

Note

On the Evolution of Codon Volatility

Jianzhi Zhang¹

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109

Manuscript received August 11, 2004
Accepted for publication September 30, 2004

ABSTRACT

Volatility of a codon is defined as the probability that a random point mutation in the codon generates a nonsynonymous change. It has been proposed that higher-than-expected mean codon volatility of a gene indicates that positive selection for nonsynonymous changes has acted on the gene in the recent past. I show that strong frequency-dependent selection (minority advantage) in large populations can increase codon volatility slightly, whereas directional positive selection has no effect on volatility. Factors unrelated to positive selection, such as expression-related or GC-content-related codon usage bias, also affect volatility. These and other considerations suggest that codon volatility has only limited utility for detecting positive selection at the DNA sequence level.

POSITIVE Darwinian selection at the DNA sequence level is usually assessed by comparing the number of synonymous substitutions per synonymous site (d_s) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) from an alignment of at least two homologous protein-coding DNA sequences (LI 1997; NEI and KUMAR 2000). A significantly higher d_N over d_s is a strong indication of positive selection (HUGHES and NEI 1988). Plotkin and co-workers recently introduced a new method for detecting positive selection using just one DNA sequence, and this method is based on the concept of codon volatility (PLOTKIN *et al.* 2004). Volatility of a codon is defined as the probability that a random point mutation in this codon is nonsynonymous. Codon volatility necessarily depends on mutational patterns, such as the ratio of transitional to transversional changes. In the simplest mutation model (SMM) where all nucleotides have equal mutation rates and all nucleotides are equally exchangeable, the volatility of a codon is the proportion of point-mutation neighbors that encode different amino acids (PLOTKIN *et al.* 2004). For instance, TTG (Leu) has a volatility of 6/8, because 6 of its 8 non-stop-codon neighbors are nonsynonymous. Table 1 lists the volatility of all 61 sense codons under the SMM. One can see that codon volatility varies from 0.5 (CGA for Arg) to 1 (TGG for Trp or ATG for Met). The volatility of a gene is the average volatility for all the codons in the gene. A gene may have an exception-

ally high volatility simply because it has a high proportion of amino acids with high volatility (*e.g.*, Trp and Arg). Controlled for amino acid composition, a gene may still have an unusually high volatility due to the frequent use of synonymous codons that are of high volatility. Four of the 20 amino acids (Arg, Gly, Leu, and Ser) contain synonymous codons of different volatilities (Table 1). PLOTKIN *et al.* (2004) argued that high volatility (after control for amino acid composition and genomic average patterns of synonymous codon usage) is a result of positive selection for nonsynonymous changes and that volatility can be used as a statistic for detecting positive selection. In particular, it was proposed that compared with a low-volatility codon, the presence of a high-volatility codon in a gene sequence indicates a greater probability that the previous substitution was nonsynonymous (PLOTKIN *et al.* 2004). Although this interpretation is valid, it does not follow that high volatility indicates positive selection. Furthermore, in theory, positive selection for nonsynonymous changes may not increase codon volatility for the following reason: The only advantage of having a high-volatility codon is its great potential of generating nonsynonymous mutations in the future. But natural selection cannot foresee the future. As long as two codons code for the same amino acid, the high-volatility codon does not confer higher fitness. Therefore, it is unclear under what conditions codon volatility would indicate positive selection. Here I address this question by computer simulation and analysis of genomic sequence data. My results suggest that positive selection rarely increases codon volatility and that factors other than positive selection can affect co-

¹Address for correspondence: Department of Ecology and Evolutionary Biology, University of Michigan, 3003 Natural Science Bldg., 830 N. University Ave., Ann Arbor, MI 48109. E-mail: jianzhi@umich.edu

TABLE 1
Codon volatility under the simplest model of mutation

| | | | | | | | | | | | |
|-----|-----|-------|-----|-----|-------|-----|------|-------|-----|------|-------|
| TTT | Phe | 0.889 | TCT | Ser | 0.667 | TAT | Tyr | 0.857 | TGT | Cys | 0.875 |
| TTC | Phe | 0.889 | TCC | Ser | 0.667 | TAC | Tyr | 0.857 | TGC | Cys | 0.875 |
| TTA | Leu | 0.714 | TCA | Ser | 0.571 | TAA | Stop | | TGA | Stop | |
| TTG | Leu | 0.750 | TCG | Ser | 0.625 | TAG | Stop | | TGG | Trp | 1.000 |
| CTT | Leu | 0.667 | CCT | Pro | 0.667 | CAT | His | 0.889 | CGT | Arg | 0.667 |
| CTC | Leu | 0.667 | CCC | Pro | 0.667 | CAC | His | 0.889 | CGC | Arg | 0.667 |
| CTA | Leu | 0.556 | CCA | Pro | 0.667 | CAA | Gln | 0.875 | CGA | Arg | 0.500 |
| CTG | Leu | 0.556 | CCG | Pro | 0.667 | CAG | Gln | 0.875 | CGG | Arg | 0.556 |
| ATT | Ile | 0.778 | ACT | Thr | 0.667 | AAT | Asn | 0.889 | AGT | Ser | 0.889 |
| ATC | Ile | 0.778 | ACC | Thr | 0.667 | AAC | Asn | 0.889 | AGC | Ser | 0.889 |
| ATA | Ile | 0.778 | ACA | Thr | 0.667 | AAA | Lys | 0.875 | AGA | Arg | 0.750 |
| ATG | Met | 1.000 | ACG | Thr | 0.667 | AAG | Lys | 0.875 | AGG | Arg | 0.778 |
| GTT | Val | 0.667 | GCT | Ala | 0.667 | GAT | Asp | 0.889 | GGT | Gly | 0.667 |
| GTC | Val | 0.667 | GCC | Ala | 0.667 | GAC | Asp | 0.889 | GGC | Gly | 0.667 |
| GTA | Val | 0.667 | GCA | Ala | 0.667 | GAA | Glu | 0.875 | GGA | Gly | 0.625 |
| GTG | Val | 0.667 | GCG | Ala | 0.667 | GAG | Glu | 0.875 | GGG | Gly | 0.667 |

don volatility. Thus, the utility of this method appears limited.

RESULTS AND DISCUSSION

Volatility varies slightly among genes: Because volatility is between 0.5 and 1 for each codon, it follows that

all genes have volatilities between 0.5 and 1. Figure 1A shows the distribution of volatility for all 3624 genes from the complete genome sequence of the K12 strain of *Escherichia coli*. The mean volatility is 0.761 and the standard deviation is 0.014. One can see that gene volatility has a narrow distribution, with the standard deviation being only 1.8% of the mean. The distribution is

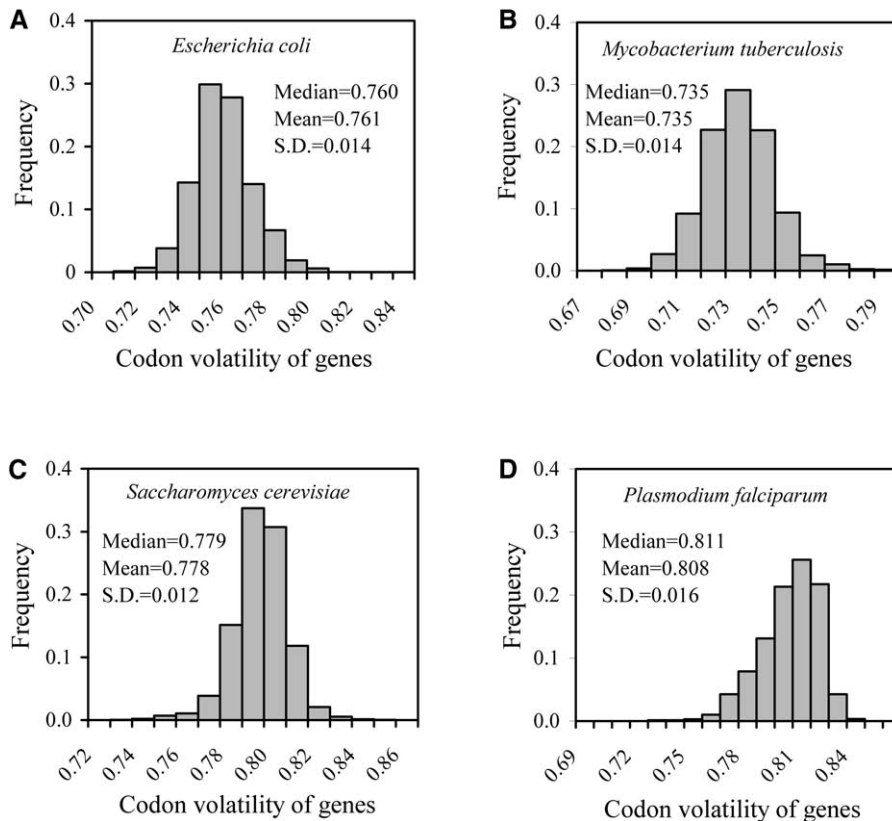


FIGURE 1.—Frequency distributions of codon volatility among all genes of (A) *E. coli*, (B) *Mycobacterium tuberculosis*, (C) *S. cerevisiae*, and (D) *P. falciparum*. Volatility server (<http://volatility.cgr.harvard.edu/cgi-bin/volatility.pl>) was used for the computation.

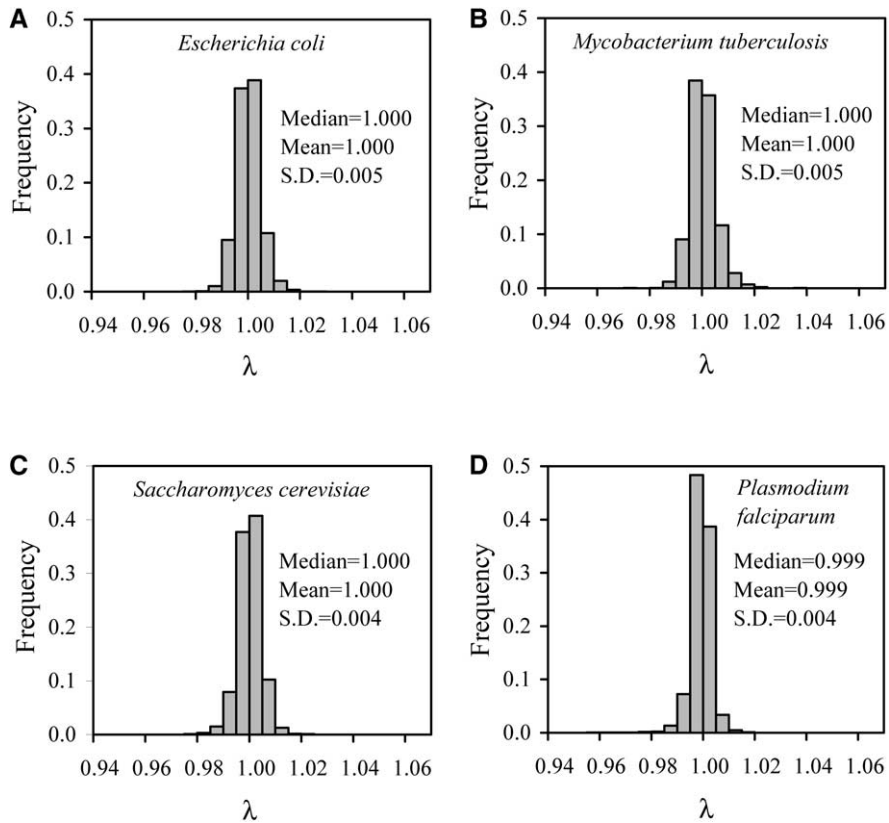


FIGURE 2.—Frequency distributions of λ (observed volatility divided by expected volatility controlled for the amino acid composition of the gene and the genomic average of synonymous codon usage) among all genes of (A) *E. coli*, (B) *M. tuberculosis*, (C) *S. cerevisiae*, and (D) *P. falciparum*. Volatility server (<http://volatility.cgr.harvard.edu/cgi-bin/volatility.pl>) was used for the computation.

also symmetric, as the median (0.760) is very close to mean (0.761). The same pattern is observed in several other prokaryotic and eukaryotic genomes examined (Figure 1), except for the malaria parasite *Plasmodium falciparum*, which appears to have an asymmetric distribution that may be due to its exceptionally high AT content in the genome.

As mentioned, gene volatility is affected by the amino acid composition of the gene as well as by synonymous codon usage. It is thus possible to compute the expected volatility of the gene using its amino acid composition and the average frequencies of synonymous codons found in the genome (PLOTKIN *et al.* 2004). The ratio (λ) of the observed volatility to the expected volatility measures the deviation in volatility due to the bias of synonymous codon usage of the gene from the genomic average. Higher volatility than expected is indicated by $\lambda > 1$ and lower volatility than expected is indicated by $\lambda < 1$. Figure 2A shows the distribution of λ for all the genes of *E. coli* K12. Again, the distribution is very narrow, with the standard deviation being only 0.46% of the mean. The same is true for other species examined (Figure 2). These patterns show that the variation in gene volatility is largely due to the among-gene variation in amino acid composition and that gene volatility becomes extremely homogenous when the amino acid composition variation is controlled for. This implies the rarity of selection that would substantially increase gene volatility.

Directional positive selection does not increase codon volatility: To investigate whether continuous positive selection favoring nonsynonymous substitutions would increase the codon volatility of a gene, I conducted a computer simulation. In the simulation, I randomly generated a sense codon and then introduced mutations randomly according to the SMM described above. I used different types of selection measured by the d_N/d_S ratio. For example, when d_N/d_S was 2, a nonsynonymous mutation was twice as likely as a synonymous mutation to be fixed. After a long period of evolution (equivalent to $d_S = 10$), I measured the volatility of the resultant codon and compared it with that of the initial codon. This simulation was repeated 20,000 times. The results showed that neither positive ($d_N/d_S > 1$) nor negative ($d_N/d_S < 1$) selection affected the evolution of codon volatility (Table 2). The average codon volatility was always 0.74–0.75 and codon frequencies did not change by the selection. For example, in the extreme case of positive selection with $d_N/d_S = 8$, the mean starting volatility was 0.74647 and the mean resultant volatility was 0.74678, and their difference was not statistically significant ($P > 0.05$; Z-test). Among the 20,000 codons simulated, volatility increased in 7771 codons during the evolution, reduced in 7740 codons, and remained unchanged in the rest of the cases. Again, the two numbers, 7771 and 7740, are not significantly different ($P > 0.05$; binomial test). These simulation results demonstrate that directional positive selection for nonsynony-

TABLE 2

Directional positive or negative selection does not affect codon volatility

| d_N/d_S | Mean starting volatility \pm SEM ^a | Mean ending volatility \pm SEM ^a | Increase ^b | Decrease ^c | Ratio ^d |
|-----------|---|---|-----------------------|-----------------------|--------------------|
| 8.0 | 0.74647 \pm 0.00085 | 0.74678 \pm 0.00085 | 7771 | 7740 | 1.004 |
| 4.0 | 0.74650 \pm 0.00085 | 0.74690 \pm 0.00085 | 7837 | 7694 | 1.019 |
| 2.0 | 0.74610 \pm 0.00085 | 0.74650 \pm 0.00085 | 7780 | 7769 | 1.001 |
| 1.0 | 0.74737 \pm 0.00085 | 0.74676 \pm 0.00085 | 7722 | 7701 | 1.003 |
| 0.5 | 0.74768 \pm 0.00085 | 0.74577 \pm 0.00085 | 7700 | 7868 | 0.979 |
| 0.2 | 0.74620 \pm 0.00085 | 0.74702 \pm 0.00085 | 7560 | 7527 | 1.004 |
| 0.1 | 0.74734 \pm 0.00085 | 0.74792 \pm 0.00085 | 6875 | 6907 | 0.995 |

^a Under none of the conditions examined is the mean ending volatility significantly different from the mean starting volatility ($P > 0.05$, Z-test).

^b Number of cases (in 20,000 simulations) in which the ending codon volatility exceeds the starting volatility.

^c Number of cases (in 20,000 simulations) in which the starting codon volatility exceeds the ending volatility.

^d Ratio of the number of increase cases to the number of decrease cases. Under none of the conditions examined is the ratio significantly different from 0 ($P > 0.05$, binomial test).

mous changes does not increase codon volatility and thus volatility cannot be used as an indicator for such selection.

Strong frequency-dependent selection may increase codon volatility: In the above simulation, I assumed that for every codon, there were at most two alleles segregating in the population at any given time. This is identical to the infinite-site model of population genetics (KIMURA 1969, 1971) applied at the codon level and it is a reasonable model for most genes in reality. For example, virtually all single-nucleotide polymorphisms observed in human populations have only two different nucleotides. However, this model does not apply to genes under strong overdominant selection or frequency-dependent selection (minority advantage), such as the mammalian major histocompatibility complex genes where more than two different nucleotides may be segregating at a single site (HUGHES and NEI 1988, 1989). I suspect that these types of selection might increase codon volatility for two reasons. First, it has been noted that high-volatility codons tend to have neighboring codons of high volatility (PLOTKIN and DUSHOFF 2003), as shown in Figure 3. This means that codon volatility is

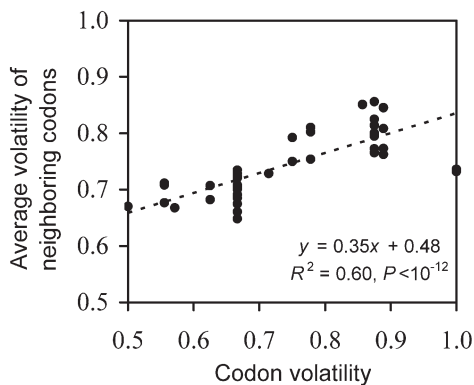


FIGURE 3.—Correlation between the volatility of a codon and the average volatility of its neighboring codons.

heritable. Second, under strong overdominant selection or frequency-dependent selection, multiple alleles at a given codon site may be segregating. This, in conjunction with the heritability of codon volatility, may allow high-volatility codons to generate more nonsynonymous codons that are positively selected for, which could increase the average volatility of the gene in the population.

To investigate this possibility, I conducted a computer simulation of frequency-dependent selection in a haploid population. Frequency-dependent selection and overdominant selection have similar effects on molecular evolution (TAKAHATA and NEI 1990). I here used the frequency-dependent selection in a haploid population due to its simplicity. I began the simulation by randomly generating a sense codon and assigning it to the entire population of N haploid individuals in generation 0. In generation t , N alleles were chosen randomly from the gene pool of generation $t - 1$, under both selection and drift. All synonymous codons encoding amino acid i had the same fitness of $f_i = 1 - sp_i$, where p_i was the total frequency of synonymous codons encoding amino acid i and s ($0 \leq s \leq 1$) was the selection coefficient. This fitness formula reflected one form of minority advantage, where the fitness of an allele was linearly determined by the allele frequency. Random mutations were then generated with a rate of u per nucleotide site per generation. Evolution continued for T generations, and the average codon volatility for the population in generation T was computed and compared with that in generation 0. I used sufficiently large T values to ensure that the population reached equilibrium at the end of the simulation. The simulation was repeated 10,000 times. Table 3 lists the simulation results under various conditions. To reduce computer time, I used small population sizes (N) but relatively large selective coefficients (s) and high mutation rates (u) so that Nu was in the range of 0.001–1 and Ns was in the range of 0–99. Nu is usually between 0.01 and 1 for prokaryotes and unicellular eukaryotes and on the order of 0.001 or lower for multisel-

TABLE 3
Increase of codon volatility by frequency-dependent selection for nonsynonymous mutations

| N^a | u^b | s^c | T^d | Starting volatility ^e | Ending volatility ^f | P^g | π^h | k^i | Increase ^j | Decrease ^k | Ratio ^l | P^m |
|-------|--------|-------|--------|----------------------------------|--------------------------------|---------|---------|-------|-----------------------|-----------------------|--------------------|---------|
| 10 | 0.0001 | 0.00 | 25,000 | 0.7466 ± 0.0012 | 0.7463 ± 0.0012 | NS | 0.002 | 1.02 | 3859 | 3826 | 1.009 | NS |
| 10 | 0.0001 | 0.10 | 25,000 | 0.7475 ± 0.0012 | 0.7461 ± 0.0012 | NS | 0.002 | 1.02 | 3851 | 3917 | 0.983 | NS |
| 10 | 0.0001 | 0.50 | 25,000 | 0.7507 ± 0.0012 | 0.7522 ± 0.0012 | NS | 0.011 | 1.07 | 3999 | 3921 | 1.020 | NS |
| 10 | 0.0001 | 0.99 | 25,000 | 0.7484 ± 0.0012 | 0.7551 ± 0.0011 | <0.0001 | 0.102 | 1.54 | 4489 | 3946 | 1.138 | <0.0001 |
| 10 | 0.0010 | 0.10 | 5,000 | 0.7465 ± 0.0012 | 0.7483 ± 0.0012 | NS | 0.023 | 1.19 | 4010 | 3893 | 1.030 | NS |
| 10 | 0.0010 | 0.50 | 5,000 | 0.7451 ± 0.0012 | 0.7520 ± 0.0011 | <0.0001 | 0.073 | 1.47 | 4285 | 3990 | 1.074 | 0.0006 |
| 10 | 0.0010 | 0.99 | 5,000 | 0.7462 ± 0.0012 | 0.7583 ± 0.0009 | <0.0001 | 0.235 | 2.10 | 4869 | 4169 | 1.168 | <0.0001 |
| 10 | 0.0100 | 0.10 | 1,000 | 0.7475 ± 0.0012 | 0.7487 ± 0.0010 | NS | 0.172 | 2.35 | 4455 | 4468 | 0.997 | NS |
| 10 | 0.0100 | 0.50 | 1,000 | 0.7474 ± 0.0012 | 0.7554 ± 0.0009 | <0.0001 | 0.278 | 2.83 | 4876 | 4434 | 1.100 | <0.0001 |
| 10 | 0.0100 | 0.99 | 1,000 | 0.7471 ± 0.0012 | 0.7618 ± 0.0008 | <0.0001 | 0.367 | 3.20 | 5257 | 4309 | 1.220 | <0.0001 |
| 100 | 0.0001 | 0.10 | 25,000 | 0.7472 ± 0.0012 | 0.7581 ± 0.0010 | <0.0001 | 0.189 | 2.22 | 4865 | 4270 | 1.139 | <0.0001 |
| 100 | 0.0001 | 0.50 | 25,000 | 0.7448 ± 0.0012 | 0.7617 ± 0.0007 | <0.0001 | 0.426 | 4.05 | 5572 | 4243 | 1.313 | <0.0001 |
| 100 | 0.0001 | 0.99 | 25,000 | 0.7468 ± 0.0012 | 0.7725 ± 0.0006 | <0.0001 | 0.504 | 5.25 | 5823 | 4138 | 1.407 | <0.0001 |
| 100 | 0.0010 | 0.10 | 5,000 | 0.7463 ± 0.0012 | 0.7606 ± 0.0008 | <0.0001 | 0.348 | 4.62 | 5403 | 4427 | 1.220 | <0.0001 |
| 100 | 0.0010 | 0.50 | 5,000 | 0.7490 ± 0.0012 | 0.7708 ± 0.0006 | <0.0001 | 0.516 | 6.60 | 5716 | 4258 | 1.342 | <0.0001 |
| 100 | 0.0010 | 0.99 | 5,000 | 0.7488 ± 0.0012 | 0.7748 ± 0.0005 | <0.0001 | 0.575 | 8.02 | 5857 | 4137 | 1.416 | <0.0001 |
| 100 | 0.0100 | 0.10 | 1,000 | 0.7485 ± 0.0012 | 0.7549 ± 0.0005 | <0.0001 | 0.585 | 14.91 | 5548 | 4449 | 1.247 | <0.0001 |
| 100 | 0.0100 | 0.50 | 1,000 | 0.7459 ± 0.0012 | 0.7674 ± 0.0004 | <0.0001 | 0.639 | 16.60 | 5908 | 4092 | 1.444 | <0.0001 |
| 100 | 0.0100 | 0.99 | 1,000 | 0.7464 ± 0.0012 | 0.7738 ± 0.0003 | <0.0001 | 0.662 | 17.70 | 5949 | 4048 | 1.470 | <0.0001 |

NS, not significant.

^aPopulation size.

^bMutation rate per site per generation.

^cSelection coefficient as defined in the main text.

^dNumber of generations of evolution.

^ePopulation volatility or the average volatility of all individuals in the population. The mean population volatility and its standard error at the beginning of the simulation are shown.

^fThe mean population volatility and its standard error at the end of simulation.

^gThe probability that the mean starting and ending volatilities are identical (one-tail Z-test).

^hMean nucleotide diversity per site at the end of evolution.

ⁱMean number of alleles in the population at the end of evolution.

^jNumber of cases (of 10,000 replications) in which the ending population volatility exceeds the starting population volatility.

^kNumber of cases (of 10,000 replications) in which the ending population volatility is lower than the starting population volatility.

^lThe ratio of the number of increase cases to the number of decrease cases.

^mThe probability that the ratio is 1 (one-tail binomial test).

Note

TABLE 4

Expected volatility in highly and lowly expressed genes

| Species | Arg | Gly | Leu | Ser |
|------------------------|-------|-------|-------|-------|
| <i>E. coli</i> | | | | |
| Highly expressed | 0.667 | 0.665 | 0.566 | 0.705 |
| Lowly expressed | 0.646 | 0.660 | 0.641 | 0.742 |
| <i>S. cerevisiae</i> | | | | |
| Highly expressed | 0.741 | 0.667 | 0.744 | 0.673 |
| Lowly expressed | 0.704 | 0.656 | 0.666 | 0.703 |
| <i>D. melanogaster</i> | | | | |
| Highly expressed | 0.669 | 0.655 | 0.597 | 0.700 |
| Lowly expressed | 0.639 | 0.656 | 0.638 | 0.715 |

The expected volatility for each amino acid is the average volatility for all synonymous codons (Table 1) weighted by codon frequencies in SHARP *et al.* (1988). Only the four listed amino acids have synonymous codons of different volatilities.

lular eukaryotes (LYNCH and CONERY 2003). Thus, our simulation was biologically meaningful. As predicted, frequency-dependent selection could increase codon volatility, and this increase became more obvious when N_s and N_u were larger. In higher eukaryotes, codon volatility is less useful as an indicator for positive selection because N and N_u are generally quite small. For instance, in humans, N is about 10^4 (TAKAHATA *et al.* 1995) and u is about 2×10^{-8} per generation (YI *et al.* 2002), and N_u is thus 2×10^{-4} . However, even under the most extreme condition examined ($N = 100$, $s = 0.99$, $u = 0.01$), the mean codon volatility increased by $<4\%$, suggesting that codon volatility responded to only a small extent to the frequency-dependent selection for nonsynonymous changes. Note that the above selection was quite strong, as it could maintain on average 17.7 alleles in the population of 100 haploids and resulted in a nucleotide diversity of 0.662. It should be noted that the above simulation scheme was somewhat different from conventional population genetic analysis, because it was the behavior of a single codon, not that of a DNA sequence, that was examined. The simulation can be regarded as a special case in which at any time of evolution only the variation at one codon of the gene affects fitness. In reality, nonadditive fitness effects of different codons of a gene would make the situation much more complicated. Nevertheless, the simple simulation demonstrates the extent to which frequency-dependent selection could affect codon volatility.

Expression-related codon usage bias potentially affects codon volatility: Although codon volatility could be enhanced by frequency-dependent selection and thus in principle may be used as an indicator for positive selection, codon volatility may also be affected by factors unrelated to positive selection for nonsynonymous changes. For instance, it is well known that synonymous codon usage bias is related to the level of gene expression in many organisms, with a greater degree of bias in highly expressed genes (LI 1997). It is believed that this phe-

nomenon is due to selection for high translational efficiency through preferential use of synonymous codons whose cognate tRNA has high concentrations. As mentioned, four amino acids have synonymous codons of different volatilities. Table 4 lists the average volatility for each of these amino acids computed by considering the average synonymous codon usage in highly and lowly expressed genes of *E. coli*, *Saccharomyces cerevisiae*, and *Drosophila melanogaster* (SHARP *et al.* 1988). It is obvious that codon volatility can be affected by gene expression level. The actual amount of effect, however, depends on the frequencies of these four amino acids in a protein, particularly Arg, Leu, and Ser. A recent analysis showed that the synonymous codon usage in genes highly expressed in one stage of the developmental cycle of *P. falciparum* differs from that in genes highly expressed in another stage (PEIXOTO *et al.* 2004). It is unclear, however, whether this difference resulted from codon volatility-related selection (PLOTKIN *et al.* 2004) or from translation efficiency-related selection that may arise if tRNA concentrations vary in different developmental stages.

Furthermore, in some organisms, GC content varies substantially in different regions of the genome (*e.g.*, mammals; BERNARDI 2004), which may lead to variable expected codon volatility for different genomic regions. Thus, use of the genomic average synonymous codon usage to compute the expected codon volatility may be inappropriate.

In this work, the simplest mutation model was used. In reality, mutational patterns are quite complicated and may vary among genes and genomes. Although codon volatility can be computed given a mutational model (PLOTKIN *et al.* 2004), the lack of knowledge of exact mutational models for a given gene or genome adds another layer of uncertainty to the application of codon volatility in detecting positive selection.

Conclusions: In contrast to the claim of PLOTKIN *et al.* (2004), I find that codon volatility does not increase by directional positive selection for nonsynonymous changes, and it increases only slightly by strong frequency-dependent selection in large populations. Given that factors unrelated to positive selection also affect codon volatility, the utility of this measure in detecting positive selection at the DNA sequence level seems limited. Nevertheless, the idea of using just one DNA sequence to detect natural selection (PLOTKIN *et al.* 2004) is novel and attractive, and it would be interesting to develop other measures that may accomplish this goal.

I thank Hunter Fraser and Xionglei He for discussions. David Webb and two anonymous reviewers provided valuable comments on the manuscript. This work was supported by a startup fund from the University of Michigan and National Institutes of Health grant GM-67030 to J.Z.

LITERATURE CITED

- BERNARDI, G., 2004 *Structural and Evolutionary Genomics: Natural Selection in Genome Evolution*. Elsevier, Amsterdam.

- HUGHES, A. L., and M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167–170.
- HUGHES, A. L., and M. NEI, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* **86**: 958–962.
- KIMURA, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**: 893–903.
- KIMURA, M., 1971 Theoretical foundation of population genetics at the molecular level. *Theor. Popul. Biol.* **2**: 174–208.
- LI, W. H., 1997 *Molecular Evolution*. Sinauer, Sunderland, MA.
- LYNCH, M., and J. S. CONERY, 2003 The origins of genome complexity. *Science* **302**: 1401–1404.
- NEI, M., and S. KUMAR, 2000 *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- PEIXOTO, L., V. FERNANDEZ and H. MUSTO, 2004 The effect of expression levels on codon usage in *Plasmodium falciparum*. *Parasitology* **28**: 245–251.
- PLOTKIN, J. B., and J. DUSHOFF, 2003 Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proc. Natl. Acad. Sci. USA* **100**: 7152–7177.
- PLOTKIN, J. B., J. DUSHOFF and H. B. FRASER, 2004 Detecting selection using a single genome sequence of *M. tuberculosis* and *P. falciparum*. *Nature* **428**: 942–945.
- SHARP, P. M., E. COWE, D. G. HIGGINS, D. C. SHIELDS, K. H. WOLFE *et al.*, 1988 Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res.* **16**: 8207–8211.
- TAKAHATA, N., and M. NEI, 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**: 967–978.
- TAKAHATA, N., Y. SATTA and J. KLEIN, 1995 Divergence time and population size in the lineage leading to modern humans. *Theor. Popul. Biol.* **48**: 198–221.
- YI, S., D. L. ELLSWORTH and W. H. LI, 2002 Slow molecular clocks in Old World monkeys, apes, and humans. *Mol. Biol. Evol.* **19**: 2191–2198.

Communicating editor: S. W. SCHAEFFER

