu\beta$

Next, we must determine the probability of fixation of neutral and advantageous non-synonymous mutations. For neutral mutations, this probability is

$$P_{\text{neutral}} = \frac{1}{2N} \quad \text{(why?).}$$

For advantageous mutations, this probability was shown by Fisher and Wright to be

$$P_{\text{advantageous}} = \frac{s}{1-e^{-2Ns}} \quad ,$$

where N is population size and s is the selection coefficient. When Ns is $\gg 1$, then

$$P_{\text{advantageous}} \approx s \; .$$

Finally, the number of non-synonymous mutations fixed per generation is

$$K_a = \text{(\# neutral mutations arising/gen)} \times \text{(prob. of fixation)}$$

$$+ \text{(\# advantageous mutations arising/gen)} \times \text{(prob. of fixation)}$$

$$= 2N\mu\alpha \frac{1}{2N} \; + \; 2N\mu\beta s \; = \; \mu\alpha \; + 2N\mu\beta s \; .$$

For synonymous mutations, which are presumed to be neutral, we have as before

$$K_s = \mu_s = \mu \; .$$

Thus,

$$\frac{K_a}{K_s} = \frac{\mu\alpha + 2N\mu\beta s}{\mu} = \alpha + 2N\beta s \; .$$

If $2N\beta s$ is large enough, then $\frac{K_a}{K_s}$ can be > 1.

Example: Let $\alpha = 0.3$, $\beta = 0.1$, $s = 0.01$, $N = 1000$.

Then $\frac{K_a}{K_s} = 0.3 + 2\,(1000)\,(0.01)\,(0.1) = 0.3 + 2 = 2.3$, which is > 1.

Conclusion: If $\frac{K_a}{K_s} > 1$, positive selection occurred.

III. A value of $\frac{K_a}{K_s} < 1$ does *not* mean positive selection is not occurring

Example: Let $\alpha = 0.3$, $\beta = 0.01$, $s = 0.01$, $N = 1000$. Since $\beta \neq 0$, positive selection occurs.

Then $\frac{K_a}{K_s} = 0.3 + 2\ (1000)\ (0.01)\ (0.01) = 0.3 + 0.2 = 0.5$, which is <1.

IV. If $\frac{K_a}{K_s} < 1$, there is constraint (i.e. $\alpha + \beta < 1$)

Let $\gamma = 1 - (\alpha + \beta)$ be the proportion of mutations that are deleterious and thus eliminated by purifying selection. If $\gamma > 0$, we say there is selective constraint, *i.e.* some mutations are eliminated by purifying selection.

We know that $\frac{K_a}{K_s} = \alpha + 2N\beta s$. The second term ($2N\beta s$) represents fixations of positively selected alleles. The condition $\frac{K_a}{K_s} < 1$ means $\alpha + 2N\beta s < 1$. But we also know that for alleles to be positively selected, it must be the case that $2Ns > 1$. This then implies that

$\alpha + \beta < 1$, which in turn implies that $1 - (\alpha + \beta) = \gamma > 0$.

V. Calculating $\frac{K_a}{K_s}$ from sequence data.

Suppose you have the following 12-bp sequence, with the properties indicated. All 1st and 2nd codon positions are non-degenerate, which means that they are classified as non-synonymous sites. The same is true for the 3rd position of Met. For the other amino acids, however, the 3rd position is degenerate. With Val and Thr, this degeneracy is four-fold, which means that any substitution is synonymous. These sites are thus classified as fully synonymous. The 3rd position of Arg is two-fold degenerate, which means that mutation to A is synonymous, whereas a mutation to T or C is non-synonymous, producing the amino acid Ser. Consequently, this site is scored as $\frac{2}{3}$ non-synonymnous and $\frac{1}{3}$ synonymous.

	Val	Met	Arg	Thr	
	G T T	A T G	A A G	A C C	Total
degeneracy	(4)		(2)	(4)	
# non-syn sites	1 1 0	1 1 1	1 1 $\frac{2}{3}$	1 1 0	$9\frac{2}{3}$
# syn sites	0 0 1	0 0 0	0 0 $\frac{1}{3}$	0 0 1	$2\frac{1}{3}$

Adding across nucleotide sites gives $9\frac{2}{3}$ non-synonymous sites in this peptide and $2\frac{1}{3}$ synonymous sites.

Next, let us assume that this peptide evolves by having three amino-acid substitutions to produce the following peptide:

	Val	Leu	Arg	Thr	
	G T A	C T G	A A A	A C C	Total
# substitutions					
non-syn	0 0 0	1 0 0	0 0 0	0 0 0	1
synonymous	0 0 1	0 0 0	0 0 1	0 0 0	2

The total number of synonymous and non-synonymous substitutions are indicated in the last column.

Now, $K_a = \frac{\text{\# non-syn substitutions}}{\text{\# non-syn sites}} = \frac{1}{9\frac{2}{3}} = 0.103$ and

$$K_s = \frac{\text{\# synon. substitutions}}{\text{\# synon. sites}} = \frac{2}{2\frac{1}{3}} = 0.857.$$

Thus,

$$\frac{K_a}{K_s} = \frac{0.103}{0.857} = 0.12 \ .$$