

Mammalian Nonsynonymous Sites Are Not Overdispersed: Comparative Genomic Analysis of Index of Dispersion of Mammalian Proteins

Seong-Ho Kim¹ and Soojin V. Yi

School of Biology, Georgia Institute of Technology

It is often stated that patterns of nonsynonymous rate variation among mammalian lineages are more irregular than expected or overdispersed under the neutral model, whereas synonymous sites conform to the neutral model. Here we reexamined genome-wide patterns of the variance to mean ratio, or index of dispersion (R), of substitutions in proteins from human, mouse, and dog. Contrary to the prevailing notion, we found that the mean index of dispersion for nonsynonymous sites of mammalian proteins is not significantly different from 1. We propose that earlier analyses were biased because the data included disproportionately more protein hormones, which tend to be more dispersed than genes in other functional categories. Synonymous sites exhibit greater degree of dispersion than nonsynonymous sites, although similar to earlier estimates and potentially due to errors associated with correction for multiple hits. Overall, our analysis identifies strong genome-wide generation-time effect and natural selection as important determinants of among-lineage variation of protein evolutionary rates. Furthermore, patterns of lineage-specific selective constraint are consistent with the nearly neutral model of molecular evolution.

Introduction

A commonly used tool to evaluate the neutral theory of molecular evolution is the so-called index of dispersion (Kimura 1983; Gillespie 1986, 1989, 1991; Ohta 1995; Nielsen 1997; Cutler 2000; Wilke 2004). This statistic measures the ratio of variance to mean of the number of substitutions among lineages, which is expected to follow a Poisson distribution according to the strict neutral theory (Kimura 1983). Thus, the “index of dispersion” should be 1 under neutrality.

Formally, index of dispersion is usually estimated as $R = \text{Var}(N_i) / E(N_i)$, where N_i is the estimated number of substitutions in the i -th lineage. Studies of index of dispersion began early in the history of molecular evolution (Langley and Fitch 1974; Kimura 1983). However, those earlier studies (before Gillespie’s [1989] analysis) suffered from several methodological problems. The first issue arises from estimating the underlying phylogeny. Earlier studies approximated the mammalian radiation as a “star” phylogeny (Kimura 1983) or reconstructed ancestral sequences using parsimony (Langley and Fitch 1974). Such methods can introduce errors associated with branch length due to phylogenetic inaccuracy. Another potential problem is systematic variation of substitution rates among lineages, due to neutral factors such as generation times or mutation rates. Furthermore, methods to correct for multiple hits introduce additional errors (Bulmer 1989; Nielsen 1997).

Gillespie (1989, 1991) improved upon these issues by using 3 species only, thereby avoiding problems associated with phylogenetic reconstruction (because there is only 1 possible star phylogeny of 3 species). Another significant technical improvement by Gillespie (1989) was the usage of lineage-weighting factors to account for the so-called lineage effect, which refers to species-specific average rate of evolution.

¹ Present address: Division of Biostatistics, School of Medicine, Indiana University, Indianapolis, IN.

Key words: index of dispersion, lineage effects, neutral theory, nearly neutral theory, comparative genomics, gene ontology.

E-mail: soojinyi@gatech.edu.

Mol. Biol. Evol. 25(4):634–642, 2008

doi:10.1093/molbev/msm281

Advance Access publication December 19, 2007

Gillespie (1989) then analyzed 20 proteins from 3 mammalian species and estimated that R for nonsynonymous sites was 6.75, much larger than the neutral expectation of 1. Gillespie’s (1989) R for synonymous sites was 4.64, considered only marginally different from 1. Thus, it was concluded that in mammals, nonsynonymous sites evolve nonneutrally, in “episodic” fashion, whereas synonymous sites were only marginally deviant of the neutral pattern (Gillespie 1989, 1991). This view has since remained authoritative and is commonly reiterated in literature (p. 231 in Li 1997; Cutler 2000; Wilke 2004).

Ohta (1995) analyzed a larger data set (49 mammalian proteins) and observed that R for synonymous sites (5.89) was comparable with, if slightly greater than, that for nonsynonymous sites (5.60). Interestingly, Zeng et al. (1998) showed that in *Drosophila*, R for nonsynonymous sites was 1.64, much less than that in mammalian proteins. In contrast, they found that R for synonymous sites was 4.37, significantly greater than 1. Zeng et al. (1998) interpreted that different evolutionary forces affect synonymous site evolution in *Drosophila* and mammals.

Sequencing of several mammalian genomes brought an opportunity to examine a genome-wide pattern of variance of evolutionary rates among lineages. Methods to correct for multiple hits have also significantly improved, and new analytical tools to investigate roles of selection and mutation on protein evolution have become available. In addition, theoretical analyses of selective and neutral models of “overdispersion” have provided new insights into the behavior of index of dispersion under a variety of evolutionary models (Gillespie 1993; Cutler 2000; Bastolla et al. 2003; Wilke 2004). Our study is motivated by these advances, aiming to gain a comprehensive understanding on the variance of the number of nucleotide substitutions in mammalian proteins.

Here, we further improve upon Gillespie’s methods of estimating the index of dispersion in mammalian proteins by 1) using a codon-substitution model to correct for multiple hits, 2) deriving a statistically superior estimate of index of dispersion, and 3) using a large number of loci to estimate “lineage effects.” For this purpose, we used genomic data from 3 lineages, primates, carnivores, and rodents.

Our analyses provide genome-wide lineage-weighting factors for each lineage, which clearly represent

generation-time effects. The numbers of nonsynonymous and synonymous substitutions in each lineage are consistent with the nearly neutral theory of molecular evolution. Interestingly, our genome-wide analysis of index of dispersion reveals that the mean index of dispersion for mammalian nonsynonymous sites is substantially lower than previous estimates and not different from one.

Materials and Methods

Sequence Data

We used data from 3 mammalian lineages, Primata (*Homo sapiens*), Rodentia (*Mus musculus*), and Carnivora (*Canis familiaris*). We obtained known gene sequences from these species from Ensembl (<http://www.ensembl.org>). The numbers of sequences from each species are 22,719 for *H. sapiens*, 26,452 for *M. musculus*, and 8,253 for *C. familiaris*. Next, we used the OrthoMCL program (Li et al. 2003) to identify orthologs among 3 mammalian lineages with default parameters except for a more stringent E value of 10^{-10} . This step identified 3,493 orthologs. We first aligned the amino acid sequences of these genes using ClustalW (Thompson et al. 1994) and then translated back to DNA sequence alignment.

Estimating Sequence Divergence

We used a codon-based maximum likelihood method to estimate the numbers of pairwise synonymous and nonsynonymous substitutions, using PAML (Yang 1997). Because estimation of sequence divergence is not reliable when pairwise sequence divergence is much greater than saturation, we constrained on pairwise d_N and d_S to curate the data set before calculating the index of dispersion for each gene.

As the constraint increases the index of dispersion decrease for both synonymous and nonsynonymous sites because the number of substitutions can “limit” the variance of evolutionary rates (Cutler 2000). In the remainder of the paper, we mainly discuss results when we restrict our data to cases when pairwise d_N and d_S are less than 1.5, unless noted otherwise. All qualitative results remained the same when we did not constrain our data or used more stringent criteria of $d_N, d_S < 1$. We then obtained lineage-specific numbers of substitutions from the pairwise estimates, as in Gillespie (1989) and Ohta (1995).

Calculating the Indices of Dispersions

Reliable estimation of sequence divergence is critical for our calculation of the indices of dispersion. As mentioned above, we constrained on pairwise d_N and d_S . In addition, we excluded genes in which the numbers of ungapped aligned sites correspond to a 95 percentile in the overall distribution of the proportion of ungapped aligned sites in the 3 species alignments. This cutoff corresponds to alignments where the numbers of ungapped aligned sites in any pairwise alignment are less than 88%. Employing different cutoff values did not change

our conclusions. The final number of alignments used to calculate the indices of dispersion (with the constraints of pairwise d_N and $d_S < 1.5$) is 2,932.

Gene Ontology

We downloaded gene ontology (GO) identities from Ensembl (<http://www.ensembl.org>) for each gene of *H. sapiens*, *M. musculus*, and *C. familiaris*. We then used the AmiGO database (<http://www.godatabase.org>) to extract the explanation for each GO identity. All GO identities were then categorized according to biological process, cellular component, and molecular function. The expectations of the index of dispersion and the number of synonymous and nonsynonymous substitutions of each GO identity were calculated.

An Improved Measure of Index of Dispersion, R_b

We improved the commonly used Gillespie’s (1989) measure of index of dispersion. We first obtained the minimum variance unbiased estimator (MVUE) of the mean number of substitution for a particular locus ($\hat{\mu}_b$). We also derived an unbiased estimator of variance of the number of substitutions for a particular locus (σ_b^2) and prove that it has less variance than that of Gillespie’s estimator. Therefore, we can improve Gillespie’s expression R_g to $R_b = \frac{\sigma_b^2}{\hat{\mu}_b}$, which has lower variance than Gillespie’s (1989) statistic. Details are presented in the Appendix.

Results

Numbers of Synonymous and Nonsynonymous Substitutions in 3 Mammalian Lineages and the Genome-Wide Lineage-Weighting Factors

We obtained lineage-specific numbers of nucleotide substitutions for each gene. Figure 1 shows the mean numbers of synonymous and nonsynonymous substitutions per site in the 3 mammalian genomes. The branch lengths are the longest for the mouse (Rodentia), followed by the dog (Carnivora) and the human (Primata), for both synonymous and nonsynonymous sites (fig. 1). Our estimates of lineage-specific branch lengths and pairwise divergences are well in accord with previous estimates (Makalowski and Boguski 1998; Mouse Genome Sequencing Consortium 2002; Lindblad-Toh et al. 2005).

The order of lineage-specific branch lengths is inversely correlated with the lengths of generation time in each lineage, supporting generation-time effect; current estimates of generation times of human, dog, and mouse are 25, 4, and 0.5 years, respectively (Keightley and Eyre-Walker 2000). Furthermore, the generation-time effect is more conspicuous for synonymous substitutions than for nonsynonymous substitutions (the ratio of branch lengths of Rodentia:Carnivora:Primata is 2.78:1.45:1 for synonymous sites, compared with 2.14:1.17:1 for nonsynonymous sites), suggesting that synonymous sites are more strongly influenced by neutral factors that differ among lineages, such as mutation rates.

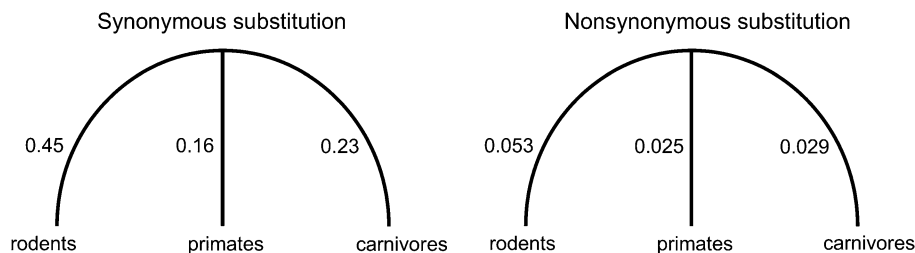


FIG. 1.—The star phylogeny of the 3 lineages analyzed in this study. Branch lengths are the estimated numbers of substitution per site, calculated using a maximum likelihood method based on a codon-substitution model using PAML.

Table 1 shows the lineage-weighting factors and the average numbers of nonsynonymous and synonymous substitutions per lineage. The lineage-weighting factors are calculated by dividing the number of lineage-specific substitutions by the mean number of substitutions in all 3 lineages, to account for the genome-wide differences in branch lengths between lineages (Gillespie 1989). Previous studies of index of dispersion in mammals (Gillespie 1989; Ohta 1995) used Artiodactyla instead of Carnivora for comparisons. It appears that Artiodactyla and Carnivora are both intermediate of rodents and primates in terms of their relative branch lengths and the resulting lineage effects. For example, among 49 genes, the lineage-weighting factors for Primata:Rodentia:Artiodactyla were 0.61:1.58:0.82 for synonymous sites (0.75:1.28:0.97 for nonsynonymous sites) (Ohta 1995). These numbers are similar to those obtained from the current analyses (table 1).

Genome-Wide Index of Dispersion (R) in Mammals

Now we examine the genome-wide indices of dispersion in mammals. We calculated the indices of dispersion using 3 different methods, Kimura's (1983) method (R_k), Gillespie's (1989) method (R_g), and our new method (R_b), which improves upon Gillespie's method (see Materials and Methods and Appendix for details). Table 2 is re-analysis of the 20 loci used in Gillespie's (1989) analysis. Overall, Gillespie's estimator and the new estimator appear similar, whereas Kimura's estimator gives the largest dispersion. This is due to the fact that Kimura's measure does not correct for lineage effects (see below).

We present genome-wide mean indices of dispersion for synonymous and nonsynonymous sites obtained by these 3 methods in table 3. Indices of dispersion when different levels of constraints were used are also presented for comparison. Note that R_k , which does not correct for lineage-specific weights, always shows the largest dispersion. As we correct for lineage effect using lineage-weighting factor (i.e., in case of R_g and R_b), indices decrease. The behaviors of R_g and R_b are similar. As expected, when

Table 1
Lineage Weights (Numbers of Substitutions) Estimated from Genome-Wide Comparison of 3 Species

	Human	Mouse	Dog
Synonymous	0.55 (47.37)	1.65 (142.16)	0.80 (69.29)
Nonsynonymous	0.67 (22.13)	1.51 (49.62)	0.82 (26.80)

we constrain data to allow smaller pairwise d_N and d_S , indices decrease.

Our new estimates of the index of dispersion are 8.04 for synonymous sites and 4.42 for nonsynonymous sites, for the data set in which pairwise d_N and d_S were less than 1.5. All results were qualitatively similar when data obtained under different constraints were used or when we considered genes with greater than 100 synonymous sites only, to avoid errors introduced by short genes (results not shown).

We tested whether the estimates of indices are significantly different from the neutral expectation of 1. Kimura (1983) suggested using a chi-square distribution for this purpose. Gillespie (1989) showed, by simulation, that chi-square approximation is appropriate for nonsynonymous sites but for synonymous sites. Therefore, we used critical values obtained in Gillespie (1989) for our test (table 4). According to these critical values, the new measure of index of dispersion for nonsynonymous sites (referred to as R_{bn}) is

Table 2
Comparison of the 3 Methods of Estimating the Index of Dispersion (R), Using 20 Loci as Included in table 1 of Gillespie (1989)

Locus	Synonymous			Nonsynonymous		
	R_k	R_g	R_b	R_k	R_g	R_b
Prolactin	10.33	1.02	1.12	26.94	12.21	11.90
Parathyroid	16.96	4.58	4.07	8.03	3.47	3.63
Proenkephalin	30.34	9.05	6.93	8.57	3.73	3.43
Proglucagon	16.99	9.39	7.02	1.58	2.13	2.08
α -globin	18.21	4.68	4.17	0.23	0.32	0.36
β -globin	4.72	0.73	0.80	7.30	3.41	3.17
Thyrotropin, β	6.77	4.89	4.11	2.80	2.26	1.97
POMC	9.84	1.46	1.21	15.10	6.44	6.74
Growth hormone	3.45	17.10	20.20	32.22	43.82	46.81
GPHA	1.00	2.52	3.73	22.43	27.74	30.04
Luteinizing, β	11.74	2.16	1.68	3.21	6.55	6.72
Relaxin	13.08	0.28	0.24	3.38	0.13	0.12
Interleukin-2	57.45	17.19	16.01	12.80	8.85	7.75
Signal peptides	18.11	7.52	6.41	9.98	3.25	3.19
CCK	5.00	1.20	1.16	2.36	0.31	0.33
ACHRG	13.40	3.71	4.60	1.52	0.61	0.57
UPA	21.38	0.25	0.22	0.01	5.11	5.95
ANF	10.05	2.17	1.80	1.94	1.64	1.50
β -crystallin	5.40	0.36	0.48	1.38	2.50	2.65
Na, K-ATPase, β	13.95	2.48	2.05	3.49	4.46	4.24
Average	14.41	4.64	4.40	8.26	6.95	7.16

NOTE.— R_k : Kimura's (1983) method, R_g : Gillespie's (1989) method, and R_b : New method, POMC: proopiomelanocortin, CCK: cholecystokinin, ACHRG: acetylcholine receptor gamma subunit, UPA: urokinase-type plasminogen activator, ANF: atrial natriuretic factor.

Table 3
Mean Index of Dispersion for Synonymous and Nonsynonymous Sites

	Nonsynonymous Sites			Synonymous Sites		
	R_k	R_g	R_b	R_k	R_g	R_b
Full data set	11.52	4.60	4.94	38.57	9.64	10.25
$d_N, d_S < 1.5$	10.88	4.08	4.42	35.87	7.69	8.04
$d_N, d_S < 1.0$	10.67	3.48	3.76	34	5.98	6.27

NOTE.—Values discussed mainly in the text are shown in bold and italic.

significantly different from 1 for 960 (at 5% significance level) genes out of 2,932 genes in the whole-data set. In comparison, index of dispersion for synonymous sites (referred to as R_{bs}) is significantly different from 1 for 875 (at 5% significance level) genes. Index of dispersion for synonymous sites ($R_{bs} = 8.04$) is larger than those of Gillespie (1989, $R = 4.6$) and Ohta (1995, $R = 5.89$). With increasing constraint, new estimate of index of dispersion for synonymous sites approaches that in Ohta (table 3). At 1% significance level, mean indices of dispersion for both synonymous and nonsynonymous sites are statistically equivalent to 1.

The index of dispersion is greater for synonymous sites than for nonsynonymous sites, the opposite of the previously reported pattern in mammals (Gillespie 1989) and qualitatively similar to the pattern observed in *Drosophila* (Zeng et al. 1998). In the least, we do not observe the pattern that was concluded by the influential analysis of Gillespie (1989), that nonsynonymous sites show clear overdispersion whereas synonymous sites evolve neutrally.

It should be noted that correcting for multiple hits introduces another source of variance for estimating index of dispersion (Bulmer 1989; Gillespie 1989; Ohta 1995; Nielsen 1997). Hence, we consider our estimates as overestimates compared with true values. It follows that our conclusion that mammalian nonsynonymous sites are not overdispersed is a conservative one (see Discussion). In comparison, index of dispersion for synonymous sites may be substantially affected by errors due to multiple hit correction.

Genome-Wide Factors on Mammalian Index of Dispersion

Overdispersion of protein molecular evolution in mammals has been attributed to effects of genetic drift and natural selection. Gillespie (1989) emphasized the role of natural selection and changing environments on overdispersion of protein molecular clock. In other words, episodic positive selection will inflate variance of evolutionary rates. On the other hand, strong purifying selection can suppress variance of evolutionary rates.

The opposing effects of positive and negative natural selection on variance of evolutionary rates predict that index of dispersion and selective constraint (measured as the ratio of rates of nonsynonymous substitution to synonymous substitutions, d_N/d_S) should be correlated. Indeed, we found that the index of dispersion for nonsynonymous sites is strongly positively correlated with d_N/d_S (Spearman's rank correlation coefficient ρ between R_{bn} and $d_N/d_S = 0.32$, $P < 10^{-4}$). This relationship did not

Table 4
Cutoff values for Significance of the Indices of Dispersion

Significance Level	χ^2 -test (degree of freedom = 2) ^a	Simulation ^b	
		Synonymous	Nonsynonymous
1%	4.605	12.15	4.90
5%	2.991	7.50	3.20

^a Kimura (1983) suggested that $2R$ follows a chi-square distribution with $df = 2$.

^b Gillespie (1989) used a simulation method to derive cutoff values.

change when we changed the constraint in the data set. In contrast, the correlation between R_{bs} and d_N/d_S was not significant, suggesting that natural selection on amino acid sequence is not a strong determinant of variance of synonymous rates.

We showed that for approximately 60% of genes, variance of nonsynonymous rates can be explained by statistical fluctuations under a simple Poisson model. It suggests that the majority of mammalian proteins are under purifying selection that is strong enough to suppress variance of evolutionary rates caused by other factors. This can be further demonstrated by the observation that average selective constraint for the genes whose R_{bn} is significantly greater than 1 (at 1% significance level) is weaker than in the rest of the data. For example, between human and mouse, average d_N/d_S of genes whose $R_{bn} > 1$ is 0.173, whereas the mean selective constraint is 0.127 for the whole data. In comparison, average selective constraint for genes whose R_{bn} is not significantly different from 1 is 0.114. The difference between mean selective constraints of 2 groups of genes (divided by whether R_{bn} is greater than 1 or not) is highly significant in all 3 pairwise species comparisons ($P < 10^{-4}$ in all comparisons). Furthermore, we show in the next section that genes encoding essential functions such as transcription regulation tend to exhibit lower index of dispersion than proteins in other functional categories, which attest the effect of purifying selection to suppress variance of nonsynonymous rates.

According to the neutral models of overdispersion (Takahata 1987; Ohta 1992), the numbers of substitutions among lineages can vary above the Poisson expectation due to factors such as varying effective population sizes, weakly deleterious mutations, and "fluctuating neutral space." For example, in Ohta's (1992, 1995) theory, different lineages may experience different rates of slightly deleterious substitutions, if their effective population sizes differ. According to this hypothesis, a significant correlation between the mean number of substitutions and the index of dispersion is expected (Ohta 1995; Zeng et al. 1998). This relationship is expected to be much stronger when an unweighted measure of index of dispersion (e.g., R_k) is used.

Indeed, R_k and the mean number of substitutions are strongly correlated in our data (correlation coefficient greater than 0.6 for both synonymous and nonsynonymous sites). When we use weighted measures of index of dispersion (i.e., R_g or R_b), we still find strong correlation between the mean number of substitutions and the index of dispersion for both synonymous (e.g., in case of R_b , Spearman's rank correlation coefficient $\rho = 0.33$, $P < 10^{-4}$) and nonsynonymous ($\rho = 0.39$, $P < 10^{-4}$) sites. This suggests

Table 5
Spearman's Rank Correlations between Lineage-Specific Branch Lengths (d_S or d_N) and Indices of Dispersion

	Correlation between d_S and R_{bs}			Correlation between d_N and R_{bn}		
	Human	Dog	Mouse	Human	Dog	Mouse
$d_N, d_S < 1.5$	0.11	0.30	0.19	0.30	0.30	0.32
No constraint	0.15	0.33	0.21	0.31	0.30	0.32

NOTE.—All correlations are significant ($P < 10^{-4}$).

that lineage-specific factors for each gene (even after removing the genome-wide lineage effects) are significant determinants of rate variation among lineages, for both synonymous and nonsynonymous sites.

Lineage-specific branch lengths (as compared with the mean number of substitutions analyzed above) and the indices of dispersion should be also correlated according to this hypothesis (Zeng et al. 1998). Consistent with this idea, all 3 lineages showed significant correlation between d_N or d_S and R_b (table 5). We also observe a significant correlation between R_{bs} and R_{bn} ($\rho = 0.1790, P < 10^{-4}$). Hence, synonymous and nonsynonymous substitutions in each protein are evolving under similar evolutionary forces. As Gillespie (1989) pointed out, this may reflect gene-specific mutation rates.

In conclusion, we find evidence that variance in the numbers of nucleotide substitutions in nonsynonymous sites among the 3 mammalian lineages is caused by positive and negative selection, lineage-specific effects, and mutation rates. Variance in the numbers of synonymous substitutions can be partially explained by gene-specific mutation rates and potential lineage-specific effects.

Relationship between Index of Dispersion and GO

Recent efforts to classify proteins according to their functions (The Gene Ontology Consortium 2000) have advanced our understanding of specific selective forces acting

on different functional classes of genes (e.g., The Chimpanzee Sequencing and Analysis Consortium 2005). Index of dispersion may vary according to functional classifications of proteins because the types and the strength of natural selection may vary in different functional categories or that proteins encoding-specific functions are subject to different neutral factors. To test this hypothesis, we divided our data into different GO categories. We chose GO terms with more than 10 genes per GO and examined indices of dispersion for synonymous and nonsynonymous sites. The results are shown in table 6.

Proteins belonging to different GO terms exhibit different degrees of dispersion in nonsynonymous sites. Among the GO terms examined, GO0006508 (proteolysis) and GO0005179 (hormone activity) show the greatest dispersion (6.95 and 8.11, respectively). In contrast, GO0006355 (regulation of transcription) and GO0003700 (transcription factor activity) exhibited little variation in evolutionary rates among lineages ($R_{bn} = 2.15$ and 2.18, respectively). Synonymous sites showed relatively little difference among GO terms, with the exception of the GO0005634 (nucleus), which exhibited a large index of dispersion ($R_{bs} = 15.96$).

We tested whether we may observe such high and low values of index of dispersion by chance, by randomly choosing the same number of genes from the total sample and calculating the index of dispersion for that group. We repeated this procedure 1,000 times. We found that the index of dispersion for nonsynonymous sites for the GO0006508 (proteolysis) was marginally significant by this analysis ($P = 0.09$). Similarly, the average index of dispersion of the GO0005179 (hormone activity) is greater than obtained by this bootstrapping approach at 8% level. In other words, genes belonging to the GO0005179 tend to show overdispersion of nonsynonymous sites compared with the rest of the genome.

We propose that different degrees of overdispersion for proteins in different functional categories can explain the qualitative discrepancy between our results and that

Table 6
Mean Index of Dispersion for Proteins in Different GO Terms (Number of Proteins Analyzed)

GO ID	Synonymous Sites		Nonsynonymous Sites		Terms
	R_g	R_b	R_g	R_b	
Biological processes					
GO:0007186 (19)	6.86	6.37	3.88	4.56	G-protein coupled receptor signaling pathway
GO:0006508 (13)	6.93	6.20	6.31	6.95	Proteolysis
GO:0006355 (12)	6.37	7.38	1.84	2.15	<i>Regulation of transcription</i>
GO:0006810 (10)	5.31	5.54	3.57	4.09	Transport
Cellular component					
GO:0016021 (37)	5.44	5.16	3.62	4.01	Integral to membrane
GO:0016020 (31)	7.81	7.81	5.45	6.06	Membrane
GO:0005634 (27)	11.92	15.96	3.34	3.47	Nucleus
Molecular function					
GO:0004872 (24)	4.93	5.21	4.40	4.92	Receptor activity
GO:0001584 (19)	6.98	6.56	3.75	4.08	Rhodopsin-like receptor activity
GO:0003700 (13)	5.94	6.93	1.89	2.18	<i>Transcription factor activity</i>
GO:0005524 (12)	8.17	7.03	2.82	2.85	Adenosine triphosphate binding
GO:0005179 (10)	4.28	4.14	7.08	8.11	<i>Hormone activity</i>
GO:0016740 (10)	6.03	5.72	3.42	4.42	Transferase activity

NOTE.—Only those GO terms with more than 10 proteins are analyzed. GO terms with particularly low or high nonsynonymous index of dispersion (as mentioned in the main text) are shown in italic and bold.

of Gillespie (1989), in that he concluded that mammalian nonsynonymous sites were significantly overdispersed. In Gillespie's data set of 20 genes, 4 proteins (prolactin, parathyroid, glycoprotein hormone alpha subunit [GPHA], and growth hormone) belong to the GO0001579 (hormone activity) (see table 2). Hence, his estimate of R may have been inflated due to functional bias in the data. If we exclude those 4 genes from his data set, the average R for nonsynonymous sites becomes 3.02, only marginally different from 1 and similar to our genome-wide estimate. Zeng et al. (1998) also noticed that Gillespie's data set contained many protein hormones and proposed that sampling bias may have inflated the estimate of index of dispersion in mammalian nonsynonymous sites.

According to our bootstrapping analysis, synonymous sites of proteins in the GO0005634 (nucleus) exhibit significant ($P = 0.02$) overdispersion. This observation is intriguing and demands future investigation of causes of synonymous rate variation in mammals (see Discussion).

Discussion

Analysis of index of dispersion has served at least 2 important functions. First, the usage of lineage-weighting factors has been informative to understand patterns and causes of lineage-specific evolutionary rates. In the current study, we provide updated estimates of the lineage factors and the genome-wide index of dispersion, obtained using a large number of loci from 3 mammalian genomes and improving upon previous statistical methods. Our measures of genome-wide lineage-weighting factor support the generation-time effect; species with longer generation time tend to have smaller lineage-weighting factor. Generation-time effect is more prominent in synonymous sites than in nonsynonymous sites, consistent with the idea that synonymous sites are under greater influence of neutral force, such as mutation rates.

We observe that selection strength has been reduced in the primate lineage compared with the other 2 mammalian lineages. The mean d_N/d_S is larger in primate lineage (0.185) than it is in rodents (0.118) or in carnivores (0.136). This is consistent with the idea that reduced effective population size (N_e) can promote fixation of weakly deleterious mutations (Ohta 1992, 1997). Even though the ancestral effective population sizes of different mammalian lineages are not known, current effective population size of rodents is considered to be in the range of 10^5 (Ideraabdullah et al. 2004), whereas in primates it is in the order of 10^4 's (Takahata 1993; Harding et al. 1997; Chen and Li 2001). The N_e of carnivores is less well understood, although some studies have indicated N_e of several 10^4 's (Nei and Graur 1984; Spong et al. 2000). Therefore, reduction of N_e in primates may have accelerated fixation of mildly deleterious amino acid substitutions.

Second, different values of indices of dispersion for synonymous and nonsynonymous sites were taken as evidence of fundamental difference in the underlying evolutionary mechanisms on them. In Gillespie's (1989) influential analysis of mammalian proteins, synonymous sites appeared to have index of dispersion not different from 1, whereas nonsynonymous sites had greater than expected

index of dispersion under the neutral model. Based upon this observation, it has been often cited that in mammals, synonymous sites are mostly evolving in neutral fashion, whereas nonsynonymous sites are overdispersed due to nonneutral forces (Gillespie 1989, 1991; Li 1997; Cutler 2000; Wilke 2004).

Contrary to the aforementioned view, we here show that mammalian nonsynonymous sites are on the whole not overdispersed as previously thought. Genes subject to stronger purifying selection (as measured by d_N/d_S) tend to have lower index of dispersion, suggesting that negative (purifying) selection is an important force to reduce variance of evolutionary rates among lineages. In addition, genes in specific functional categories as defined by GO terms exhibit different among-lineage variation of evolutionary rates. For example, genes involved in regulation of transcription tend to show lower index of dispersion, whereas protein hormones tend to have large index of dispersion (table 6). In particular, earlier estimates may have been biased because they included analyses of protein hormones, which tend to exhibit greater variation among lineages than the rest of the genome.

Many protein hormones exhibit lineage-specific episodic molecular evolution in mammals (Wallis 2001; Maston and Ruvolo 2002; Opazo et al. 2005; Yi and Li 2007). The observation that protein hormones generally show significant overdispersion is therefore consistent with Gillespie's (1986, 1989, 1991) proposal that bursts of adaptive evolution can cause overdispersion. However, even though this hypothesis is well in accord with the data presented here, there are some theoretical problems with this proposal. Namely, to cause an overdispersion, the rate of environmental fluctuation should be roughly the same as the rate of a substitution (Gillespie 1993; Cutler 2000), which is estimated to be in millions of years for a typical nonsynonymous change in mammals (Smith and Eyre-Walker 2003), appearing too slow compared with the perceived timescale of ecological habitat changes (see Smith and Eyre-Walker [2003] for alternative possibilities).

Estimates of index of dispersion can be inflated due to errors associated with correction for multiple hits, even with sophisticated evolutionary models (Bulmer 1989; Goldman 1994; Nielsen 1997; Yang and Nielsen 1998). Given that the number of nonsynonymous substitutions per site is generally much below saturation within mammals, multiple hit correction is not likely to be a significant source of error for nonsynonymous sites. For example, Gillespie (1989) showed, by simulation, that in case of Jukes-Cantor method, the extent of inflation of the index of dispersion due to multiple hit correction is only approximately $\sim 10\%$ for nonsynonymous sites. Note that variance due to multiple hit correction will make our main conclusion (that the mean index of dispersion for mammalian nonsynonymous sites is not significantly different from 1) a conservative one. However, for synonymous sites, having generally undergone much greater number of substitutions than nonsynonymous sites, errors associated with multiple hit correction can inflate index of dispersion substantially.

Thus, we refrain from overinterpreting our observation that index of dispersion is generally greater for synonymous sites than nonsynonymous sites. Nevertheless, it should be

noted that a variety of factors that affect synonymous rates in lineage- and gene-specific manner have been discovered in recent years. For example, neutral evolutionary rates can vary greatly due to different CpG contents (Kim et al. 2006) or changes in recombination rates (Perry and Ashworth 1999; Montoya-Burgos et al. 2003; Meunier and Duret 2004; Yi and Li 2005). Synonymous sites may be under selective constraint related to mechanisms such as RNA editing, microRNA binding, and conservation of splice signals (Chamary et al. 2006). Remarkably, a synonymous polymorphism in the human MDR1 locus has been shown to change the protein’s biochemical properties, in this case its ability to pump certain molecules (Kimchi-Sarfaty et al. 2007). Thus, synonymous sites are potentially subject to positive selection. If true, it has significant implications on several aspects of molecular evolutionary analyses. For example, the assumption of molecular clock for synonymous sites in evolutionary inference or utilizing synonymous rates as a reference to detect positive selection need to be employed with caution. Analyses of proteins with significantly overdispersed synonymous sites may offer insights on these issues.

Conclusions

As stated above, our study refutes the canonical view that mammalian nonsynonymous sites exhibit episodic evolution. One caveat of our study is that we considered genes that remained single copy since the divergence of the 3 mammalian lineages. Lineage-specific gene duplication occurs frequently, and they may acquire new function and undergo episodic molecular evolution.

Several possible causes of overdispersion for synonymous and nonsynonymous sites are discussed. The fact that different functional categories of proteins can exhibit different degrees of among-lineage variation of rates attests the role of natural selection on dispersion of evolutionary rates. Index of dispersion may be useful in determining selective and neutral factors that lead to episodic, positive selection for both synonymous and nonsynonymous sites of mammalian proteins.

Acknowledgments

We thank the members of the Yi laboratory, Dr. Koichiro Tamura, and 2 anonymous reviewers for valuable comments. Dr. Tomoko Ohta has kindly provided comments and pointed out a mistake in an earlier version of the manuscript. This study is supported by funds from the Georgia Institute of Technology.

Appendix

Statistical Properties of the New Measure of Index of Dispersion, R_b

Let the number of substitutions at the m -th locus on the i -th lineage be $N_{m,i}$. Assuming that the $N_{m,i}$ ’s are independent random variables, we will write their moments in a form that allows the removal of lineage effects: $N_{m,i} \sim \text{Poisson}(w_i \lambda_{m,i})$, $E(N_{m,i}) = w_i \lambda_{m,i}$, $\sum_{i=1}^n w_i = n$, with

$$\text{probability mass function (pmf)} \quad p(N_{m,i}) = \frac{e^{-w_i \lambda_{m,i}} (w_i \lambda_{m,i})^{N_{m,i}}}{(N_{m,i})!}; \quad N_{m,i} = 0, 1, \dots; \quad 0 \leq \lambda_{m,i} < \infty.$$

We now consider testing the null hypothesis $H_0 : N_{m,i} \sim \text{Poisson}(w_i \lambda_{m,i})$, $\lambda_{m,i} = \mu_m$ for all i , versus the alternative hypothesis $H_1 : N_{m,i} \sim \text{Poisson}(w_i \lambda_{m,i})$, $\sum_{i=1}^n (\lambda_{m,i} - \bar{\lambda})^2 > 0$, where $\bar{\lambda} = \frac{1}{n} \sum_{i=1}^n \lambda_{m,i}$. For brevity, we substitute $y_{m,i}$ for $\frac{N_{m,i}}{w_i}$ in all expression above. Then $H_0 : \lambda_{m,i} = \mu_m$ for all i , versus $H_1 : \sum_{i=1}^n (\lambda_{m,i} - \bar{\lambda})^2 > 0$, where $\bar{\lambda} = \frac{1}{n} \sum_{i=1}^n \lambda_{m,i}$ and $y_{m,1}, y_{m,2}, \dots, y_{m,n}$ are random variables with pmf $p(y_{m,i}) = \frac{e^{-w_i \lambda_{m,i}} (w_i \lambda_{m,i})^{w_i y_{m,i}}}{(w_i y_{m,i})!}; \quad E(y_{m,i}) = \lambda_{m,i}; \quad \sum_{i=1}^n w_i = n.$

A New Estimator of Mean

We aim to find the MVUE of the mean number of substitution for a particular locus ($\hat{\mu}_m$) under H_0 . Because $\lambda_{m,i} = \mu_m$ for all i , we have the log-likelihood function

$$\begin{aligned} \ln p(y_{m,1}, \dots, y_{m,n} | H_0) &= (\ln \mu_m) \sum_{i=1}^n w_i y_{m,i} - \sum_{i=1}^n w_i \mu_m + \sum_{i=1}^n w_i y_{m,i} \ln(w_i) \\ &\quad - \sum_{i=1}^n \ln\{(w_i y_{m,i})!\}. \end{aligned}$$

Because we have been an exponential family, we obtain

$$E_{\mu_m} \left(\left(\frac{\partial}{\partial \mu_m} \ln p(y_{m,1}, \dots, y_{m,n} | \mu_m) \right)^2 \right) = \frac{n}{\mu_m}. \quad (1)$$

using Lemma 7.3.1 on p. 312 in Casella and Berger (1990).

As a first speculation, consider the stastic $W(y_{m,1}, \dots, y_{m,n}) = \frac{1}{n} \sum_{i=1}^n w_i y_{m,i}$. The pmf of $W(y_{m,1}, \dots, y_{m,n})$ is $p(y_{m,i})$ with $E(y_{m,i}) = \lambda_{m,i}$ and $\sum_{i=1}^n w_i = n$. Thus,

$$E_{\mu_m} (W(y_{m,1}, \dots, y_{m,n})) = \frac{1}{n} \sum_{i=1}^n w_i E_{\mu_m}(y_{m,i}) = \mu_m. \quad (2)$$

Therefore, $W(y_{m,1}, \dots, y_{m,n})$ is an unbiased estimator of μ_m . We next calculate

$$\begin{aligned} \text{Var}_{\mu_m} (W(y_{m,1}, \dots, y_{m,n})) &= \frac{1}{n^2} \sum_{i=1}^n w_i^2 \text{Var}(y_{m,i}) = \frac{\mu_m}{n} \left(\text{Because } \text{Var}(y_{m,i}) = \frac{\mu_m}{w_i} \text{ under } H_0 \right). \end{aligned} \quad (3)$$

According to the equation (2), we find $\frac{d}{d\mu_m} E_{\mu_m} (W(y_{m,1}, \dots, y_{m,n})) = \frac{d}{d\mu_m} \mu_m = 1$. Then, by the Cramer–Rao theory (theorem 7.3.1, p. 308, Casella and Berger, 1990) and equations (1) and (3), $\text{Var}_{\mu_m} (W(y_{m,1}, \dots, y_{m,n})) \geq \frac{\mu_m}{n}$.

Therefore, $W(y_{m,1}, \dots, y_{m,n})$ is a MVUE, that is, a best unbiased estimation. Furthermore, we can infer that $W(y_{m,1}, \dots, y_{m,n})$ is the unique MVUE because $\sum_{i=1}^n w_i y_{m,i}$ is a complete sufficient statistic.

Alternatively, we can directly see that the variance of the new estimator of the mean number of substitution for a particular locus under H_0 is less than Gillespie's (1989) estimator. We know that Gillespie's estimator of the mean number is $\hat{\mu}_g = \frac{1}{n} \sum_{i=1}^n \frac{N_{m,i}}{w_i}$ and the new estimator is $\hat{\mu}_b = \frac{1}{n} \sum_{i=1}^n N_{m,i}$ because $y_{m,i} = \frac{N_{m,i}}{w_i}$. The difference $Var(\hat{\mu}_g) - Var(\hat{\mu}_b)$ of variances of 2 estimators is $\frac{\mu_m}{n^2} \left(\sum_{i=1}^n \frac{1}{w_i} - n \right)$. We then have $Var(\hat{\mu}_g) - Var(\hat{\mu}_b)$ from the fact that the arithmetic mean is greater than or equal to the harmonic mean.

A New Estimator of Variance

We would next like to find the estimator of the variance that has less variance than that of Gillespie's estimator. Let $\hat{\sigma}_g^2 = \left(\frac{n^2}{n-1} \left(\sum_{i=1}^n \frac{1}{w_i} \right)^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \left(\frac{N_{m,i}}{w_i} \right)^2 - \hat{\mu}_g^2 \right\} \right)$ be Gillespie's estimator of the variance and $\hat{\sigma}_b^2 = \left(\frac{\sum_{i=1}^n w_i}{n-1} \left\{ \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n \frac{N_{m,i}^2}{w_i} - \hat{\mu}_b^2 \right\} \right)$ the new estimator of the variance, where $\sum_{i=1}^n w_i = n$, $w_i > 0$ for all i , $E(N_{m,i}) = w_i \mu_m$, and $Var(N_{m,i}) = w_i \sigma_m^2$.

First, we look for the mean of the new estimator of the variance. The mean of $\hat{\sigma}_b^2$ is $E(\hat{\sigma}_b^2) = \frac{1}{n-1} \sum_{i=1}^n \frac{E(N_{m,i}^2)}{w_i} - \frac{n}{n-1} \left(\frac{\sigma_m^2}{n^2} \sum_{i=1}^n w_i + \mu_m^2 \right) = \sigma_m^2$, revealing that the new estimator is an unbiased estimator of the variance.

Hence, the difference of 2 variances above is

$$Var(\hat{\sigma}_g^2) - Var(\hat{\sigma}_b^2) = \frac{4\mu_m^3}{(n-1)^2 n^2} \left(\sum_{i=1}^n \frac{1}{w_i} \right)^{-1} - A + \mu_m^2 B + \mu_m C,$$

where

$$A = n^2(n-1)(n-2) - 6 \left(\sum_{i=1}^n \frac{1}{w_i} \right) \sum_{i < j < k} w_i w_j w_k,$$

$$B = \frac{2(n-2)}{n-1} - \frac{4}{\sum_{i=1}^n \frac{1}{w_i}} \left(\frac{n(n-2)}{(n-1)^2} \right) \sum_{i < j} \frac{1}{w_i w_j}, \text{ and}$$

$$C = \left(\sum_{i=1}^n \frac{1}{w_i} \right)^{-2} \sum_{i=1}^n \frac{1}{w_i^3} - \frac{1}{(n-1)^2} \left(\sum_{i=1}^n \frac{1}{w_i} \right) + \frac{2n-1}{n(n-1)^2}.$$

Therefore, $Var(\hat{\sigma}_g^2) \geq Var(\hat{\sigma}_b^2)$ because $\mu_m \geq 0$ under H_0 and A , B , and C are greater than or equal to zero.

Literature Cited

- Bastolla U, Porto M, Roman HE, Vendruscolo M. 2003. Connectivity of neutral networks, overdispersion, and structural conservation in protein evolution. *J Mol Evol.* 56: 243–254.
- Bulmer M. 1989. Estimating the variability of substitution rates. *Genetics.* 123:615–619.
- Casella G, Berger RL. 1990. *Statistical inference.* Belmont (CA): Duxbury Press.
- Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet.* 7:98–108.
- Chen FC, Li W-H. 2001. Genomic divergence between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet.* 68:444–456.
- Cutler DJ. 2000. Understanding the overdispersed molecular clock. *Genetics.* 154:1403–1417.
- Gillespie JH. 1986. Natural selection and the molecular clock. *Mol Biol Evol.* 3:138–155.
- Gillespie JH. 1989. Lineage effects and the index of dispersion of molecular evolution. *Mol Biol Evol.* 6:636–647.
- Gillespie JH. 1991. *The causes of molecular evolution.* Oxford (UK): Oxford University Press.
- Gillespie JH. 1993. Substitution processes in molecular evolution I. Uniform and clustered substitutions in a haploid model. *Genetics.* 134:971–981.
- Goldman N. 1994. Variance to mean ratio, $R(t)$, for poisson processes on phylogenetic trees. *Mol Phylogenet Evol.* 3:230–239.
- Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Mouline DS, Clegg JB. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet.* 60:772–789.
- Ideraabdullah FY, de la Casa-Esperon E, Bell TA, Detwiler DA, Magnuson T, Sapienza C, de Villena FP-M. 2004. Genetic and haplotype diversity among wild-derived mouse inbred strains. *Genome Res.* 14:1880–1887.
- Keightley PD, Eyre-Walker A. 2000. Deleterious mutations and the evolution of sex. *Science.* 290:331–333.
- Kim S-H, Elango N, Warden CD, Vigoda E, Yi S. 2006. Heterogenous genomic molecular clocks in primates. *PLoS Genet.* 2:e163.
- Kimchi-Sarfaty C, Oh JM, Kim I-W, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM. 2007. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science.* 315:525–528.
- Kimura M. 1983. *The neutral theory of molecular evolution.* Cambridge (UK): Cambridge University Press.
- Langley CH, Fitch WM. 1974. An examination of the constancy of the rate of molecular evolution. *J Mol Evol.* 3:161–177.
- Li L, Stoeckert CJJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Li W-H. 1997. *Molecular evolution.* Sunderland (MA): Sinauer.
- Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. (46 co-authors). 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature.* 438:803–819.
- Makalowski W, Boguski MS. 1998. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc Natl Acad Sci USA.* 95:9407–9412.
- Maston GA, Ruvolo M. 2002. Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol Biol Evol.* 19:320–335.
- Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. *Mol Biol Evol.* 21:984–990.
- Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. *Trends Genet.* 19:128–130.
- Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 420:520–562.

- Nei M, Graur D. 1984. Extent of protein polymorphism and the neutral mutation theory. *Evol Biol.* 17:73–118.
- Nielsen R. 1997. Robustness of the estimator of the index of dispersion for DNA sequences. *Mol Phylogenet Evol.* 7:346–351.
- Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst.* 23:263–286.
- Ohta T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J Mol Evol.* 40:56–63.
- Ohta T. 1997. Role of random genetic drift in the evolution of interactive systems. *J Mol Evol.* 44:S9–S14.
- Opazo JC, Palma RE, Melo F, Lessa EP. 2005. Adaptive evolution of the insulin gene in Caviomorph rodents. *Mol Biol Evol.* 22:1290–1298.
- Perry J, Ashworth A. 1999. Evolutionary rate of a gene affected by chromosomal position. *Curr Biol.* 9:987–989.
- Smith NG, Eyre-Walker A. 2003. Partitioning the variation in mammalian substitution rates. *Mol Biol Evol.* 20:10–17.
- Spong G, Johansson M, Björklund M. 2000. High genetic variation in leopards indicates large and long-term stable effective population size. *Mol Ecol.* 9:1773–1782.
- Takahata N. 1987. On the overdispersed molecular clock. *Genetics.* 116:169–179.
- Takahata N. 1993. Allelic genealogy and human evolution. *Mol Biol Evol.* 10:2–22.
- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature.* 437:69–87.
- The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet.* 25:25–29.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Wallis M. 2001. Episodic evolution of protein hormones in mammals. *J Mol Evol.* 53:10–18.
- Wilke CO. 2004. Molecular clock in neutral protein evolution. *BMC Genetics.* 5:25.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.
- Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol.* 46:409–418.
- Yi S, Li W-H. 2005. Molecular evolution of recombination hotspots and highly recombining pseudoautosomal regions in hominoids. *Mol Biol Evol.* 22:1223–1230.
- Yi S, Li W-H. 2007. Episodic molecular evolution of some protein hormones in primates and its implications for primate adaptation. In: Ravosa MJ, Dagosto M, editors. *Primate origins: adaptations and evolution.* New York: Springer. p. 739–773.
- Zeng L, Comeron J, Chen B, Kreitman M. 1998. The molecular clock revisited: the rate of synonymous vs. replacement change in *Drosophila*. *Genetica.* 102–103:369–382.

Koichiro Tamura, Associate Editor

Accepted December 13, 2007