Synonymous Codon Bias Is Not Caused by Mutation Bias in G+C-Rich Genes in Humans

Nick G. C. Smith and Adam Eyre-Walker

Centre for the Study of Evolution and School of Biological Sciences, University of Sussex, Brighton, England

It is has been suggested that synonymous codon bias is a consequence of mutation bias in mammals. We tested this hypothesis in humans using single-nucleotide polymorphism data. We found a pattern of polymorphism which was inconsistent with the mutation bias hypothesis in G+C-rich genes. However, the data were consistent with the action of natural selection or biased gene conversion. Similar patterns of polymorphism were also observed in noncoding DNA, suggesting that natural selection or biased gene conversion may affect large tracts of the human genome.

Introduction

It is well established that selection acts on synonymous codon use in many groups of organisms, including bacteria, fungi, and insects (Sharp et al. 1992). However, in those groups of organisms, the degree of synonymous codon bias is correlated to the level of gene expression (Gouy and Gautier 1982; Ikemura 1985; Duret and Mouchiroud 1999), and this is not observed in mammals (Duret and Mouchiroud 1999). Instead, the pattern of synonymous codon use is correlated to the G+C content of the genomic region in which the gene resides; genes in the G+C-rich regions of the genome preferentially use G- and C-ending codons, while those in the A+T-rich regions use A- and T-ending codons (Bernardi et al. 1985; Bernardi 1995). This pattern has led to the suggestion that synonymous codon bias is caused by mutation bias in mammals (Filipski 1987; Suoeka 1988; Wolfe, Sharp, and Li 1989).

We can test whether synonymous codon bias is caused by mutation bias using population genetic data. Let u be the mutation rate from G : C base pairs to A : T base pairs, and let v be the mutation rate in the opposite direction. If mutation rates are low (i.e., $N_e u \ll 1$ and $N_e v \ll 1$, where N_e is the effective population size) and constant, and no other evolutionary forces affect base composition, then the equilibrium frequency of G : C base pairs in a sequence is $f = v/(v + u)$ (Suoeka 1962). Therefore, the probability that we will observe an A or T mutation segregating at a site which was ancestrally G or C, henceforth referred to as a GC→AT mutation, is $M_{GC\rightarrow AT} = fuH(n)$, where $H(n)$ is the probability of observing a neutral mutation in a sample of n sequences, and the probability of observing a G or C mutation at a site which was ancestrally A or T, henceforth referred to as an AT→GC mutation, is $M_{AT\rightarrow GC} = (1 - f)vH(n)$. It is not difficult to show that $M_{GC\rightarrow AT} = M_{AT\rightarrow GC}$; i.e., the number of AT→GC mutations segregating in a sample is expected to be equal to the number of GC→AT mutations if mutation bias is the sole cause of synonymous codon bias (Eyre-Walker 1997, 1999).

A recent analysis showed that there were more GC→AT mutations than AT→GC mutations segregating at synonymous sites in mammalian MHC genes, suggesting that mutation bias was not solely responsible for synonymous codon bias (Eyre-Walker 1999). However, it was not possible to demonstrate conclusively that the data conformed to the infinite-sites model (the requirement that mutation rates are low), and the results lacked generality, since for each species, all the studied genes came from a small region of a single chromosome.

A large number of single-nucleotide polymorphisms (SNPs) from human protein-coding genes, dispersed throughout the genome, have recently been published (Cargill et al. 1999; Hacia et al. 1999). For many of these SNPs, the corresponding sites have been sequenced in chimpanzees. Since the divergence between humans and chimpanzees is low (~0.015 at fourfold-degenerate synonymous sites; Eyre-Walker and Keightley 1999), the chimpanzee sequence can be used to infer the ancestral state in humans (i.e., whether an SNP segregating X and Y is due to an X→Y or a Y→X mutation). Furthermore, the average nucleotide diversity at fourfold-degenerate sites in human genes is sufficiently low (~0.001; Li and Stadler 1991; Cargill et al. 1999) for the data to conform to the infinite-sites model, even at CpG dinucleotides which mutate approximately 10–20 times as fast as other sites (Bulmer 1986; Sved and Bird 1990).

In this paper, we test the mutation bias hypothesis (i.e., whether mutation bias is responsible for synonymous codon bias) in humans by analyzing the pattern of polymorphism in synonymous SNPs.

Materials and Methods

Data

SNP data from two recent studies were obtained from their respective websites (http://waldo.wi.mit.edu/cvar_snps/ for Cargill et al. [1999]; http://genome.nhgri.nih.gov/apes/ for Hacia et al. [1999]). Both data sets provided the following information: the nucleotides segregating in humans, the nucleotide(s) at the same site in chimpanzee sequences, reference names for the sequences containing the SNPs, and the nucleotide se-

Abbreviations: EST, expressed sequence tag; MHC, major histocompatibility complex; SNP, single-nucleotide polymorphism; STS, sequence tagged site; UTR, untranslated region.

Key words: human, synonymous codons, mutation bias.

Address for correspondence and reprints: Adam Eyre-Walker, Centre for the Study of Evolution and School of Biological Sciences, University of Sussex, Brighton BN1 9QG, United Kingdom. E-mail: a.c.eyre-walker@sussex.ac.uk.

Mol. Biol. Evol. 18(6):982–986. 2001

quences flanking the SNP. By assuming the chimpanzee nucleotide to be the ancestral state, SNPs were classified as GC→AT or AT→GC according to the mutation which generated them. We ignored those SNPs at which A/T or G/C were segregating and the few sites for which ancestral state reconstruction was ambiguous (if the chimpanzee site was polymorphic or if the chimpanzee nucleotide differed from both human nucleotides). We obtained the sequence containing each SNP by using the accession number from the Human SNP database (www-genome.wi.mit.edu/SNP/human/index.html) or by using the SNP-flanking sequences in a BLAST search. Annotations in the GenBank sequences allowed us to classify the SNPs into four classes: there were 125 synonymous, 60 intron, 60 3′ untranslated region (UTR), and 49 anonymous STS SNPs. All of the synonymous and intron SNPs came from Cargill et al. (1999), while all of the STS SNPs came from Hacia et al. (1999); exactly half of the 3′ UTR SNPs came from each data set. STS SNPs not found in exons, introns, ESTs, or UTRs were categorized as anonymous; we ignored SNPs for which no clear annotation was available (e.g., unannotated EST sequences). For each synonymous SNP, we calculated the third position G+C content (GC_3) of the exon in which the SNP was contained, or GC_3 for the complete coding sequence if the intron/exon boundaries were not known. For SNPs in introns and UTRs, we calculated the G+C content for the intron or UTR concerned, and for STSs, we used the longest sequence available, up to 500 bp either side of the SNP.

CpG Islands

CpG islands were identified by calculating the expected number of CpG's based on the base composition and comparing this number to the level observed. If the observed/expected ratio was >50%, the SNP was inferred to be in a CpG island. For CpG analysis, we used the longest available contiguous sequence. If the sequence length was >600 bp, then a sliding-window analysis was performed (window length = 300 bp, step length = 50 bp), and the maximum observed/expected value overlapping the SNP was taken. For shorter contiguous sequences, we took the observed/expected value for the entire sequence.

Results and Discussion

Synonymous SNPs

There are 125 GC↔AT synonymous SNPs in the data set of Cargill et al. (1999) for which the chimpanzee sequence has been determined and a complete human cDNA sequence exists. As in a previous analysis of synonymous SNPs in MHC genes (Eyre-Walker 1999), there is a clear excess of GC→AT mutations segregating at synonymous sites (88 GC→AT and 37 AT→GC mutations, $P < 0.00001$). The excess of GC→AT mutations is particularly evident in those genes which preferentially use G- and C-ending codons (for SNPs in exons with $GC_3 > 0.6$, 60 GC→AT and 11 AT→GC mutations, $P < 0.00001$) (table 1); there is no evidence of an excess of GC→AT mutations in genes with low GC_3.

Table 1
The Numbers of GC→AT and AT→GC Synonymous Mutations Segregating in Human Genes

GC_3	GC→AT	AT→GC	*P*
0.20–0.30	2	0	NS
0.30–0.40	9	7	NS
0.40–0.50	9	12	NS
0.50–0.60	8	7	NS
0.60–0.70	19	4	0.003
0.70–0.80	18	4	0.005
0.80–0.90	23	3	0.0001
Total.....	88	37	6×10^{-6}

NOTE.—The data are divided according to the GC_3 (G+C content at the third codon position) of the exon containing the single-nucleotide polymorphisms, and the *P* value is for the test of $M_{GC\to AT} = M_{AT\to GC}$ obtained from a binomial distribution $B[M_{GC\to AT} + M_{AT\to GC}, 0.5]$.

Sampling Bias

While this result would seem to be inconsistent with the mutation bias hypothesis in the G+C-rich genes, there are a number of explanations for the excess of GC→AT mutations which need to be considered: sampling bias, hypermutable sites, and a recent change in the pattern of mutation. It seems unlikely that our results were due to biases in the methods used to detect the SNPs for several reasons (i.e., ascertainment bias). First, Cargill et al. (1999) estimate that they detected at least 85% of all SNPs. Second, we would not expect the excess of GC→AT mutations to increase with increasing G+C content, as we see in the data (table 1); since under the mutation bias hypothesis we expect equal numbers of GC→AT and AT→GC mutations at all G+C contents, we would therefore expect a similar level of ascertainment bias at all compositional levels. Third, a similar excess of synonymous GC→AT mutations was observed in MHC genes, where the mutations were detected by a different method, direct sequencing (Eyre-Walker 1999).

Hypermutation

Hypermutable sites potentially have two effects; they could lead to problems with parsimony, and they could violate the infinite-sites assumption. In each case, if the hypermutable sites had elevated rates of AT→GC mutation, they would tend to generate an excess of GC→AT mutations, as we see in the data. The reasons for this rather counterintuitive behavior are fully discussed elsewhere (Eyre-Walker 1998, 1999). However, three lines of evidence suggest that hypermutable sites were not responsible for the excess of GC→AT mutations we observed. First, we are not aware of any evidence of AT→GC hypermutable sites in mammals; the one well-known class of hypermutable sites, CpG dinucleotides, are expected to cause a bias in the opposite direction of that required to explain the data: CpG dinucleotides generate C→T and G→A transitions at elevated rates, and such mutations will tend to appear as

T→C and A→G changes, respectively, in the data (Eyre-Walker 1998, 1999). Second, it is possible to demonstrate that the excess in GC→AT mutations is not due to a problem with parsimony, since we can dispense with the chimpanzee sequence and infer the direction of mutation from the frequencies of the alleles segregating at a site; the rarer allele is assumed to be more recent. This method is unbiased under the null hypothesis (synonymous codon bias is caused by mutation bias) and the infinite-sites assumption (Eyre-Walker 1999). Using allele frequencies, we infer that there have been 65 GC→AT mutations, compared with 37 AT→GC mutations over all genes ($P = 0.007$) and 45 GC→AT versus 15 AT→GC mutations ($P = 0.0001$) for genes with $GC_3 > 0.6$; the sample sizes are smaller because frequency data are available for only a subset of the SNPs. Third, the infinite-sites assumption would only be seriously compromised in this context if the rate of mutation were some 100 times as high as the average nucleotide diversity observed (Eyre-Walker 1999), and with that level of hypermutability, we would expect to see an excess of GC→AT substitutions inferred by parsimony (Eyre-Walker 1998) over even short timescales, such as the divergence along the human lineage since we split from chimpanzees (Eyre-Walker and Keightley 1999). In a sample of 28 genes sequenced in humans, chimpanzees, and gorillas (Eyre-Walker and Keightley 1999), there have been identical numbers of GC→AT and AT→GC synonymous substitutions along the human lineage (22 substitutions in each direction inferred by parsimony, 18 GC→AT and 15 AT→GC substitutions for genes with $GC_3 > 0.6$), just as we expect for a sequence of stationary base composition.

Mutation Pattern

The excess of GC→AT mutations segregating in human SNPs could be the result of a recent change in the mutation pattern from a GC bias to an AT bias, but this seems unlikely for three reasons. First, a change in the mutation pattern would manifest itself as an excess of GC→AT substitutions over AT→GC substitutions unless the change in the mutation pattern had been very recent. As we showed above, there appear to have been similar numbers of GC→AT and AT→GC substitutions along the human lineage since the split from chimpanzees. Second, a dramatic change in the mutation pattern is required to explain the data. For example, there are 18 GC→AT mutations and 4 AT→GC mutations for the SNPs in exons with GC_3 between 70% and 80%, and the change in the mutation process needed to cause this pattern would eventually reduce GC_3 to ~40% (calculated using eq. 8 in Eyre-Walker [1997]). Third, we would require several independent changes in the mutation pattern in the same direction to explain the excess of GC→AT synonymous polymorphisms in the MHC genes of other mammals (Eyre-Walker 1999).

Selection and Biased Gene Conversion

It therefore seems that mutation bias is not responsible for synonymous codon bias in human genes. However, there are at least two other possibilities: natural selection and biased gene conversion; biased gene conversion is a process which leads to the biased transmission of alleles; for example, if biased gene conversion is very strong and G+C-biased, 100% of all gametes from a C/T heterozygote will be C. Both selection and biased gene conversion are expected to generate an excess of GC→AT mutations. This can be seen using the following simple argument: Let us imagine there is no mutation bias, and selection has elevated the G+C content of a sequence to 80%. Since there is no mutation bias, 80% of the new mutations will be GC→AT, and 20% will be AT→GC (ignoring G↔C and A↔T mutations). Unfortunately, the situation is more complicated, because selection may affect the probability of detecting a mutation; for example, if directional selection had elevated the G+C content to 80% in the previous example, each GC→AT mutation would be slightly deleterious, while each AT→GC would be slightly advantageous; we would therefore expect to detect the AT→GC mutations more readily, because they would segregate at slightly higher frequencies, on average, than the GC→AT mutations.

To demonstrate formally that selection and biased gene conversion are expected to generate an excess of GC→AT mutations, we derived the expected proportion of GC→AT mutations segregating in a sample of sequences, $P_{GC\to AT}$, under two models: a model of weak directional selection, which is equivalent to a model of biased gene conversion (Nagylaki 1983); and a model of strong stabilizing selection. Let f' (or f'') be the frequency of sites fixed for G : C base pairs, u be mutation rate from G : C to A : T base pairs, and v be the mutation rate in the opposite direction. We will assume that selection or biased gene conversion favors high G+C. First, consider weak directional selection and biased gene conversion, two processes which can be described by a single parameter s, since they are dynamically identical (Nagylaki 1983). Under semidominant directional selection, s is the strength of selection in favor of G : : C base pairs, and under biased gene conversion, s is the strength of biased gene conversion, where $(s + 1)/2$ of the alleles from a G : C/A : T heterozygote are G or C. If mutation rates are low enough that the infinite-sites assumption holds (i.e., $N_e u \ll 1$, $N_e v \ll 1$), the equilibrium proportion of sites fixed for G : C in a diploid is

$$f' = \frac{1}{1 + (u/v)e^{-S}} \tag{1}$$

(Li et al. 1987; Bulmer 1991), where $S = 4N_e s$. Therefore, in a sample of n sequences, the expected proportions of GC→AT and AT→GC mutations are given by

$$M'_{GC\to AT} = f'uQ(n, -S)$$
$$M'_{AT\to GC} = (1 - f')vQ(n, S) \tag{2}$$

where

$$Q(n, S) = \sum_{i=1}^{n-1} \int_0^1 \frac{n!}{i!(n-i)!} \frac{(1 - e^{-S(1-x)})}{1 - e^{-S}} x^{i-1}(1-x)^{n-i-1}\, dx$$

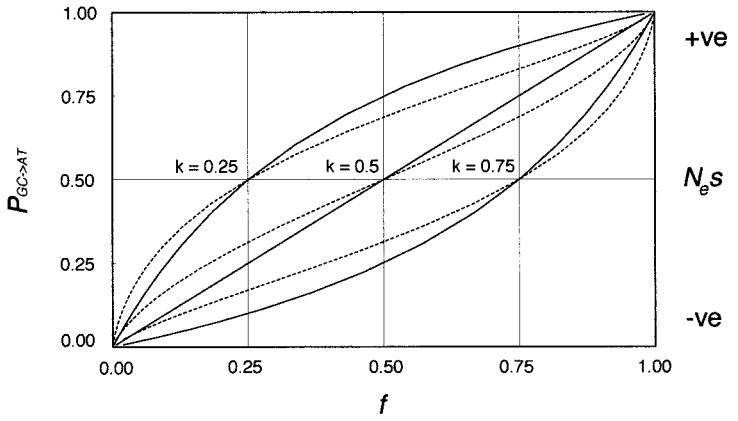

FIG. 1.—The expected proportion of GC→AT mutations segregating in a sample of sequences ($P_{GC\to AT}$) plotted against the G+C content of the sequence for a variety of mutation biases ($k = v/(u + v)$) for two models: weak directional selection/biased gene conversion (solid lines) (eq. 3) and stabilizing selection (dashed lines) (eq. 4). Note that for the first model, $P'_{GC\to AT}$ is a function of k, S, and f', but f' is itself a function of k and S, so $P'_{GC\to AT}$ can be plotted parametrically against f' as S varies. In each case, the sample size is assumed to be 10 sequences; the results are very similar for other sample sizes.

(Sawyer and Hartl 1992), assuming there is free recombination. The proportion of GC→AT mutations segregating in the sample is simply

$$P'_{GC\to AT} = \frac{M'_{GC\to AT}}{M'_{GC\to AT} + M'_{AT\to GC}}. \quad (3)$$

Second, consider a stabilizing-selection model in which selection is acting to maintain the G+C content of a sequence at f''. If we assume that selection is sufficiently strong that the average G+C content of the sequence in the population is at the optimum, f'', and that mutation rates are sufficiently low that each mutation in the sequence under stabilizing selection appears, segregates, and is removed before the next occurs, then each mutation, whether it is GC→AT or AT→GC, is deleterious. If we assume that selection is symmetrical about the optimum, then each mutation will be subject to the same level of selection; let the strength of selection be s against the mutation. Then, we have

$$M''_{GC\to AT} = f''uQ(n, -S)$$

$$M''_{AT\to GC} = (1 - f'')vQ(n, -S)$$

$$P''_{GC\to AT} = \frac{M''_{GC\to AT}}{M''_{GC\to AT} + M''_{AT\to GC}}. \quad (4)$$

As figure 1 shows, when selection favors increased G+C, we expect an excess of GC→AT mutations under both models, and when selection favors increased A+T, we expect a deficit of GC→AT mutations. This is likely to be the pattern we expect under most models of selection, since the stabilizing- and directional-selection models lie at opposite ends of a continuum; as selection becomes weak in the stabilizing-selection model, mutation pressure will push the population away from the optimum; if selection becomes very weak, then the population will be sufficiently far below the optimum that the model becomes a weak directional-selection model.

CpG Dinucleotides

While both selection and biased gene conversion are consistent with the data presented here, there are few data which can discriminate between them at present. We can test two simple selective hypotheses: that selection is acting on synonymous codon use, but only to maintain (1) CpG islands, ~1-kb sequences which have high levels of the dinucleotide CpG and high G+C content, or (2) methylated CpG dinucleotides. Both CpG islands and methylated CpGs have been implicated in the regulation of gene expression (Lewis and Bird 1991) and might therefore be targets of natural selection. However, while the excess of GC→AT mutations is very apparent for both CpG islands and CpG dinucleotides (CpG islands: 14 GC→AT mutations and 1 AT→GC mutation, $P = 0.0005$; CpG dinucleotides: 47 GC→AT and 15 AT→GC mutations, $P = 0.0001$ at SNPs segregating C/T at a site flanked 3′ by G, or G/A at a site flanked 5′ by C), there is an excess of GC→AT mutations both for non-CpG island DNA and for dinucleotides other than CpG (non-CpG island: 73 GC→AT and 36 AT→GC mutations, $P = 0.0005$; other dinucleotides: 41 GC→AT and 15 AT→GC mutations, $P = 0.023$).

Noncoding DNA

It is likely that whatever affects synonymous codon bias also affects large regions of the genome, since in mammals synonymous codon bias is correlative with the base composition of the chromosomal region in which the gene is situated—i.e., GC_3 is strongly correlated to the G+C content of the 5′ and 3′ UTR regions, introns, and isochores (Bernardi et al. 1985; Clay et al. 1996). As expected, there is an excess of GC→AT mutations segregating in intron, 3′ UTR, and anonymous STS sequences (i.e., STS sequences which are not known to be within or flanking a protein-coding sequence), particularly in those sequences which are G+C rich (table 2). It therefore seems that either natural selection or biased gene conversion also affects the base composition of G+C rich noncoding DNA and therefore has a profound effect on the structure of the human genome, since large sections of the genome are G+C-rich, while others are G+C-poor (Bernardi 1995).

Table 2
The Numbers of GC→AT and AT→GC Single-Nucleotide Polymorphisms (SNPs) Segregating in Introns, 3′ Untranslated Regions (UTRs), and Anonymous Sequence Tagged Site (STS) Sequences

G+C Content	Intron GC→AT : AT→GC	3′ UTR GC→AT : AT→GC	STS GC→AT : AT→GC
0.20–0.30	3:5	3:2	2:1
0.30–0.40	7:5	14:14	11:8
0.40–0.50	8:3	10:5	15:9
0.50–0.60	8:0**	6:3	2:1
0.60–0.70	13:8	3:0	0:0
Total	39:21*	36:24	30:19
Overall total		105:64**	

NOTE.—SNPs are divided according to the G+C content of the sequence containing them, and significant P values for the binomial test of $M_{GC\rightarrow AT} = M_{AT\rightarrow GC}$ are indicated (* $P < 0.05$; ** $P < 0.01$).

Acknowledgments

We thank Eric Lander, Francis Collins, and their groups for making their data available, and Gil McVean, Laurence Hurst, and Peter Keightley for comments and helpful discussion. This work was supported by the BBSRC (N.G.C.S., A.E.-W.) and the Royal Society (A.E.-W.).

LITERATURE CITED

BERNARDI, G. 1995. The human genome: organization and evolutionary history. Annu. Rev. Genet. **29**:445–476.

BERNARDI, G., B. OLOFSSON, J. FILIPSKI, M. ZERIAL, J. SALINAS, G. CUNY, M. MEUNIER-ROTIVAL, and F. RODIER. 1985. The mosaic genome of warm blooded vertebrates. Science **228**:953–958.

BULMER, M. 1986. Neighbouring base effects on substitution rates in pseudogenes. Mol. Biol. Evol. **3**:322–329.

———. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics **129**:897–907.

CARGILL, M., D. ALTSHULER, J. IRELAND et al. (17 co-authors). 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat. Genet. **22**:231–238.

CLAY, O., S. CACCIO, Z. ZOUBAK, D. MOUCHIROUD, and G. BERNARDI. 1996. Human coding and noncoding DNA: compositional correlations. Mol. Phylogenet. Evol. **5**:2–12.

DURET, L., and D. MOUCHIROUD. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis, Drosophila,* and *Arabidopsis.* Proc. Natl. Acad. Sci. USA **96**:4482–4487.

EYRE-WALKER, A. 1997. Differentiating selection and mutation bias. Genetics **147**:1983–1987.

———. 1998. Problems with parsimony in sequences of biased base composition. J. Mol. Evol. **47**:686–690.

———. 1999. Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics **152**:675–683.

EYRE-WALKER, A., and P. D. KEIGHTLEY. 1999. High genomic deleterious mutation rates in hominids. Nature **397**:344–347.

FILIPSKI, J. 1987. Correlation between molecular clock ticking, codon usage, fidelity of DNA repair, chromosome banding and chromtin compactness in germline cells. FEBS Lett. **217**:184–186.

GOUY, M., and C. GAUTIER. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10**: 7055–7074.

HACIA, J. G., J.-B. FAN, O. RYDER et al. (16 co-authors). 1999. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat. Genet. **22**:164–167.

IKEMURA, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**:13–34.

LEWIS, J., and A. P. BIRD. 1991. DNA methylation and chromatin structure. FEBS Lett. **205**:155–159.

LI, W.-H., and L. A. STADLER. 1991. Low nucleotide diversity in man. Genetics **129**:513–523.

LI, W.-H., M. TANIMURA, and P. M. SHARP. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. **25**:330–342.

NAGYLAKI, T. 1983. Evolution of a finite population under gene conversion. Proc. Natl. Acad. Sci. USA **80**:6278–6281.

SAWYER, S. A., and D. L. HARTL. 1992. Population genetics of polymorphism and divergence. Genetics **132**:1161–1176.

SHARP, P. M., C. J. BURGESS, A. T. LLOYD, and K. J. MITCHELL. 1992. Selective use of termination and variation in codon choice. Pp. 397–425 *in* D. L. HATFIELD, B. J. LEE, and R. M. PIRTLE, eds. Transfer RNA in protein synthesis. CRC Press, Boca Raton, Fla.

SUOEKA, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA **48**:582–592.

———. 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85**:2653–2657.

SVED, J., and A. P. BIRD. 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc. Natl. Acad. Sci. USA **87**:4692–4696.

WOLFE, K. H., P. M. SHARP, and W.-H. LI. 1989. Mutation rates differ among regions of the mammalian genome. Nature **337**:283–285.

MANOLO GOUY, reviewing editor

Accepted January 4, 2001