

7. Sharp, P. M. *J. Mol. Evol.* **33**, 23–33 (1991).
 8. Akashi, H. *Curr. Opin. Gen. Dev.* **11**, 660–666 (2001).
 9. Goldman, N. & Yang, Z. *Mol. Biol. Evol.* **11**, 725–736 (1994).
 Reply: J. B. Plotkin, J. Dushoff and H. B. Fraser reply to this communication (doi:10.1038/nature03224).

Evolutionary genomics

Detecting selection needs comparative data

Positive selection at the molecular level is usually indicated by an increase in the ratio of non-synonymous to synonymous substitutions (dN/dS) in comparative data. However, Plotkin *et al.*¹ describe a new method for detecting positive selection based on a single nucleotide sequence. We show here that this method is particularly sensitive to assumptions regarding the underlying mutational processes and does not provide a reliable way to identify positive selection.

Plotkin *et al.*¹ use a measure for detecting selection known as the volatility index, whereby a codon with high volatility is more likely to have arisen by a non-synonymous mutation than a codon with low volatility; so, for high dN/dS , there should be more codons of high volatility. Positive selection should be detectable simply by examining the volatility index in a single sequence.

However, this argument is flawed because high rates of non-synonymous mutation will increase the rate of substitution both into and out of codons with high volatility. In models in which the substitution process is reversible over time, these two factors will cancel each other out, and variations in the strength of selection at the amino-acid level do not affect the expected volatility. Although most models used in studies of molecular evolution are time-reversible², the true substitution process probably is not, because of the specifics of the mutational and population-level processes.

To examine the effect of the substitution model on the volatility index, we simulated random-substitution models in which the rate of substitution between different nucleotides was sampled from a uniform random variable between zero and one. For these models, we then calculated the equilibrium frequencies of the 61 sense codons in a Markov chain model that resulted from simulations having varying synonymous and non-synonymous substitution rates. Based on the equilibrium frequencies, we could then calculate the expected value of the volatility index.

Our results indicate that the volatility index can be either an increasing or a decreasing function of dN/dS , or have a minimum or maximum at an intermediate value of this ratio (Fig. 1). We also find that the dN/dS ratio only marginally affects the volatility index — particularly for values of

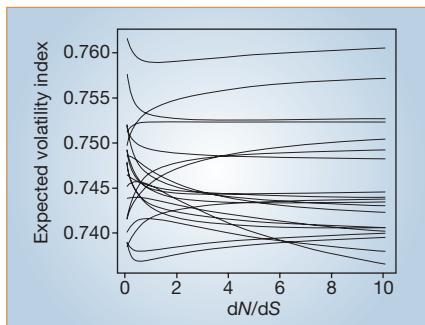


Figure 1 Expected value of the volatility index, as defined by Plotkin *et al.*¹, as a function of the dN/dS ratio for 20 random substitution models.

$dN/dS > 1$. Although models can be constructed in which strong stabilizing selection on particular amino acids has a marked effect on the volatility index, there is no evidence that the volatility index captures much information regarding positive selection. Realistic models of positive selection will predict an increased rate of substitution both in and out of codons with high volatility.

What then explains the results of Plotkin *et al.*¹, in which the volatility index correlates with the rate of amino-acid substitution in comparative data and with the amount of expression? Non-random codon usage is common in most organisms, particularly in bacteria and yeast^{3–7}, and may be caused by selection for optimal codon usage and affected by variation in the nucleotide composition and other factors. In bacteria, the strength of codon-usage bias is correlated with the amount of expression^{3–5} and with the extent of amino-acid substitution^{6,7}; this may be because highly expressed genes tend to be more conserved at the amino-acid level and have more codon-usage bias than genes with low expression. The degree of amino-acid substitution might also correlate with local nucleotide frequencies because regions that differ in this respect could have different rates and patterns of mutation.

To investigate the extent to which the volatility index is sensitive to local nucleotide content, we took advantage of the fact that only codons with sixfold degeneracy or with stop codons as neighbours can contribute to the volatility index. Using all other codons we obtained independent estimates of the nucleotide frequencies. We also calculated a *P* value for a one-tailed test of increase in the frequency of a particular nucleotide by using the methodology of Plotkin *et al.*¹, but calculated only for codons that do not contribute to the volatility index.

Applying this approach to the *Plasmodium falciparum* data analysed by Plotkin *et al.*¹, the correlation coefficient between the log *P* value of the volatility index and the log *P* value associated with the percentage of thymine is 0.29. Variation in third-position nucleotide content is one of the factors explaining the distribution of volatility

index-related *P* values in *P. falciparum*. Correlation of the volatility index with the amount of amino-acid substitution could be caused by the presence of covariates such as nucleotide frequencies, selection for optimal codon usage bias and/or expression levels.

The results of Plotkin *et al.*¹ might also be explained by variation in the amino-acid frequencies among genes. If the true evolutionary model is not time-reversible, these frequencies should influence codon usage and the volatility *P* value. Indeed, many of the amino-acid frequencies show correlation with the volatility *P* values calculated by Plotkin *et al.*¹. For example, the correlation coefficient between the frequency of glutamine and the log volatility *P* values is -0.32 . All codons for glutamine have the same volatility, but this amino acid is one mutational step away from arginine and leucine, which both affect the volatility index. The volatility index in models that are not time-reversible can therefore be affected by stabilizing selection on particular amino acids, because such selection affects the amino-acid frequency. But whether the volatility index correlates positively or negatively with such selection depends on which amino acid is the target of selection. Positive selection that increases the rate of amino-acid substitution does not have the same impact on the volatility index.

We argue that the volatility index cannot be applied to detect positive selection as it is under greater influence from other factors, such as amino-acid and nucleotide frequencies. However, the results of Plotkin *et al.*¹ should spur efforts to identify the causes of non-random codon usage in bacteria and other organisms.

Rasmus Nielsen*†, Melissa J. Hubisz‡

*Centre for Bioinformatics, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
 e-mail: rasmus@binf.ku.dk

†Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, USA
 doi:10.1038/nature03222

1. Plotkin, J. B., Dushoff, J. & Fraser, H. B. *Nature* **428**, 942–945 (2004).
2. Lio, P. & Goldman, N. *Genome Res.* **8**, 1233–1244 (1998).
3. Ikemura, T. *J. Mol. Biol.* **151**, 389–409 (1981).
4. Grantham, R., Gautier, C., Gouy, M., Jacobson, M. & Mercier, R. *Nucl. Acids Res.* **9**, 43–74 (1981).
5. Grosjean, H. & Fiers, W. *Gene* **18**, 199–209 (1982).
6. Sharp, P. M. *J. Mol. Evol.* **33**, 23–33 (1991).
7. Akashi, H. & Gojobori, T. *Proc. Natl. Acad. Sci. USA* **99**, 3695–3670 (2002).

Reply: J. B. Plotkin, J. Dushoff and H. B. Fraser reply to this communication (doi:10.1038/nature03224).

Evolutionary genomics

Codon volatility does not detect selection

Plotkin *et al.*¹ introduce a method to detect selection that is based on an index called codon volatility and that