J Mol Evol (2000) 50:56–68
DOI: 10.1007/s002399910007

JOURNAL OF MOLECULAR EVOLUTION

Rates of Conservative and Radical Nonsynonymous Nucleotide Substitutions in Mammalian Nuclear Genes

Jianzhi Zhang*

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park, PA 16802, USA

Received: 6 April 1999 / Accepted: 16 August 1999

Abstract. To understand the process and mechanism of protein evolution, it is important to know what types of amino acid substitutions are more likely to be under selection and what types are mostly neutral. An amino acid substitution can be classified as either conservative or radical, depending on whether it involves a change in a certain physicochemical property of the amino acid. Assuming Kimura's two-parameter model of nucleotide substitution, I present a method for computing the numbers of conservative and radical nonsynonymous (amino acid altering) nucleotide substitutions per site and estimate these rates for 47 nuclear genes from mammals. The results are as follows. (1) The average radical/conservative rate ratio is 0.81 for charge changes, 0.85 for polarity changes, and 0.49 when both polarity and volume changes are considered. (2) The radical/conservative rate ratio is positively correlated with the nonsynonymous/synonymous rate ratio for charge changes or when both polarity and volume changes are considered. (3) Both the conservative/synonymous rate ratio and the radical/synonymous rate ratio are lower in the rodent lineage than in the primate or artiodactyl lineage, suggesting more intense purifying selection in the rodent lineage, for both conservative and radical nonsynonymous substitutions. (4) Neglecting transition/transversion bias would cause an underestimation of both radical and conservative rates and the ratio thereof. (5) Transversions induce more dramatic genetic alternations than transitions in that transversions produce more amino acid altering changes and among which, more radical changes.

Key words: Substitution rate — Conservative change — Radical change — Positive selection — Amino acid property — Transition bias

Introduction

The 20 amino acids can be classified into groups according to their physicochemical properties such as charge, polarity, and volume. Amino acid substitutions within groups are called conservative substitutions whereas those between groups are radical. For example, with respect to the amino acid charge in physiological environments, a substitution from lysine to arginine is called a conservative substitution because the original and resultant amino acids are both positively charged, but a substitution from lysine to isoleucine is radical because the resultant isoleucine is neutral (uncharged). It has been known for a long time that there are more conservative amino acid substitutions than radical substitutions in terms of charge or polarity in protein evolution (Zuckerkandl and Pauling 1965; Epstein 1967; Clarke 1970; Dayhoff et al. 1972). This difference in quantity is usually explained by a higher intensity of purifying selection on radical mutations than on conservative mutations.

**Present address* Jianzhi Zhang, Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Building 10, Room 11N104, 9000 Rockville Pike, Bethesda, MD 20892, USA
Correspondence to: Jianzhi Zhang; *e-mail:* jzhang@niaid.nih.gov

However, the genetic code table is of such a structure that a random mutation is more likely to be conservative than radical if it is nonsynonymous (amino acid altering) (Goldberg and Wittes 1966; Epstein 1966, 1967). Therefore, to investigate whether radical amino acid changes are more likely to be deleterious than conservative changes, one has to consider the underlying pattern of nucleotide mutations.

By assuming the nucleotide mutation (substitution) model of Jukes and Cantor (1969), Hughes et al. (1990) developed a method for computing the rates of conservative and radical substitutions, i.e., the number of conservative nonsynonymous nucleotide substitutions per conservative nonsynonymous site and the number of radical nonsynonymous nucleotide substitutions per radical nonsynonymous site between two homologous gene sequences. In practice, however, the rate of transitional nucleotide substitution is often greater than that of transversional substitution (known as the transition/transversion bias or transition bias), violating the assumption of the Jukes–Cantor model. Neglecting transition bias causes overestimation of the number of nonsynonymous sites (Li 1993; Ina 1995) and is expected to affect the estimation of the conservative and radical substitution rates as well. A method that takes into account the transition bias is therefore preferred.

A significantly higher rate of radical nonsynonymous substitution than conservative substitution has been taken as evidence for positive Darwinian selection on radical substitutions even without an observation of a significantly higher rate of nonsynonymous than synonymous substitution (Hughes 1992, 1994; Hughes and Hughes 1993). This interpretation relies on the assumption that the intensity of purifying selection on radical substitutions is equal to or greater than that on conservative substitutions. The validity of this assumption as a general rule has not been thoroughly examined (Li et al. 1985).

Comparison of the rates of synonymous substitution, conservative (nonsynonymous) substitution, and radical (nonsynonymous) substitution may also shed light on the verification of the nearly neutral theory (Ohta 1992). It has been shown that the nonsynonymous/synonymous rate ratio varies among mammalian orders with a lower value for rodents and higher values for primates and artiodactyls (Ohta 1995). It is therefore interesting to examine the conservative and radical substitution rates of genes from primates, artiodactyls, and rodents in order to see whether the intensities of purifying selection for these two classes of amino acid substitutions vary among the three evolutionary lineages.

In this article, I first extend Hughes and co-workers' (1990) method of computing the conservative and radical substitution rates by considering transition bias. I then compute these rates for 47 nuclear genes of mammals to examine whether the radical substitution rate is smaller than the conservative rate, as generally believed. Finally, I compare the conservative and radical substitution rates of the 47 genes in the primate, rodent, and artiodactyl lineages and discuss the results in the light of the nearly neutral theory.

Classification by charge

Positive: R, H, K
Negative: D, E
Neutral: A, N, C, Q, G, I, L, M, F, P, S, T, W, Y, V

Classification by polarity

Polar: R, N, D, C, Q, E, G, H, K, S, T, Y
Nonpolar: A, I, L, M, F, P, W, V

Classification by polarity and volume

Special: C
Neutal and small: A, G, P, S, T
Polar and relatively small: N, D, Q, E
Polar and relatively large: R, H, K
Nonpolar and relatively small: I, L, M, V
Nonpolar and relatively large: F, W, Y

Fig. 1. Three classifications of amino acids according to certain physicochemical properties.

Data and Methods

Sequence Data

I analyzed DNA sequences of 47 nuclear genes of mammals. For each gene, three orthologous sequences from a primate, an artiodactyl, and a rodent were examined. These data sets were part of the 49 genes originally compiled and analyzed by Ohta (1995). The opsin gene in the original compilation was not used here because the alignment was found to be unreliable (Yang and Nielsen 1998). The interleukin-2 gene was not used because it has apparently been under positive selection (Zhang and Nei, unpublished) and therefore is unsuitable for examining the intensities of purifying selection on conservative and radical substitutions.

Computation of d_C *and* d_R *Between Homologous Sequences Under Kimura's (1980) Model*

In this study, I considered three classifications of amino acids with respect to the (1) amino acid charge, (2) polarity, and (3) polarity and volume, respectively (Fig. 1). In the following, I describe the method using the charge-based classification as an example. The 20 amino acids are classified into three groups, with group I (R, H, K) being positively charged, group II (D, E) being negatively charged, and group III (all other amino acids) being neutral. All amino acid substitutions within groups are referred to as conservative and between groups radical. For the *i*th codon of a DNA sequence *m* codons in length, I first compute the number of conservative nonsynonymous sites (c_i) for the codon, assuming that nucleotide substitutions follow Kimura's (1980) model. Here $c_i = \sum_{j=1}^{3} c_{ij}$, where c_{ij} is the expected number of conservative nonsynonymous changes when a nucleotide change occurs at the *j*th ($j = 1$, 2, or 3) position of the *i*th codon. The number (r_i) of

radical nonsynonymous sites for codon i is computed similarly. As will be illustrated in an example, the computation of c_i and r_i is analogous to the computation of the numbers of synonymous and nonsynonymous sites in the modified Nei–Gojobori method (Nei and Gojobori 1986; Zhang et al. 1998). It is obvious that $c_i + r_i = n_i$, where n_i is the number of nonsynonymous sites for the ith codon. For example, a codon TTT has a n_i of $(3 + 2q)/(1 + q)$, where q is the ratio of transitions to transversions in Kimura's (1980) model (Zhang et al. 1998). For this codon, $c_i = 1 + 0 + 1/(1 + q) = (2 + q)/(1 + q)$, because all possible nucleotide substitutions are conservative nonsynonymous if they occur at the first position of the codon TTT, none are conservative nonsynonymous if they occur at the second position, and transversions at the third position are conservative nonsynonymous. Obviously, $r_i = n_i - c_i = 1$. It is interesting to note that in this case r_i does not depend on q. So, for a TTT codon, when the transition bias is ignored, as by Hughes et al. (1990), c_i is overestimated but r_i is not affected. The c_i and r_i are computed for every codon of the sequence and the total numbers of conservative (C) and radical (R) nonsynonymous sites of the sequence are obtained by $C = \sum_{i=1}^{m} c_i$ and $R = \sum_{i=1}^{m} r_i$, respectively. Obviously, $C + R = N$, where $N = \sum_{i=1}^{m} n_i$ is the total number of nonsynonymous sites for the sequence. It is clear that when there is no transition bias ($q = 0.5$), C and R values computed in the new method are identical to those obtained by Hughes and co-workers' (1990) method. It should be noted that some authors prefer using the transition/transversion rate ratio κ to the transition/transversion ratio q. These two quantities have the relationship of $\kappa = 2q$ under Kimura's model.

The numbers of conservative (C_d) and radical (R_d) nonsynonymous differences between two homologous sequences are computed as described by Hughes et al. (1990). This computation is analogous to the computation of the numbers of synonymous and nonsynonymous differences in Nei and Gojobori's (1986) method. The proportion of conservative nonsynonymous differences and that of radical nonsynonymous differences are then computed as $p_C = C_d/C$ and $p_R = R_d/R$, respectively, where C and R are the average numbers of conservative and radical sites of the two sequences in comparison. When p_C and p_R are small (say, less than 0.3), Jukes–Cantor's formula can be used to correct multiple hits. That is, $d_C = 0.75\ln[1 - (4p_C/3)]$ and $d_R = 0.75\ln[1 - (4p_R/3)]$. In short, d_C and d_R are called the conservative and radical distances between two sequences, respectively, and can be interpreted as the numbers of conservative nonsynonymous substitutions per conservative nonsynonymous site and the number of radical nonsynonymous substitutions per radical nonsynonymous site. The variances of p_C, p_R, d_C, and d_R can be derived by an analogy to Ota and Nei's (1994) formulation. They are

$$\mathrm{Var}(p_C) = \sum_{i=1}^{m} \frac{(c_{di} - p_C c_i)^2}{C^2}$$

$$\mathrm{Var}(p_R) = \sum_{i=1}^{m} \frac{(r_{di} - p_R r_i)^2}{R^2}$$

$$\mathrm{Var}(d_C) = \frac{\mathrm{Var}(p_C)}{\left(1 - \frac{4p_C}{3}\right)^2}$$

$$\mathrm{Var}(d_R) = \frac{\mathrm{Var}(p_R)}{\left(1 - \frac{4p_R}{3}\right)^2}$$

where c_{di} and r_{di} are the numbers of conservative and radical nonsynonymous differences between the two sequences in the ith codon, respectively. Covariances of the distances can also be obtained similarly. It needs to be pointed out that the variances may be underestimated by the above formulas because the variance of q is assumed to be 0 in the computation. This underestimation, however, is likely to be trivial when the numbers of transitional and transversional differences between sequences are not too small.

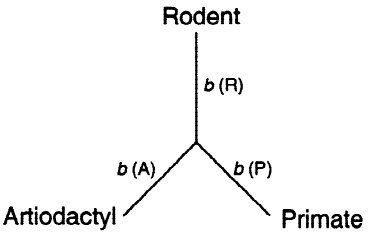

Fig. 2. The phylogeny of primates, artiodactyls, and rodents.

Data Analysis

For the data sets used in this paper, the transition/transversion rate ratio κ for each gene has been estimated by Yang and Nielsen (1998), and these estimates were used in the present analysis. After d_C and d_R are obtained for pairwise comparisons of the primate, artiodactyl, and rodent sequences of a gene, we can compute d_C and d_R for each branch of the phylogenetic tree of these sequences by using the least-squares method (see Rzhetsky and Nei 1993). In the present case, the least-squares solution is rather simple because there are only three sequences. Let us denote primates, artiodactyls, and rodents by P, A, and R, respectively. Let $d_C(\mathrm{PA})$, $d_C(\mathrm{PR})$, and $d_C(\mathrm{AR})$ be the pairwise conservative distances between P and A, P and R, and A and R, respectively, $b_C(\mathrm{P})$, $b_C(\mathrm{A})$, and $b_C(\mathrm{R})$ be the branch lengths in terms of the number of conservative substitutions per site for the three branches linking the interior node and the three present-day sequences, and $b_C(\mathrm{T})$ be the total branch length (Fig. 2). Then

$$\begin{aligned} b_C(\mathrm{P}) &= [d_C(\mathrm{PA}) + d_C(\mathrm{PR}) - d_C(\mathrm{AR})]/2 \\ b_C(\mathrm{A}) &= [d_C(\mathrm{PA}) + d_C(\mathrm{AR}) - d_C(\mathrm{PR})]/2 \\ b_C(\mathrm{R}) &= [d_C(\mathrm{PR}) + d_C(\mathrm{AR}) - d_C(\mathrm{PA})]/2 \\ b_C(\mathrm{T}) &= [d_C(\mathrm{PA}) + d_C(\mathrm{PR}) + d_C(\mathrm{AR})]/2 \end{aligned} \quad (1)$$

The corresponding quantities for radical nonsynonymous substitutions can be similarly derived. Because of stochastic errors, some branch lengths may be estimated to be negative and they are treated as zero in further analysis. Variances of branch lengths can also be estimated from the variances and covariances of pairwise distances. For example,

$$\begin{aligned} \mathrm{Var}[b_C(\mathrm{P})] = {} & \frac{1}{4}\mathrm{Var}[d_C(\mathrm{PA})] + \frac{1}{4}\mathrm{Var}[d_C(\mathrm{PR})] + \frac{1}{4}\mathrm{Var}[d_C(\mathrm{AR})] \\ & + \frac{1}{2}\mathrm{Cov}[d_C(\mathrm{PA}), d_C(\mathrm{PR})] \\ & - \frac{1}{2}\mathrm{Cov}[d_C(\mathrm{PA}), d_C(\mathrm{AR})] \\ & - \frac{1}{2}\mathrm{Cov}[d_C(\mathrm{PR}), d_C(\mathrm{AR})] \end{aligned} \quad (2)$$

Results

Rates of Conservative and Radical Substitutions with Respect to Charge Changes

The numbers (b_C and b_R) of conservative and radical nonsynonymous substitutions per site with respect to

Table 1. Numbers of conservative and radical substitutions per site for the evolutionary lineages of primates (P), artiodactyls (A), rodents (R) and their sums (T): Charge changes are considered

Gene No.	Gene name	Codons	b_C (× 100)				b_R (× 100)				b_R/b_C			
			P	A	R	T	P	A	R	T	P	A	R	T
1	Acetylcholine receptor α	456	1.03	0.80	2.36	4.19	2.03	0.45	0.78	3.26	1.97	0.56	0.33	0.78
2	Acetylcholine receptor β	500	2.58	2.19	4.24	9.00	2.08	2.03	3.09	7.20	0.81	0.93	0.73	0.80
3	Acid phosphatase type 5	322	3.74	5.25	7.22	16.21	1.70	6.02	3.21	10.93	0.45	1.15	0.44	0.67
4	Albumin	606	5.73	10.10	12.05	27.87	5.13	8.42	9.83	23.38	0.90	0.83	0.82	0.84
5	Alkaline phosphatase intestine	495	4.75	9.22	9.32	23.28	4.23	9.67	8.48	22.38	0.89	1.05	0.91	0.96
6	Alkaline phosphatase liver	523	3.45	3.34	4.74	11.53	2.20	2.42	4.17	8.79	0.64	0.73	0.88	0.76
7	Aspartate aminotransferase cytosolic	412	1.97	3.12	2.92	8.01	1.12	2.07	2.86	6.05	0.57	0.66	0.98	0.75
8	Aspartate aminotransferase mitochondrial	429	2.22	2.66	1.67	6.56	0.00	1.44	0.00	1.44	0.00	0.54	0.00	0.22
9	ATP synthase α	543	1.12	0.63	1.06	2.81	0.12	0.36	0.36	0.84	0.11	0.58	0.34	0.30
10	ATP synthase β	357	0.05	0.77	0.78	1.60	0.00	0.00	0.36	0.36	0.00	0.00	0.46	0.22
11	β-1,4-Galactosyl transferase	396	2.96	6.64	6.33	15.93	1.92	6.98	3.94	12.84	0.65	1.05	0.62	0.81
12	Carboxypeptidase	432	0.59	2.54	2.33	5.46	0.00	1.49	1.50	2.99	0.00	0.59	0.64	0.55
13	Connexin	381	0.97	0.38	1.16	2.51	0.00	0.62	0.00	0.62	0.00	1.64	0.00	0.25
14	Corticotropin-releasing factor	182	2.02	13.63	8.56	24.21	2.30	8.95	10.52	21.77	1.14	0.66	1.23	0.90
15	Dopamine receptor D2	442	1.19	0.54	1.42	3.14	0.84	1.19	0.16	2.20	0.71	2.21	0.11	0.70
16	Fibrinogen α	433	4.09	8.93	8.92	21.94	2.05	4.39	4.82	11.26	0.50	0.49	0.54	0.51
17	Glucose transporter	491	0.98	0.71	0.91	2.60	1.12	0.59	0.77	2.48	1.14	0.83	0.85	0.96
18	Growth hormone	189	20.29	5.02	4.95	30.26	16.21	1.53	1.49	19.23	0.80	0.30	0.30	0.64
19	Growth hormone receptor	636	6.07	4.25	14.74	25.06	4.77	1.27	9.96	15.99	0.78	0.30	0.68	0.64
20	Hexokinas I	915	2.49	2.65	2.66	7.80	1.38	3.63	1.27	6.28	0.56	1.37	0.48	0.81
21	IGF binding protein 1	258	10.64	11.87	10.25	32.75	11.37	4.91	6.25	22.52	1.07	0.41	0.61	0.69
22	IGF binding protein 3	287	6.73	9.12	7.23	23.07	2.38	1.85	6.14	10.37	0.35	0.20	0.85	0.45
23	Insulin-like growth factor 1	114	0.66	1.34	2.78	4.77	0.00	0.00	1.52	1.52	0.00	0.00	0.55	0.32
24	Insulin-like growth factor 2	149	4.67	7.02	4.73	16.41	0.32	4.00	4.77	9.08	0.07	0.57	1.01	0.55
25	Interleukin 1α	260	7.91	6.96	15.70	30.58	7.60	10.07	15.47	33.14	0.96	1.45	0.99	1.08
26	Interleukin 1β	263	7.05	19.34	12.50	38.88	10.05	12.30	10.15	32.50	1.43	0.64	0.81	0.84
27	Interleukin 6	205	18.64	18.20	31.65	68.49	16.24	15.59	39.43	71.26	0.87	0.86	1.25	1.04
28	Interleukin 7	153	8.30	7.67	11.10	27.08	10.58	5.56	7.39	23.53	1.27	0.73	0.67	0.87
29	Lactate dehydrogenase A	331	2.49	1.53	1.89	5.92	0.89	1.83	1.11	3.82	0.36	1.19	0.59	0.65
30	Lactoferrin	662	7.01	12.51	12.71	32.22	7.16	14.70	12.77	34.63	1.02	1.18	1.01	1.07
31	Luteinizing hormone receptor	685	4.14	3.17	5.08	12.40	3.69	2.68	4.91	11.28	0.89	0.85	0.97	0.91
32	Myelin proteolipid protein	148	0.65	1.30	0.00	1.94	1.75	0.00	0.00	1.75	2.71	0.00	NA	0.90
33	Neuroleukin	557	2.16	2.30	4.48	8.94	1.34	1.27	5.11	7.72	0.62	0.55	1.14	0.86
34	Neurophysin I	162	4.70	5.10	7.60	17.39	0.91	0.68	7.48	9.06	0.19	0.13	0.98	0.52
35	Neurophysin II	116	7.27	5.29	2.49	15.04	0.00	15.08	0.07	15.15	0.00	2.85	0.03	1.01
36	Ornithine decarboxylase	460	1.93	2.05	3.98	7.96	1.23	1.16	3.53	5.91	0.64	0.56	0.89	0.74
37	Plasminogen activator inhibitor	386	3.10	5.07	8.00	16.17	4.02	3.63	10.00	17.65	1.30	0.71	1.25	1.09
38	Prolactin	197	6.82	11.09	25.92	43.83	3.03	11.19	18.32	32.54	0.44	1.01	0.71	0.74
39	Proopiomelanocortin	211	1.44	2.23	8.43	12.10	2.87	2.20	6.04	11.11	1.99	0.99	0.72	0.92
40	Protein disulfide isomerase	505	1.77	1.93	3.44	7.14	1.66	0.58	1.82	4.06	0.94	0.30	0.53	0.57
41	Terminal transferase	506	4.71	4.08	9.53	18.32	3.35	2.36	6.30	12.01	0.71	0.58	0.66	0.66
42	Thrombomodulin	341	11.68	13.61	10.88	36.16	9.61	11.11	13.33	34.05	0.82	0.82	1.23	0.94
43	Transforming growth factor β1	315	1.37	2.52	6.93	10.82	1.60	0.86	4.63	7.09	1.17	0.34	0.67	0.66
44	Transforming growth factor β2	413	0.55	0.00	2.40	2.95	0.15	0.44	1.03	1.62	0.26	NA	0.43	0.55
45	Transforming growth factor β3	408	0.27	4.72	1.39	6.38	0.20	4.04	0.37	4.61	0.74	0.85	0.27	0.72
46	Transforming growth factor β3 receptor	843	3.82	5.55	7.13	16.49	3.65	4.67	4.35	12.67	0.96	0.84	0.61	0.77
47	Urokinase-plasminogen activator	403	8.45	4.66	12.03	25.14	7.41	10.05	13.22	30.69	0.88	2.16	1.10	1.22
Mean		393	4.41	5.40	6.99	16.79	3.45	4.39	5.81	13.65	0.72	0.83	0.70	0.73

charge changes were estimated for the 47 genes for the three branches leading to primates, artiodactyls, and rodents, respectively (Table 1, Fig. 1). For both conservative and radical substitutions, the branch of the rodent lineage is longer than that of the artiodactyl or primate lineages. Using a sign test (Sokal and Rohlf 1995, p. 444), I found that these differences are statistically significant (Table 2). This result is consistent with the consensus that rodents diverged from primates and artiodactyls before the latter two diverged from each other. The branch of the artiodactyl lineage appears to be longer

Table 2. Comparisons of conservative and radical distances for the three lineages: The property of charge is considered

Comparison	b_C Counts	b_C p^a	b_R Counts	b_R p	b_R/b_C Counts	b_R/b_C p
P < A	29 (47)[b]	NS	27 (47)	NS	21 (46)	NS
P < R	38 (47)	<0.0001	38 (47)	<0.0001	24 (46)	NS
A < R	32 (47)	0.0186	31 (47)	0.0400	24 (45)	NS

[a] Probability in two-tail sign tests. NS, not significant.

[b] Number of comparisons (genes) for which the inequality $b_C(P) < b_C(A)$ holds. The total number of comparisons is given in parentheses.

Table 3. Number of conservative and radical substitutions per site for the evolutionary lineages of primates (P), artiodactyles (A), and rodents (R): Polarity changes are considered

	b_C (×100)				b_R (×100)				b_R/b_C			
Gene No.	P	A	R	T	P	A	R	T	P	A	R	T
1	1.51	0.48	2.21	4.20	1.08	1.08	1.08	3.24	0.72	2.24	0.49	0.77
2	2.75	1.72	4.24	8.71	1.71	2.95	2.99	7.65	0.62	1.71	0.70	0.88
3	2.46	4.26	5.97	12.69	4.11	8.35	5.28	17.74	1.67	1.96	0.89	1.40
4	5.19	10.12	12.68	27.98	6.05	7.82	7.77	21.63	1.17	0.77	0.61	0.77
5	4.11	9.99	10.02	24.12	5.37	8.24	7.08	20.69	1.31	0.82	0.71	0.86
6	2.95	3.42	4.29	10.66	2.99	2.09	5.00	10.08	1.01	0.61	1.17	0.95
7	2.34	3.30	3.38	9.02	0.33	1.66	2.00	3.99	0.14	0.50	0.59	0.44
8	1.55	1.80	1.56	4.90	1.01	3.01	0.00	4.02	0.65	1.67	0.00	0.82
9	0.98	0.65	0.78	2.41	0.27	0.28	0.84	1.39	0.28	0.43	1.07	0.58
10	0.05	0.54	0.54	1.13	0.00	0.39	0.79	1.18	0.00	0.73	1.46	1.05
11	2.40	7.44	5.25	15.09	2.94	5.40	5.80	14.13	1.22	0.73	1.10	0.94
12	0.44	2.58	1.95	4.96	0.14	1.12	2.06	3.32	0.32	0.43	1.06	0.67
13	0.53	0.35	0.89	1.77	0.72	0.72	0.36	1.80	1.35	2.04	0.41	1.01
14	2.18	13.09	8.32	23.58	2.11	9.33	11.32	22.75	0.97	0.71	1.36	0.96
15	0.78	0.94	1.34	3.06	1.69	0.33	0.33	2.36	2.16	0.35	0.25	0.77
16	3.22	7.45	8.13	18.79	3.34	6.12	5.28	14.75	1.04	0.82	0.65	0.78
17	1.37	0.71	0.70	2.78	0.30	0.61	1.21	2.12	0.22	0.86	1.73	0.76
18	19.09	1.77	4.81	25.67	17.66	8.13	0.70	26.49	0.93	4.60	0.14	1.03
19	6.03	3.34	13.20	22.57	4.61	2.64	12.25	19.50	0.77	0.79	0.93	0.86
20	2.34	3.50	2.41	8.25	1.38	2.11	1.41	4.90	0.59	0.60	0.59	0.59
21	12.64	9.53	7.80	29.97	7.89	8.25	10.24	26.38	0.62	0.87	1.31	0.88
22	4.18	5.60	6.64	16.43	6.66	7.15	7.26	21.08	1.59	1.28	1.09	1.28
23	0.00	0.00	1.55	1.56	1.31	2.61	3.88	7.80	NA	NA	2.50	5.02
24	1.85	4.93	3.39	10.17	5.55	7.79	7.91	21.25	3.00	1.58	2.33	2.09
25	7.77	9.22	17.81	34.79	7.80	5.81	11.09	24.69	1.00	0.63	0.62	0.71
26	10.03	19.05	12.64	41.72	4.66	10.58	8.88	24.12	0.46	0.56	0.70	0.58
27	15.19	16.07	39.48	70.74	23.60	20.27	23.70	67.57	1.55	1.26	0.60	0.96
28	8.43	6.64	11.48	26.54	11.22	7.30	4.91	23.43	1.33	1.10	0.43	0.88
29	2.17	2.06	1.97	6.20	1.17	0.70	0.70	2.57	0.54	0.34	0.36	0.41
30	7.17	15.18	13.79	36.13	6.76	9.68	10.56	27.00	0.94	0.64	0.77	0.75
31	3.94	3.81	5.05	12.81	4.10	1.46	4.98	10.55	1.04	0.38	0.99	0.82
32	1.01	1.02	0.00	2.03	0.83	0.83	0.00	1.66	0.82	0.81	NA	0.82
33	2.10	1.80	5.43	9.33	1.31	2.05	3.43	6.79	0.62	1.14	0.63	0.73
34	3.45	2.19	6.75	12.39	2.49	6.32	9.77	18.58	0.72	2.89	1.45	1.50
35	5.95	7.74	1.58	15.27	1.43	11.46	1.34	14.23	0.24	1.48	0.85	0.93
36	1.97	1.90	4.29	8.15	1.02	1.33	2.79	5.14	0.52	0.70	0.65	0.63
37	4.01	5.91	8.31	18.23	2.28	1.86	9.51	13.64	0.57	0.31	1.14	0.75
38	6.29	13.00	24.35	43.64	2.82	7.59	19.26	29.67	0.45	0.58	0.79	0.68
39	2.12	1.87	7.83	11.82	1.79	3.08	6.51	11.37	0.84	1.65	0.83	0.96
40	1.85	1.82	2.77	6.43	1.47	0.30	2.69	4.45	0.80	0.16	0.97	0.69
41	4.90	3.42	9.08	17.39	2.60	3.30	6.27	12.17	0.53	0.97	0.69	0.70
42	11.06	13.53	11.32	35.90	10.42	10.60	13.02	34.05	0.94	0.78	1.15	0.95
43	1.97	2.29	6.64	10.90	0.35	0.87	4.60	5.82	0.18	0.38	0.69	0.53
44	0.57	0.24	1.73	2.54	0.00	0.00	2.20	2.20	0.00	0.00	1.28	0.87
45	0.34	5.15	1.39	6.88	0.00	2.79	0.00	2.79	0.00	0.54	0.00	0.41
46	4.21	5.71	6.53	16.45	2.89	4.25	5.21	12.35	0.69	0.74	0.80	0.75
47	9.66	8.24	13.91	31.81	4.87	4.01	9.58	18.46	0.50	0.49	0.69	0.58
Mean	4.19	5.22	6.94	16.35	3.73	4.57	5.61	13.90	0.82	1.00	0.87	0.93

than that of the primate lineage for both conservative and radical substitutions (Table 1), though the differences are not statistically significant (Table 2).

When the total branch length of the three-species tree is considered, the average number of radical substitutions per site [$\bar{b}_R(T)$] for the 47 genes is about 81.3% (13.65/16.79; Table 1) that [$\bar{b}_C(T)$] of conservative substitutions, and a sign test shows that the difference is highly significant ($p < 0.0001$). This indicates that on average, the rate of radical nonsynonymous substitution is lower than that of conservative substitution. For the 47 genes examined, the ratio [$\lambda(T)$] of $b_R(T)$ to $b_C(T)$ varies from 0.22 to 1.22, with a mean of 0.73 and a standard error of 0.24. I also compared λ for each branch of the tree. Let us use $\bar{X}$ to denote the average of a quantity X for the 47 genes. The values of $\bar{\lambda}(P)$, $\bar{\lambda}(A)$, and $\bar{\lambda}(R)$ are then 0.75,

Table 4. Comparisons of conservative and radical distances for the three lineages: The property of polarity is considered

	b_C		b_R		b_R/b_C	
Comparison	Counts	p^a	Counts	p	Counts	p
P < A	28 (47)[b]	NS	33 (47)	0.0079	26 (46)	NS
P < R	40 (47)	<0.0001	37 (47)	0.0001	24 (45)	NS
A < R	31 (47)	0.0400	32 (47)	0.0186	23 (45)	NS

[a] Probability in two-tail sign tests. NS, not significant.
[b] Number of comparisons (genes) for which the inequality $b_C(P) < b_C(A)$ holds. The total number of comparisons is given in parentheses.

0.81, and 0.69, respectively, for the primate, artiodactyl, and rodent lineages (Table 1). A sign test shows that the above three λ values are not significantly different from each other (Table 2).

Because the numbers of conservative and radical changes are rather small for some genes, estimates of λ may have large sampling errors. We therefore computed $\bar{b}_R/\bar{b}_C$ for the three branches. These values are 0.78 (3.45/4.41), 0.81 (4.39/5.40), and 0.83 (5.81/6.99) for primates, artiodactyl, and rodents, respectively. We also computed $\bar{b}_C$ and $\bar{b}_R$ values for each branch by taking into account the sequence length of each gene. That is, we weighted individual b_C and b_R values by the numbers of conservative and radical sites of the gene, respectively. In this case, we obtained weighted $\bar{b}_R/\bar{b}_C$ ratios of 0.82, 0.87, and 0.84, respectively, for the primate, artiodactyl, and rodent lineages. By using the estimated b_C and b_R values for individual genes, we estimated that the total numbers of conservative substitutions for all 47 genes are 959, 1203, and 1604 for the three lineages, respectively, and the total numbers of radical substitutions are 482, 638, and 825, respectively. Using Fisher's exact test (see Zhang et al. 1997), we found that the ratios of the number of radical substitutions to that of conservative substitutions are not significantly different between any pair of the three lineages. These results suggest that the radical/conservative rate ratios are similar among the lineages.

Rates of Conservative and Radical Substitutions with Respect to Polarity

I computed b_C and b_R with respect to polarity changes of amino acids (Fig. 1) for the primate, artiodactyl, and rodent lineages (Table 3). Again, b(R) is significantly greater than b(A) and b(P) for both conservative and radical substitutions (Table 4). In addition, b(A) is found to be significantly greater than b(P) for radical substitutions. For conservative substitutions, however, b(A) is greater than b(P) with the difference being statistically insignificant (Table 4).

When the total branch length of the tree is considered, the average number of radical substitutions per site $[\bar{b}_R(T)]$ for the 47 genes is about 85.0% (13.90/16.35; Table 3) of the average number of conservative substitutions per site $[\bar{b}_C(T)]$, and a sign test shows that the difference is highly significant ($p < 0.0001$). This means that, on average, the radical nonsynonymous substitution rate is lower than the conservative nonsynonymous rate with respect to polarity. For the 47 genes examined, the ratio $[\lambda(T)]$ of $b_R(T)$ to $b_C(T)$ varies from 0.41 to 5.02, with a mean of 0.93 and a standard error of 0.67. The values of $\bar{\lambda}(P)$, $\bar{\lambda}(A)$, and $\bar{\lambda}(R)$ are 0.82, 0.99, and 0.87, respectively (Table 3), and their differences are not statistically significantly (sign tests; Table 4).

The values of $\bar{b}_R/\bar{b}_C$ are 0.89 (3.73/4.19), 0.88 (4.57/5.22), and 0.81 (5.61/6.94) for the primate, artiodactyl, and rodent lineages, respectively. When we weighted each b_C and b_R values by C and R of the gene, respectively, we obtained weighted $\bar{b}_R/\bar{b}_C$ values of 0.83, 0.77, and 0.78, respectively, for the three lineages. For the 47 genes, the total numbers of conservative substitutions are estimated to be 1034, 1349, and 1772 for the primate, artiodactyl, and rodent lineages, respectively, whereas the numbers of radical substitutions are 407, 492, and 656. Fisher's exact test shows that the ratios of the number of radical substitutions to that of conservative substitutions are not significantly different between any pair of the three lineages.

Rates of Conservative and Radical Substitutions with Respect to Polarity and Volume

According to Miyata and co-workers' (1979) classification of amino acids (Fig. 1), which is based on the polarities and volumes of amino acid residues, we computed b_C and b_R for the 47 genes (Table 5). We found that b(R) is significantly greater than b(P) for both conservative and radical substitutions (Table 6). The difference between b(P) and b(A) and that between b(A) and b(R), however, are not significant for either conservative or radical substitutions (Table 6).

The average number of radical substitutions per site $[\bar{b}_R(T)]$ is about 49.3% (11.80/23.94; Table 5) that $[\bar{b}_C(T)]$ of conservative substitutions, and their difference is highly significant (sign test, $p < 0.0001$). This suggests that, on average, the rate of radical nonsynonymous substitution is lower than the rate of conservative substitution with regard to the present classification of amino acids. For the 47 genes examined, $\lambda(T)$ varies from 0.15 to 0.73, with a mean of 0.46 and a standard error of 0.14. We also compared the λ values for each branch of the tree. The values of $\bar{\lambda}(P)$, $\bar{\lambda}(A)$, and $\bar{\lambda}(R)$ are 0.47, 0.53, and 0.46, respectively. A sign test suggests that the λ values are not significantly different among the three branches (Table 6). The values of $\bar{b}_R/\bar{b}_C$ are 0.45 (2.92/6.54), 0.48 (3.75/7.76), and 0.53 (5.12/9.63), for the pri-

Table 5. Numbers of conservative and radical substitutions per site for the evolutionary lineages of primates (P), artiodactyles (A), and rodents (R): Both polarity and volume are considered

	b_C (×100)				b_R (×100)				b_R/b_C			
Gene No.	P	A	R	T	P	A	R	T	P	A	R	T
1	0.96	0.66	4.79	6.41	1.57	0.70	0.31	2.58	1.64	1.06	0.06	0.40
2	2.93	2.32	5.27	10.51	2.14	2.04	3.13	7.31	0.73	0.88	0.59	0.70
3	5.46	7.65	10.81	23.92	1.77	4.49	3.35	9.60	0.32	0.59	0.31	0.40
4	10.49	16.07	17.67	44.23	3.38	6.65	8.36	18.39	0.32	0.41	0.47	0.42
5	5.88	14.43	13.16	33.48	3.79	6.85	6.88	17.52	0.64	0.47	0.52	0.52
6	4.61	4.74	5.31	14.65	2.10	2.04	4.10	8.24	0.46	0.43	0.77	0.56
7	2.72	4.64	5.79	13.14	1.13	1.83	1.57	4.53	0.42	0.39	0.27	0.34
8	3.20	4.41	3.04	10.65	0.45	1.11	0.04	1.60	0.14	0.25	0.01	0.15
9	1.03	1.42	1.54	3.99	0.59	0.03	0.38	1.01	0.57	0.02	0.25	0.25
10	0.08	0.79	1.14	2.01	0.00	0.31	0.32	0.63	0.00	0.40	0.28	0.31
11	3.74	9.94	8.51	22.18	1.96	5.27	3.98	11.21	0.52	0.53	0.47	0.51
12	0.56	4.15	2.66	7.37	0.22	1.15	1.68	3.05	0.40	0.28	0.63	0.41
13	1.42	0.46	1.60	3.48	0.22	0.48	0.31	1.00	0.15	1.04	0.19	0.29
14	4.64	20.55	12.34	37.52	0.94	7.84	7.78	16.56	0.20	0.38	0.63	0.44
15	1.64	0.68	2.11	4.43	0.78	0.78	0.45	2.00	0.47	1.15	0.21	0.45
16	7.18	11.20	10.12	28.48	1.55	5.24	5.87	12.66	0.22	0.47	0.58	0.44
17	1.45	1.47	1.90	4.82	0.80	0.26	0.34	1.40	0.55	0.18	0.18	0.29
18	34.18	8.17	7.77	50.12	12.84	1.75	1.67	16.25	0.38	0.21	0.22	0.32
19	7.81	5.19	16.71	29.71	4.47	2.07	11.03	17.57	0.57	0.40	0.66	0.59
20	4.44	3.53	3.37	11.33	0.85	2.82	1.45	5.12	0.19	0.80	0.43	0.45
21	18.25	19.94	15.08	53.28	7.87	4.92	5.98	18.77	0.43	0.25	0.40	0.35
22	8.90	13.50	9.41	31.82	3.12	2.77	5.56	11.45	0.35	0.20	0.59	0.36
23	1.26	2.66	3.99	7.92	0.00	0.00	1.54	1.54	0.00	0.00	0.39	0.20
24	8.82	8.10	5.39	22.31	0.24	4.75	4.42	9.41	0.03	0.59	0.82	0.42
25	7.89	11.94	18.67	38.50	7.62	6.33	14.20	28.14	0.97	0.53	0.76	0.73
26	11.32	24.72	13.99	50.04	6.92	12.73	10.38	30.03	0.61	0.51	0.74	0.60
27	25.65	26.09	43.97	95.70	14.20	13.35	30.79	58.34	0.55	0.51	0.70	0.61
28	12.83	6.09	18.68	37.59	7.84	6.97	5.91	20.71	0.61	1.14	0.32	0.55
29	4.57	1.90	2.12	8.59	0.56	1.53	1.33	3.41	0.12	0.80	0.63	0.40
30	12.72	21.50	17.14	51.36	4.67	10.07	10.78	25.52	0.37	0.47	0.63	0.50
31	4.16	3.76	7.59	15.51	3.91	2.66	3.80	10.37	0.94	0.71	0.50	0.67
32	0.92	1.38	0.00	2.29	0.96	0.72	0.00	1.68	1.04	0.53	NA	0.73
33	3.66	3.52	7.31	14.48	0.96	1.11	3.52	5.59	0.26	0.32	0.48	0.39
34	4.77	5.22	15.08	25.07	2.48	2.56	4.54	9.57	0.52	0.49	0.30	0.38
35	8.77	7.47	3.13	19.37	2.83	9.46	0.81	13.09	0.32	1.27	0.26	0.68
36	2.76	2.53	4.84	10.13	1.10	1.30	3.27	5.67	0.40	0.51	0.68	0.56
37	4.77	5.79	11.86	22.42	2.72	3.89	7.09	13.70	0.57	0.67	0.60	0.61
38	9.74	15.84	29.59	55.17	3.20	9.09	19.79	32.07	0.33	0.57	0.67	0.58
39	3.75	2.92	10.89	17.56	1.21	1.85	5.89	8.95	0.32	0.63	0.54	0.51
40	2.06	2.55	4.56	9.17	1.56	0.76	1.86	4.19	0.76	0.30	0.41	0.46
41	5.16	5.43	11.49	22.08	3.68	2.43	6.68	12.78	0.71	0.45	0.58	0.58
42	16.97	25.51	18.72	61.19	8.18	7.76	9.17	25.11	0.48	0.30	0.49	0.41
43	2.40	3.54	9.44	15.38	1.05	1.09	4.47	6.61	0.44	0.31	0.47	0.43
44	1.05	0.00	3.71	4.77	0.08	0.25	1.00	1.33	0.08	NA	0.27	0.28
45	0.25	6.87	1.55	8.67	0.24	3.32	0.70	4.26	0.94	0.48	0.45	0.49
46	4.30	6.64	10.10	21.04	3.47	4.51	4.10	12.09	0.81	0.68	0.41	0.57
47	15.35	7.07	18.97	41.39	5.25	6.59	10.06	21.90	0.34	0.93	0.53	0.53
Mean	6.54	7.76	9.63	23.94	2.92	3.75	5.12	11.80	0.47	0.53	0.47	0.46

mate, artiodactyl, and rodent lineages, respectively. The weighted $\overline{b}_R/\overline{b}_C$ ratios are 0.48, 0.49, and 0.53, respectively. The total numbers of conservative substitutions for all 47 genes were estimated to be 732, 944, and 1187 for the primate, artiodactyl, and rodent lineages, respectively, and the total numbers of radical substitutions were 705, 923, and 1265. Fisher's exact test shows that the ratios of the number of radical substitutions to that of conservative substitutions are not significantly different between any pair of the three lineages.

Rates of Synonymous and Nonsynonymous Substitutions

It will be interesting to compare the rates of conservative and radical nonsynonymous substitutions with the rate of

Table 6. Comparisons of conservative and radical distances for the three lineages: Both polarity and volume are considered

	b_C		b_R		b_R/b_C	
Comparison	Counts	p^a	Counts	p	Counts	p
P < A	30 (47)[b]	NS	29 (47)	NS	22 (46)	NS
P < R	39 (47)	<0.0001	36 (47)	0.0003	25 (46)	NS
A < R	30 (47)	NS	29 (47)	NS	24 (45)	NS

[a] Probability in two-tail sign tests. NS, not significant.
[b] Number of comparisons (genes) for which the inequality $b_C(P) < b_C(A)$ holds. The total number of comparisons is given in parentheses.

synonymous substitution for the 47 genes used in this article. Ohta (1995) estimated the synonymous rates for these genes using Ina's (1995) method and Yang and Nielsen (1998) recently reestimated them using a maximum-likelihood method. To be comparable with the estimates of the conservative and radical substitution rates presented above, I estimated the synonymous and nonsynonymous rates (Table 7) using the modified Nei–Gojobori method (Zhang et al. 1998), which takes into account the transition bias in the same way as in the computation of d_C and d_R.

The estimates of the branch lengths in terms of nonsynonymous substitutions (b_N) are similar to those obtained by Ohta (1995) and Yang and Nielsen (1998). For the branch lengths in terms of synonymous substitutions (b_S), my estimates and Ohta's are quite close, and both are, in general, smaller than Yang and Nielsen's. For comparison of the branch lengths of the primate, artiodactyl, and rodent lineages, however, all three methods gave similar results. For example, I obtained $\bar{b}_S(P) = 0.14$, $\bar{b}_S(A) = 0.18$, and $\bar{b}_S(R) = 0.34$ for the 47 genes. These numbers are in the proportion of 0.64:0.82:1.55, similar to the previous estimates of 0.57:0.87:1.57 (Yang and Nielsen 1998) and 0.61:0.82:1.58 (Ohta 1995). It should be pointed out that Yang and Nielsen used one more gene (interleukin-2) than the 47 genes used here, and Ohta used two more genes (interleukin-2 and opsin). The sign test shows that the b_S values are significantly different between each pair of the three branches (Table 8). For nonsynonymous substitutions, I obtained $\bar{b}_N(P) = 0.040$, $\bar{b}_N(A) = 0.050$, and $\bar{b}_N(R) = 0.065$. These numbers are in the proportion of 0.77:0.97:1.26, almost-identical to the values of 0.75:0.98:1.27 of Yang and Nielsen (1998) and 0.75:0.97:1.28 of Ohta (1995). The sign test shows that the difference in the nonsynonymous branch length is significant between the primate and the rodent lineages and between the artiodactyl and the rodent lineages.

The average nonsynonymous substitution rate for the 47 genes [$\bar{b}_N(T)$] is about 23.3% (15.54/66.62; Table 8) of the average synonymous rate [$\bar{b}_N(T)$]. The ratio [$\omega(T)$] of the nonsynonymous rate $b_N(T)$ to the synonymous rate $b_S(T)$ varies from 0.02 to 0.90, with a mean of 0.23 and a standard error of 0.19. The values of $\bar{\omega}(P)$, $\bar{\omega}(A)$, and $\bar{\omega}(R)$ are 0.282, 0.274, and 0.191, respectively. A sign test suggests that ω is significantly smaller in the rodent lineage than in the primate or artiodactyl lineage (Table 8). The same conclusion was obtained when a Fisher's exact test was used for the estimated total numbers of synonymous and nonsynonymous substitutions of the 47 genes for the three branches.

Discussion

Is the Rate of Radical Substitution Lower Than That of Conservative Substitution?

For all three classifications of amino acids considered, nonsynonymous nucleotide substitutions that change the physicochemical properties of amino acids occur with a lower rate than those that do not change the properties. This result is consistent with the common wisdom that physicochemical properties of amino acids are relevant to protein functions and that radical changes are more likely to be subject to purifying selection than conservative ones. For instance, the radical substitution rate is about half of the conservative rate when amino acids are classified by polarity and volume. This means that under this classification, a conservative nonsynonymous mutation is on average twice as likely to be neutral as a radical one. The fact that radical mutations are more likely to be under purifying selection than conservative mutations may suggest that radical mutations are also more likely to contribute to evolutionary changes in protein function if fixed. It is, however, somewhat surprising that the difference between the rates of conservative and radical substitutions with respect to either charge or polarity is quite small. The radical/conservative rate ratio is about 0.81 to 0.85. This mild difference suggests that the intensity of purifying selection is not substantially greater for radical changes than for conservative changes when charge or polarity is considered. This further suggests that charge or polarity alone may not be such a good property as previously thought in predicting functionally important substitutions.

The nonsynonymous/synonymous rate ratio is about 0.23. If we assume that a nonsynonymous mutation is either neutral or deleterious and that all synonymous mutations are neutral, we can infer that about 23% of the nonsynonymous substitutions are neutral and 77% are deleterious and therefore are removed from population. Since the rate of conservative nonsynonymous substitution with respect to charge is about 25% [$\bar{b}_C(T)/\bar{b}_S(T) = 16.79/66.62$] of the rate of synonymous substitution, we can infer that about 25% of conservative nonsynony-

Table 7. Numbers of synonymous and nonsynonymous substitutions per site for the evolutionary lineages of primates (P), artiodactyles (A), and rodents (R)

	b_S (×100)				b_N (×100)				b_N/b_S			
Gene No.	P	A	R	T	P	A	R	T	P	A	R	T
1	12.25	12.01	29.65	53.92	1.36	0.68	1.83	3.88	0.111	0.057	0.062	0.072
2	15.54	15.35	32.18	63.08	2.40	2.13	3.82	8.36	0.155	0.139	0.119	0.132
3	25.62	12.97	37.67	76.26	2.98	5.55	5.75	14.29	0.116	0.428	0.153	0.187
4	25.71	15.59	57.95	99.25	5.48	9.38	11.10	25.96	0.213	0.601	0.192	0.262
5	21.58	24.90	35.10	81.58	4.55	9.39	9.00	22.93	0.211	0.377	0.256	0.281
6	14.77	17.21	36.95	68.93	2.96	2.99	4.52	10.47	0.201	0.174	0.122	0.152
7	14.43	18.99	34.60	68.02	1.65	2.72	2.90	7.27	0.114	0.143	0.084	0.107
8	15.92	18.38	29.88	64.18	1.37	2.20	1.03	4.60	0.086	0.120	0.035	0.072
9	11.59	16.73	36.24	64.55	0.75	0.53	0.80	2.08	0.065	0.032	0.022	0.032
10	9.22	15.57	32.52	57.32	0.03	0.49	0.62	1.15	0.003	0.032	0.019	0.020
11	9.97	19.93	27.36	57.25	2.57	6.77	5.43	14.77	0.258	0.340	0.199	0.258
12	13.43	23.92	48.61	85.97	0.34	2.09	1.99	4.42	0.025	0.088	0.041	0.051
13	22.35	10.81	40.12	73.28	0.59	0.47	0.71	1.78	0.027	0.044	0.018	0.024
14	11.19	25.19	27.51	63.89	2.13	11.86	9.30	23.29	0.191	0.471	0.338	0.365
15	11.78	10.65	20.82	43.25	1.07	0.75	1.02	2.84	0.091	0.070	0.049	0.066
16	9.04	19.64	52.27	80.96	3.26	7.03	7.20	17.49	0.360	0.358	0.138	0.216
17	8.35	17.17	30.08	55.60	1.02	0.67	0.87	2.56	0.122	0.039	0.029	0.046
18	27.84	17.08	22.11	67.03	18.65	3.66	3.59	25.89	0.670	0.214	0.162	0.386
19	6.42	14.19	30.57	51.18	5.56	3.11	12.89	21.55	0.865	0.219	0.422	0.421
20	13.25	18.74	43.90	75.89	2.03	3.05	2.09	7.17	0.153	0.163	0.048	0.095
21	26.77	27.10	52.24	106.11	10.99	9.08	8.64	28.71	0.411	0.335	0.165	0.271
22	9.17	35.26	28.29	72.72	4.91	6.06	6.83	17.80	0.535	0.172	0.241	0.245
23	8.15	18.18	36.51	62.84	0.41	0.83	2.32	3.56	0.050	0.046	0.064	0.057
24	11.37	20.54	20.08	51.99	2.95	5.83	4.74	13.53	0.259	0.284	0.236	0.260
25	16.53	14.64	32.38	63.55	7.78	8.15	15.63	31.56	0.470	0.557	0.483	0.497
26	11.99	28.75	29.64	70.39	8.33	16.37	11.48	36.18	0.695	0.569	0.387	0.514
27	15.34	18.08	43.50	76.91	17.69	17.21	34.62	69.52	1.154	0.952	0.796	0.904
28	6.27	9.48	22.67	38.42	9.24	6.79	9.58	25.62	1.474	0.717	0.423	0.667
29	10.93	13.35	52.40	76.68	1.87	1.65	1.59	5.11	0.171	0.123	0.030	0.067
30	15.95	29.51	36.59	82.04	7.07	13.37	12.73	33.17	0.443	0.453	0.348	0.404
31	12.30	12.81	33.80	58.91	4.00	3.02	5.03	12.04	0.325	0.236	0.149	0.204
32	3.26	7.25	11.21	21.73	0.94	0.95	0.01	1.88	0.289	0.130	0.001	0.087
33	18.51	13.83	29.24	61.57	1.83	1.89	4.74	8.45	0.099	0.137	0.162	0.137
34	6.94	8.31	24.76	40.01	3.19	3.35	7.60	14.14	0.460	0.403	0.307	0.353
35	9.66	17.45	21.74	48.85	4.57	8.88	1.49	14.93	0.473	0.509	0.069	0.306
36	23.97	19.04	28.23	71.24	1.66	1.71	3.81	7.18	0.069	0.090	0.135	0.101
37	18.17	17.32	40.74	76.24	3.43	4.55	8.72	16.70	0.189	0.263	0.214	0.219
38	19.23	24.75	36.68	80.66	5.20	11.21	22.73	39.13	0.270	0.453	0.620	0.485
39	8.68	19.74	23.45	51.86	2.03	2.21	7.45	11.69	0.233	0.112	0.318	0.225
40	19.79	26.26	41.22	87.27	1.73	1.34	2.74	5.82	0.087	0.051	0.067	0.067
41	15.92	7.47	43.84	67.23	4.16	3.39	8.19	15.73	0.261	0.453	0.187	0.234
42	16.84	23.48	53.53	93.85	10.85	12.61	11.86	35.31	0.644	0.537	0.222	0.376
43	15.94	19.20	35.11	70.24	1.47	1.85	5.99	9.31	0.092	0.096	0.171	0.133
44	11.06	14.82	33.29	59.17	0.39	0.17	1.87	2.43	0.036	0.011	0.056	0.041
45	9.69	15.77	23.18	48.64	0.24	4.44	0.98	5.66	0.025	0.282	0.042	0.116
46	13.76	31.25	33.89	78.90	3.77	5.22	6.10	15.08	0.274	0.167	0.180	0.191
47	16.19	16.86	28.75	61.81	8.04	6.84	12.50	27.39	0.496	0.406	0.435	0.443
Mean	14.31	18.20	34.11	66.62	4.03	4.99	6.52	15.54	0.298	0.269	0.192	0.231

mous mutations are neutral. For radical mutations, this number is about 21%. The corresponding values for the other two amino acid classifications are given in Table 9. Although different proteins and different sites have distinct substitution patterns, the above numbers, derived from a total of 55,434 nucleotide sites of 47 genes, should provide a rough idea about the relative likelihood of a certain type of mutation being functionally relevant.

Can an Observation of $\lambda > 1$ *Be Used as Evidence for Positive Selection?*

As stated earlier, the 47 genes examined here are used as a sample of genes in which positive selection can be neglected. This assumption is probably valid because all 47 genes have ω(T) values lower than 1. For individual branches, however, the ω(P) values for interleukin 6 and

Table 8. Comparisons of synonymous and nonsynonymous distances for the three lineages

	b_S		b_N		b_N/b_S	
Comparison	Counts	p^a	Counts	p	Counts	p
P < A	35 (47)[b]	0.0010	30 (47)	NS	22 (47)	NS
P < R	47 (47)	<0.0001	39 (47)	<0.0001	15 (47)	0.0186
A < R	45 (47)	<0.0001	31 (47)	0.0400	13 (47)	0.0031

[a] Probability in two-tail sign tests. NS, not significant.
[b] Number of comparisons (genes) for which the inequality $b_S(P) < b_S(A)$ holds. The total number of comparisons is given in parentheses.

interleukin 7 are found to be greater than 1. But neutral evolution cannot be rejected statistically in either case by the b_N/b_S test (Zhang et al. 1998).

A significantly higher rate of radical nonsynonymous substitution than of conservative substitution (i.e., $\lambda > 1$) has been regarded as evidence for positive selection (Hughes 1992, 1994; Hughes and Hughes 1993). In fact, an observation of $\lambda > 1$ can have two possible explanations. The first explanation is that positive selection favors radical changes to conservative changes and the other is that purifying selection somehow removes more conservative mutations than radical ones. Although we have shown that the average intensity of purifying selection is higher for radical substitutions than for conservative substitutions, the difference is small when the property of charge or polarity is considered. This provides the possibility of a higher intensity of purifying selection for conservative mutations than for radical mutations in some genes. For example, if a change in amino acid volume is prohibited at some positions in a protein structure, a conservative substitution with respect to charge such as Asn → Tyr may be prohibited because the two amino acids have different volumes, whereas a radical change (with respect to charge) such as Asn → Asp may be allowed because they have similar volumes. If there are many such sites in a protein, the overall intensity of purifying selection may be lower for radical substitutions than for conservative substitutions when charge changes are considered. Furthermore, the R/C ratio may be different among different regions of a sequence. Use of the average ratio over the whole sequence may give false impressions. For instance, if a sequence consists of two regions with $R/C = 70/100$ for region I and $R/C = 40/100$ for region II, and $R_d/C_d = 18/29$ for region I and $R_d/C_d = 0/1$ for region II, then for the whole sequence, we have $R_d/C_d = (18 + 0)/(29 + 1) > (70 + 40)/(100 + 100) = R/C$, whereas for each of the two regions $R_d/C_d < R/C$. Although this example appears extreme, it can happen. In fact, the R/C ratio for charge changes varies from 0.37 to 0.74 for the 47 genes examined. So, it seems that an observation of $\lambda > 1$ alone is not proof of positive selection. Nevertheless, such an observation does point to the likelihood of occurrence of positive selection, which may be further investigated.

We examined all 47 genes to see if there is any gene with $\lambda(T) > 1$. With respect to polarity and volume, there are no genes that have $\lambda(T) > 1$. For individual branches, although some λ values are slightly greater than 1, the null hypothesis of $b_C = b_R$ cannot be rejected in any of these cases. For example, $b_C(P) = 0.0096 \pm 0.0052$ (1 SE) and $b_R(P) = 0.0157 \pm 0.0053$ for the acetylcholine receptor α gene, but the difference between $b_C(P)$ and $b_R(P)$ is not statistically significant ($t = 0.82 < t_{0.05}$). For the classification according to charge, there are six genes with $\lambda(T) > 1$, but in no case is $\lambda(T)$ significantly greater than 1. For individual branches, however, we found that $\lambda(A)$ is significantly greater than 1 for the neurophysin II gene ($t = 1.87$, $p < 0.05$) and the urokinase–plasminogen activator gene ($t = 2.40$, $p = 0.01$). Evolution of these two genes in the artiodactyl lineage might have been under positive selection, but more studies are needed to verify this hypothesis. For the classification according to polarity, there are eight genes with $\lambda(T) > 1$, among which $\lambda(T)$ is significantly greater than 1 for the insulin-like growth factor 1 ($t = 1.87$, $p < 0.05$) and 2 ($t = 2.02$, $p < 0.05$) genes. In addition to these two genes, λ is significantly greater than 1 for the acid phosphatase type 5 gene in the artiodactyl lineage ($t = 1.74$, $p < 0.05$). These three genes may be under positive selection and more studies are needed for verification.

Are λ and ω Correlated?

The nonsynonymous/synonymous rate ratio is a measure of the intensity of purifying selection when there is no involvement of positive selection. It will be interesting to examine whether the radical/conservative rate ratio is related with the selection intensity. We studied the correlation between $\omega(T)$ and $\lambda(T)$ for the 47 genes. For the classification according to charge and that according to polarity and volume, there is a positive correlation between $\omega(T)$ and $\lambda(T)$, with the correlation coefficients being 0.51 ($p < 0.001$) and 0.45 ($p < 0.001$), respectively (Figs. 3 and 4). However, for the classification by polarity, there is no correlation between $\omega(T)$ and $\lambda(T)$ (correlation coefficient $= -0.05$, $p > 0.5$). Even after we remove all the genes whose $\lambda(T)$ values are larger than 1, the correlation between $\omega(T)$ and $\lambda(T)$ is rather weak (correlation coefficient $= 0.37$, $p = 0.016$). It should be pointed out that the slopes of the regression lines between $\omega(T)$ and $\lambda(T)$ are smaller than 1, even though there is a significant correlation when charge or polarity and volume is considered (Figs. 3 and 4). This suggests that $\lambda(T)$ is not so sensitive to intensity of purifying selection as $\omega(T)$.

Is There Evidence for the Nearly Neutral Theory?

By comparing the synonymous and nonsynonymous substitution rates of 49 nuclear genes, Ohta (1995) found

Table 9. The rates of conservative and radical nonsynonymous substitutions relative to the rate of synonymous substitution[a]

	Charge				Polarity				Polarity and volume			
	P	A	R	T	P	A	R	T	P	A	R	T
$\bar{b}_C/\bar{b}_S$	0.31**	0.30**	0.21	0.25	0.29*	0.29**	0.20	0.25	0.46*	0.43**	0.28	0.36
$\bar{b}_R/\bar{b}_S$	0.24	0.24*	0.17	0.21	0.26**	0.25*	0.16	0.21	0.20	0.21	0.15	0.18

[a] The hypotheses of $b_C(P)/b_S(P) > b_C(R)/b_S(R)$ and $b_C(A)/b_S(A) > b_C(R)/b_S(R)$ are examined by one-tail sign tests on the 47 genes and the results are shown in the P and A columns, respectively. Similar hypotheses for radical/synonymous rates are examined. *5% significance; **1% significance.

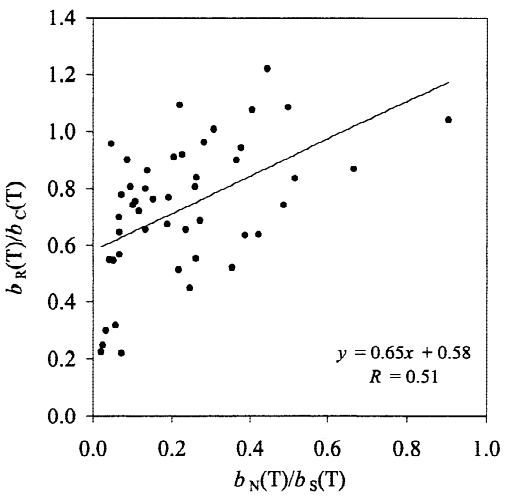

Fig. 3. Correlation between $\bar{b}_R/\bar{b}_C$ and $\bar{b}_N/\bar{b}_S$ for the 47 genes. Classification according to the amino acid charge is considered.

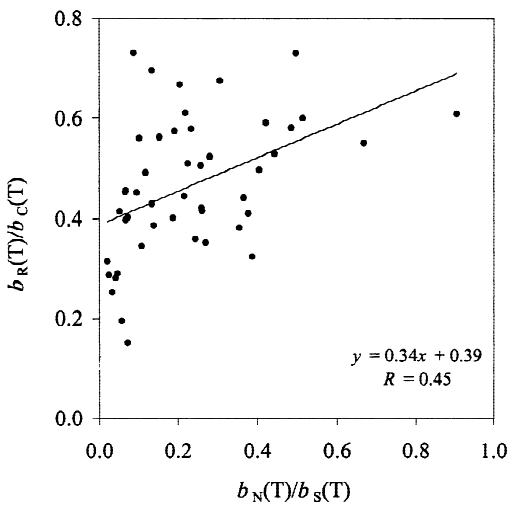

Fig. 4. Correlation between $\bar{b}_R/\bar{b}_C$ and $\bar{b}_N/\bar{b}_S$ for the 47 genes. Classification according to the amino acid polarity and volume is considered.

that the average ω value is higher for primates and artiodactyls than for rodents. In the above, we also showed that the difference in ω among the three lineages is statistically significant. This observation is not consistent with the neutral theory. It has been explained by Ohta (1995) to be consistent with the nearly neutral theory because (1) synonymous substitutions are generally neutral, but most nonsynonymous substitutions are thought to be slightly deleterious, (2) slightly deleterious mutations have a higher probability of fixation in small populations than in large ones, and (3) the effective population size for primates and artiodactyls is generally smaller than that for rodents (but see Nei and Graur 1984). In this context, it is interesting to examine the rates of conservative and radical nonsynonymous substitutions in relation to the rate of synonymous substitution in primates, artiodactyls, and rodents. For all three classifications, we can see that the ratio $\bar{b}_C/\bar{b}_S$ is significantly lower in rodents than in primates and artiodactyls (Table 9). The same pattern is observed for the ratio of $\bar{b}_R/\bar{b}_S$, though the difference between rodents and primates (or artiodactyls) is somewhat smaller. Also, in some cases the difference is not statistically significant (Table 9), which is due partly to the large sampling error of b_R. At any rate, the results suggest that for both conservative and radical nonsynonymous changes, the intensity of purifying selection is reduced in primates and artiodactyls compared to rodents, and the reduction seems to be more apparent for conservative substitutions than radical substitutions.

We have shown that ω and λ are positively correlated when the property of charge or polarity and volume is considered. Therefore, one may expect to see a higher value of λ in primates and artiodactyls than in rodents. However, inconsistent patterns were observed for the relationships of the λ values in the three lineages. That is, we observed $\bar{\lambda}$ (A) > $\bar{\lambda}$ (P) > $\bar{\lambda}$ (R) for charge, $\bar{\lambda}$ (A) > $\bar{\lambda}$ (R) > $\bar{\lambda}$ (P) for polarity, and $\bar{\lambda}$ (A) > $\bar{\lambda}$ (P) = $\bar{\lambda}$ (R) for polarity and volume. In none of these cases was the difference in $\bar{\lambda}$ among lineages statistically significant (Tables 2, 4, and 6). Similarly, inconsistent patterns are observed when $\bar{b}_R/\bar{b}_C$ or weighted $\bar{b}_R/\bar{b}_C$ is compared among the three lineages, and in no comparisons was the difference statistically significant. This unexpected result may be explained by the fact that λ is not as sensitive to the purifying selection intensity as ω, as discussed earlier, and that the sampling variances of b_C and b_R are generally greater than those of b_S and b_N. If this explanation is correct, we might see a higher value of λ(P) and

Table 10. The total numbers of conservative and radical sites for the 47 genes computed and co-workers' using Hughes method and the new method

	Charge			Polarity			Polarity and volume		
	C	*R*	*R/C*	*C*	*R*	*R/C*	*C*	*R*	*R/C*
Method									
Hughes et al.	25332	16903	0.667	27262	14972	0.549	13930	28305	2.032
New	24873	15220	0.612	27239	12853	0.472	13397	26696	1.993
Ratio[a]	1.018	1.111	1.090	1.001	1.165	1.164	1.040	1.060	1.020

[a] The *C, R,* or *R/C* value computed by the Hughes et al. method divided by the corresponding value computed by the new method.

λ(A) than λ(R) when the number of genes studied is increased. Further studies are needed to explore this possibility.

Transition Bias and Estimation of d_C *and* d_R

In this paper, a method that takes into account the transition bias in computation of d_C and d_R has been introduced. This method is expected to give more accurate results than Hughes and co-workers' (1990) method if there is transition bias. However, it is unclear how different these two estimators are for real sequence data. Since the difference in these two methods is the computation of the numbers of conservative (*C*) and radical (*R*) sites, I compared the sums of *C* and *R*, respectively, between the two methods, for the 47 genes. The results are presented in Table 10. For all three classifications, *C* and *R* are overestimated when the transition bias is neglected (in Hughes and co-workers' method). The bias, however, is much more serious for *R* than for *C* in all three classifications. As a result, d_C, d_R, and d_R/d_C are all underestimated in Hughes and co-workers' method. For example, in the classification according to charge, there is on average about 1.8% overestimation of *C*, but 11.1% overestimation of *R*. So, *R*/*C* is overestimated to an extent of 9.0%. This means that d_R/d_C will be underestimated to a similar extent. For the 47 genes used here, the average *q* value computed is approximately 1.5. Nevertheless, it should be pointed out that it is the mutational bias of transitions and transversions that should be considered in computing *C* and *R*, but the *q* values used in the present analysis measure the substitutional bias (Yang and Nielsen 1998). The substitutional bias is expected to be higher than the mutational bias because transversions are more likely to be nonsynonymous and therefore are removed by selection, thus increasing the transition/transversion ratio. Ideally, *q* should be estimated from fourfold degenerate sites only. But this will inevitably increase the sampling error when the DNA sequences considered are not very long. In the method presented here, nucleotide mutations are assumed to occur following the Kimura model. In reality, however, the mutational pattern is much more complex than the assumed model (e.g., Gojobori et al. 1982; Li et al. 1984) and the pattern may vary among DNA regions (Casane et al. 1997) and species. The influence of these factors on the estimation of the conservative and radical rates has not been well studied and is certainly worth exploring in the future (Li et al. 1984).

To study the difference between transitions and transversions in producing conservative and radical changes, we computed the numbers of synonymous, nonsynonymous, conservative nonsynonymous, and radical nonsynonymous mutations that are expected when one transitional mutation or one transversional mutation occurs (Fig. 5). These numbers are estimated under the codon frequencies of the 47 genes analyzed. We can see that

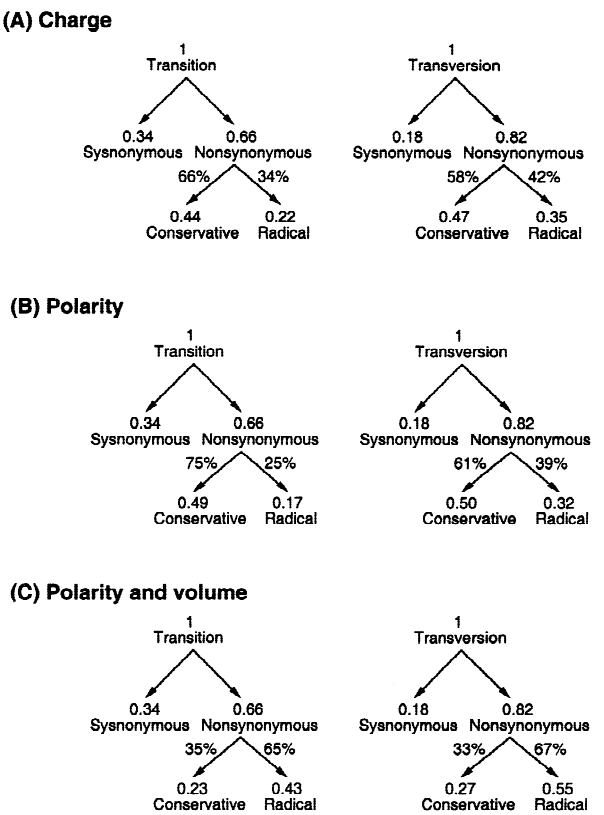

Fig. 5. Expected numbers of synonymous, nonsynonymous, conservative nonsynonymous, and radical nonsynonymous mutations when either one transition or one transversion occurs. The codon frequencies observed from the 47 genes are used in the computation.

one transitional mutation is expected to produce 0.34 synonymous mutations and 0.66 nonsynonymous mutations. Among the latter, 66% are conservative and 34% are radical when charge change is considered. That is, 0.44 conservative and 0.22 radical nonsynonymous mutations are expected (Fig. 5A). In contrast, a transversion causes 0.18 synonymous mutations and 0.82 nonsynonymous mutations. Among the latter, 58% are conservative and 42% are radical. That is, 0.47 conservative and 0.35 radical nonsynonymous mutations are expected (Fig. 5A). Similar patterns are observed for the other two classifications of amino acids (Figs. 5B and C). In sum, (1) a transversion is more likely to cause a nonsynonymous change than a transition, as known before; and (2) a nonsynonymous transversion is more likely to be radical than a nonsynonymous transition. This means that the genetic code table is of such a structure that transversions induce more dramatic changes than transitions at two different levels: they produce more nonsynonymous changes and a higher proportion of radical changes.

Program Availability

A program for computing p_C, p_R, d_C, d_R, and their variances can be downloaded from www.bio.psu.edu/People/Faculty/Nei/Lab/.

Acknowledgments. I thank Tomoko Ohta and Ziheng Yang for providing the alignments of the 47 genes used in this study. I am grateful to Masatoshi Nei, Alex Rooney, Chen Su, and two anonymous reviewers for their comments on an early version of this paper. This work was supported by NIH and NSF research grants to M. Nei.

References

Casane D, Boissinot S, Chang BHJ, Shimmin LC, Li WH (1997) Mutation pattern variation among regions of the primate genome. J Mol Evol 45:216–226

Clarke B (1970) Selective constraints on amino-acid substitutions during the evolution of proteins. Nature 228:159–160

Dayhoff MO, Eck RV, Park CM (1972) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, MD, pp 89–100

Epstein CJ (1966) Role of the amino-acid "code" and of selection for conformation in the evolution of proteins. Nature 210:25–28

Epstein CJ (1967) Non-randomness of amino-acid changes in the evolution of homologous proteins. Nature 215:355–359

Gojobori T, Li WH, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18:360–369

Goldberg AL, Wittes RE (1966) Genetic code: Aspects of organization. Science 153:420–424

Hughes AL (1992) Coevolution of the vertebrate integrin α- and β-chain genes. Mol Biol Evol 9:216–234

Hughes AL (1994) Evolution of cysteine proteinases of eukaryotes. Mol Phylogenet Evol 3:310–321

Hughes AL, Hughes MK (1993) Adaptive evolution in the rat olfactory receptor gene family. J Mol Evol 36:249–254

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major histocompatibility complex genes in mammals. Mol Biol Evol 7:491–514

Ina Y (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J Mol Evol 40:190–226

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, p 21

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitutions. J Mol Evol 36:96–99

Li WH, Wu CI, Luo CC (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitutions in protein evolution. J Mol Evol 12:219–236

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Graur D (1984) Extent of protein polymorphism and the neutral mutation theory. Evol Biol 17:73–118

Ohta, T (1992) The nearly neutral theory of molecular evolution. Annu Rev Syst Ecol 23:263–286

Ohta, T (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J Mol Evol 40: 56–63

Ota T, Nei M (1994) Variance and covariance of the numbers of synonymous and nonsynonymous substitutions per site. Mol Biol Evol 11:613–619

Sokal RR, Rohlf FJ (1995) Biometry. Freeman, New York

Rzhetsky A, Nei M (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol Biol Evol 10: 1073–1095

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Zhang J, Kumar S, Nei M (1997) Small-sample tests of episodic adaptive evolution: A case study of primate lysozyme genes. Mol Biol Evol 14:1335–1338

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA 95:3708–3713

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–116