Genes Genet. Syst. (2007) **82**, p. 341–360

Inferring natural selection operating on conservative and radical substitution at single amino acid sites

Yoshiyuki Suzuki*

Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University

(Received 25 April 2007, accepted 2 July 2007)

Natural selection operating on amino acid substitution at single amino acid sites can be detected by comparing the rates of synonymous (r_S) and nonsynonymous (r_N) nucleotide substitution at single codon sites. Amino acid substitutions can be classified as conservative or radical according to whether they retain the properties of the substituted amino acid. Here methods for comparing the rates of conservative (r_C) and radical (r_R) nonsynonymous substitution with r_S at single codon sites were developed to detect natural selection operating on these substitutions at single amino acid sites. A method for comparing r_C and r_R at single codon sites was also developed to detect biases toward these substitutions at single amino acid sites. Charge was used as the property of the amino acids. In a computer simulation, false-positive rates of these methods were always < 5%, unless termination sites were included in the computation of the numbers of sites and estimates of transition/transversion rate ratio were highly biased. The frequency of detection of natural selection operating on conservative substitution was almost independent of the presence of natural selection operating on radical substitution, and *vice versa*. Natural selection operating specifically on conservative and radical substitution was detected more efficiently by comparing r_S with r_C and r_S with r_R than by comparing r_S with r_N. These methods also appeared to be robust against the occurrence of recombination during evolution. In an analysis of class I human leukocyte antigen, negative selection operating on conservative substitution, but not positive selection operating on radical substitution, was observed at some of the codon sites with $r_R > r_C$, suggesting that $r_R > r_C$ may not necessarily be an indicator of positive selection operating on radical substitution.

Key words: conservative substitution, radical substitution, positive selection, negative selection, single site

INTRODUCTION

Natural selection operating on amino acid substitution can be detected by comparing the rates of synonymous (r_S) and nonsynonymous (r_N) nucleotide substitution. Under the assumption that synonymous substitutions are selectively nearly neutral, $r_N > r_S$ and $r_N < r_S$ indicate positive and negative selection operating on amino acid substitution, respectively (Hughes and Nei, 1988). Since the biological functions of different amino acid sites in a protein usually vary, the direction and magnitude of natural selection operating on them are also considered to vary. The parsimony (Suzuki and Gojobori, 1999), likelihood (Suzuki, 2004; Kosakovsky Pond and Frost, 2005; Massingham and Goldman, 2005), and Bayesian (Yang et al., 2000; Huelsenbeck and Dyer, 2004) methods have been developed for detecting natural selection operating on amino acid substitution at single amino acid sites, by comparing r_S and r_N at single codon sites.

To understand the molecular mechanisms of natural selection, it may be useful to examine patterns of amino acid substitution. Amino acid substitutions, as well as nonsynonymous substitutions, can be classified as conservative or radical according to whether they retain properties of the substituted amino acid such as charge, polarity, hydrophobicity, and volume (Taylor, 1986). It is therefore interesting to compare the rates of conservative (r_C) and radical (r_R) nonsynonymous substitution at single codon sites to detect biases toward these substitutions at single amino acid sites. In addition, since both conservative and radical substitutions are nonsynonymous substitutions, it is possible to detect natural selection operating on these substitutions separately by comparing r_S with r_C

Edited by Yoko Satta

* Corresponding author. E-mail: yis1@psu.edu

and r_S with r_R. $r_C > r_S$ and $r_C < r_S$ indicate positive and negative selection operating on conservative substitution, and $r_R > r_S$ and $r_R < r_S$ indicate positive and negative selection operating on radical substitution.

Inferring natural selection operating on conservative and radical substitution at single amino acid sites may be useful to identify the targets of natural selection (Wong et al., 2006). Practically, amino acid sites in the epitopes of viral proteins where positive selection operates on conservative substitution but not on radical substitution have been proposed as the candidates for immunization (Hanada et al., 2006). In addition, $r_R > r_C$ has been considered as an indicator of positive selection operating on radical substitution under the assumption that radical substitution is more likely to be positively selected than conservative substitution (Zhang, 2000).

Hughes et al. (1990) developed a method for estimating the average numbers of conservative and radical substitutions per site over the entire region of a pair of protein-coding nucleotide sequences. In this method, the numbers of conservative and radical differences as well as those of conservative and radical sites per sequence are computed in a similar manner to the method of Nei and Gojobori (1986), which is intended to estimate the numbers of synonymous and nonsynonymous substitutions. The number of substitutions is computed as the ratio of the number of differences to the number of sites. Hughes et al. (2000) and Zhang (2000) modified this method by taking into account the transition/transversion rate ratio (r_I/r_V). However, Dagan et al. (2002) indicated that results obtained with these methods may be affected by the amino acid composition, r_I/r_V, codon usage bias, and the genetic code, among which the first two appeared to exert the strongest effects. Smith (2003) further modified these methods by taking into account r_I/r_V, r_R/r_C, and equilibrium codon frequencies.

The purpose of the present study was to develop methods for detecting natural selection operating on conservative and radical substitution and for detecting biases toward these substitutions at single amino acid sites. A computer simulation was conducted to examine the performance of these methods. The methods were also applied to the analysis of the class I major histocompatibility complex (MHC) of humans (human leukocyte antigen, HLA).

MATERIALS AND METHODS

Methods In the present study, methods for detecting natural selection operating on conservative and radical substitution and for detecting biases toward these substitutions at single amino acid sites were developed by extending the parsimony method for detecting natural selection operating on amino acid substitution at single amino acid sites. It should be noted that the likelihood method can also be extended for the same purpose by incorporating r_C/r_S, r_R/r_S, and r_R/r_C into the model, and the Bayesian method has been extended in a similar manner (Wong et al., 2006). Charge was used as the property of the amino acids, although other properties such as polarity, hydrophobicity, and volume could also be used. This was because only charge was associated with natural selection in an analysis of class I MHC (Hughes et al., 1990), which was also investigated in the present study. Amino acids were classified as positively charged (histidine [H], lysine [K], and arginine [R]), negatively charged (aspartic acid [D] and glutamic acid [E]), or neutral (alanine [A], cysteine [C], phenylalanine [F], glycine [G], isoleucine [I], leucine [L], methionine [M], asparagine [N], proline [P], glutamine [Q], serine [S], threonine [T], valine [V], tryptophan [W], and tyrosine [Y]).

In the parsimony method, multiple protein-coding nucleotide sequences are compared by taking into account the phylogenetic relationship among them. For each codon site, the total numbers of synonymous (c_S) and nonsynonymous (c_N) differences as well as the average numbers of synonymous (s_S) and nonsynonymous (s_N) sites per codon over the phylogenetic tree are computed according to the maximum parsimony principle (Fitch, 1971). Any pattern of nucleotide mutation can be assumed in the computation of s_S and s_N (Suzuki, 1999). The total numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions per site over the phylogenetic tree are computed as c_S/s_S and c_N/s_N, respectively. The null hypothesis of selective neutrality for amino acid substitution ($r_S = r_N$) is tested by computing the probability (p) of obtaining the observed or more biased values for c_S and c_N, which are assumed to follow a binomial distribution with the probabilities of occurrence of synonymous and nonsynonymous substitution given by $s_S/(s_S + s_N)$ and $s_N/(s_S + s_N)$, respectively. Positive and negative selection are inferred to operate on amino acid substitution when $d_N > d_S$ and $d_N < d_S$ with $p < 0.05$, respectively (two-tailed test).

Natural selection operating on conservative and radical substitution at single amino acid sites was assessed by comparing r_S with r_C and r_S with r_R, and biases toward these substitutions by comparing r_C with r_R at single codon sites. To compare r_C with r_R, the total numbers of conservative (c_C) and radical (c_R) differences as well as the average numbers of conservative (s_C) and radical (s_R) sites per codon over the phylogenetic tree were computed in a similar manner to the parsimony method. It should be noted that $c_C + c_R = c_N$ and $s_C + s_R = s_N$. The total numbers of conservative (d_C) and radical (d_R) substitutions per site over the phylogenetic tree were computed as c_C/s_C and c_R/s_R, respectively. The null hypothesis of no bias toward conservative or radical substitution ($r_C = r_R$) was tested by computing p of obtaining the observed or more biased values for c_C and c_R, which were assumed to follow a binomial distribution with the probabilities of occurrence of conser-

vative and radical substitution given by $s_C/(s_C + s_R)$ and $s_R/(s_C + s_R)$, respectively. Biases toward conservative and radical substitution were inferred when $d_R < d_C$ and $d_R > d_C$ with $p < 0.05$, respectively (two-tailed test). The null hypothesis of selective neutrality for conservative substitution ($r_S = r_C$) was tested by computing p of obtaining the observed or more biased values for c_S and c_C, which were assumed to follow a binomial distribution with the probabilities of occurrence of synonymous and conservative substitution given by $s_S/(s_S + s_C)$ and $s_C/(s_S + s_C)$, respectively. Positive and negative selection operating on conservative substitution were inferred when $d_C > d_S$ and $d_C < d_S$ with $p < 0.05$, respectively (two-tailed test). Positive and negative selection operating on radical substitution were inferred in a similar manner, by replacing r_C, c_C, s_C, and d_C with r_R, c_R, s_R, and d_R. These methods were implemented in the program package ADAPTSITE (version 1.4) (Suzuki et al., 2001), which is available at http://www.bio.psu.edu/People/Faculty/Nei/Lab/software.htm and http://www.cib.nig.ac.jp/dda/yossuzuk/welcome.html.

Since multiple substitutions are not corrected for in these methods, c_S, c_C, c_R, and c_N may be underestimated, especially when the branch lengths of the phylogenetic tree are large. Therefore, these methods are considered to be suitable for the analysis of closely related sequences. It should be noted, however, that the parsimony method is known to be conservative even when the branch lengths are relatively large (Suzuki and Gojobori, 1999; Suzuki and Nei, 2002; Suzuki, 2004; Wong et al., 2004). In the present study, the degree of underestimation for c_S, c_C, c_R, and c_N appeared to be negligible for all the data sets analyzed, because the branch lengths were generally small, as indicated below (Saitou, 1989).

Computer simulation A computer simulation was conducted to examine the performance of the methods developed in the present study. A detailed simulation scheme is described in Suzuki and Gojobori (1999). An ancestral sequence with 300 codon sites was generated using pseudo-random numbers under the assumption that the equilibrium frequencies for 61 sense codons were the same (1/61). This sequence was evolved according to a symmetrical phylogenetic tree until the number of sequences (n) reached 64, 128, or 256. The branch length (b) was set so that the average number of synonymous substitutions per site over all codon sites was expected to be 0.01. r_I/r_V was assumed to be 1. r_C/r_S and r_R/r_S were assumed to be 0.2 (negative selection), 1 (no selection), or 5 (positive selection).

In the analysis of the sequences generated, the phylogenetic relationship among them was assumed to be unknown, and the neighbor-joining method (Saitou and Nei, 1987) with the p distance was used to construct the phylogenetic tree. The p distance is known to be useful for constructing reliable phylogenetic trees when large numbers of closely related sequences are analyzed (Nei and Kumar, 2000; Takahashi and Nei, 2000). Indeed, reliable results were obtained by the parsimony method both in the computer simulation and real data analysis when phylogenetic trees were constructed in this manner (Suzuki, 2004). r_I/r_V was also estimated as the ratio of the transitional/transversional nucleotide diversity at all sites using the two-parameter method (Kimura, 1980). Each scheme was iterated 200 times, so that the total number of codon sites analyzed was 60,000.

The results obtained with the methods developed in the present study may be affected by the amino acid composition, r_I/r_V, codon usage bias, and the genetic code, among which the first two may exert the strongest effects, as indicated above. Therefore, the effects of amino acid composition and r_I/r_V on results were examined. For the former purpose, an ancestral sequence was generated as a repeat of each of 61 sense codons, and was evolved under the assumptions that $n = 256$, $r_C/r_S = r_R/r_S = 1$, and $r_I/r_V = 1$. For the latter purpose, the entire simulation was replicated under the assumption that $r_I/r_V = 10$.

Data analysis The methods developed in the present study were applied to the analysis of class I HLA (HLA-A, -B, and -C). HLA-A, -B, and -C comprise 365, 362, and 366 amino acid sites, respectively. Position 319 of HLA-C is missing from HLA-A and -B, and positions 363–365 of HLA-A, which are homologous to positions 364–366 of HLA-C, are missing from HLA-B. These proteins comprise a signal peptide (positions [–24]–[–1] of HLA-A), three extracellular domains (α1 [positions 1–90], α2 [positions 91–182], and α3 [positions 183–274]), a transmembrane region (positions 275–313), and a cytoplasmic tail (positions 314–341). They are expressed on most adult somatic cells. They bind to an intracellularly processed peptide and present it to $CD8^+$ cytotoxic T lymphocytes (CTLs), together with β_2-microglobulin, to induce cytotoxicity against infected cells. Fifty-seven amino acid sites in the α1 and α2 domains of class I HLA are associated with this process and are called antigen recognition sites (ARSs) (Bjorkman et al., 1987a, 1987b). Parham et al. (1988) divided the ARSs into three groups: the sites that interact with peptides (peptide-binding sites) (29 sites), the sites that interact with T-cell receptors (TCRs) (TCR-binding sites) (18 sites), and the sites that do not interact either with peptides or TCRs (non-binding sites) (10 sites). It has been well-established that positive (overdominant) selection operates on ARSs (Hughes and Nei, 1988). Positive selection has been found to operate mainly on the peptide-binding sites, where $r_R > r_C$ for the entire region, suggesting that positive selection operated on radical substitution (Hughes et al., 1990). In the present study, r_S, r_C, r_R, and r_N were compared at each codon site of class I HLA to examine whether positive

selection operated only on radical substitution and whether $r_R > r_C$ is an indicator of positive selection operating on radical substitution.

Nucleotide sequences encompassing the entire protein-coding region for 93, 164, and 56 alleles of HLA-A, -B, and -C were retrieved from the IMGT/HLA database (release 2.14.1) (Robinson et al., 2003). To estimate r_I/r_V, the ratios of transitional/transversional nucleotide diversity at four-fold degenerate sites were computed for HLA-A, -B, and -C using the two-parameter method. The numbers of four-fold degenerate sites were 168, 183, and 175, and the ratios were 10.68, 1.42, and 4.88 for HLA-A, -B, and -C, respectively. r_I/r_V was estimated as the weighted average of the ratios, with weights proportional to the number of four-fold degenerate sites; its value was 5.53.

It should be noted that HLA-A, -B, and -C may be subject to intralocus and interlocus gene conversion (recombination) during evolution. If this has been the case, the phylogenetic relationship inferred using the entire protein-coding region may not necessarily apply to all codon sites. To examine the effect of recombination on results, a computer simulation was conducted in a similar manner as above, but by using incorrect phylogenetic trees in the analysis. The ancestral sequence (at generation [g] 0) was generated under the assumption that the equilibrium frequencies for 61 sense codons were the same, and was evolved under the assumptions that n = 128 or 256, $r_C/r_S = r_R/r_S$ = 0.2, 1, or 5, and r_I/r_V = 1 or 10. In the correct phylogenetic tree, the number of sequences is doubled at each successive g. If the sequences at g are denoted as 1_g, 2_g, ..., $2^g{}_g$, the topology of the phylogenetic tree can be described such that $(2x + 1)_g$ and $(2x + 2)_g$ (x = 0, 1, ..., $2^{g-1} - 1$) are attached to $(x + 1)_{g-1}$ (g = 1, 2, ..., 7 and 1, 2, ..., 8 for n = 128 and 256, respectively). Incorrect phylogenetic trees were produced by exchanging the positions of $(4x + 1)_g$ and $(4x + 3)_g$ (x = 0, 1, ..., $2^{g-2} - 1$) at each g (g = 2, 3, ..., 7 and 2, 3, ..., 8 for n = 128 and 256, respectively). Positions were exchanged between contemporary and closely related sequences to simulate real recombination events. The topological distance (d_T) between correct and incorrect phylogenetic trees was computed as twice the number of different partitions that were supported by interior branches (Robinson and Foulds, 1981; Penny and Hendy, 1985; Rzhetsky and Nei, 1992).

RESULTS

Results of computer simulation Table 1 summarizes the results of the computer simulation conducted under the assumptions that the equilibrium frequencies for 61 sense codons were the same when generating the ancestral sequence, and r_I/r_V = 1. The estimates of r_I/r_V fluctuated to some extent according to changes in r_C/r_S and r_R/r_S, because the proportions of transition and transversion are different among synonymous, conservative, and radical substitutions (Zhang, 2000). However, the frequency of detecting positive and negative selection operating on conservative and radical substitution always increased with n, while keeping the false-positive rate < 5%. In addition, the frequency of detecting positive and negative selection operating on conservative substitution was not affected to any large extent by the presence of positive and negative selection operating on radical substitution, and *vice versa*. Natural selection operating specifically on conservative and radical substitution was detected more efficiently by comparing r_S with r_C and r_S with r_R than by comparing r_S with r_N, although natural selection operating on both conservative and radical substitution was detected more efficiently by the latter than by the former. The frequency of detecting biases toward conservative and radical substitution also always increased with n and the difference between r_C/r_S and r_R/r_S, while keeping the false-positive rate < 5%. Similar results were obtained when r_I/r_V = 10 (Table 2).

Table 3 summarizes the results of the computer simulation conducted under the assumptions that the ancestral sequence was generated as a repeat of each of 61 sense codons, and r_I/r_V = 1. On average, 78–83% of 256 sequences generated retained the ancestral codon at each site. False-positive rates of detecting positive and negative selection operating on conservative and radical substitution and biases toward these substitutions were > 5% when TCA, TAT, TAC, TGG, CAA, CAG, and GGA were used to generate the ancestral sequence. Interestingly, all of these codons were different from the termination codon (TAA, TAG, or TGA) only at one nucleotide position. Similar results were obtained when r_I/r_V = 10 (Table 4); false-positive rates were > 5% when TGG, CAA, and CAG were used to generate the ancestral sequence.

It has been suggested that, for the codons that are different from the termination codon only at one nucleotide position, the fraction of possible nucleotide mutations that produce a termination codon, which is defined as the termination site in the present study, should be excluded from the computation of s_S and s_N (Yang and Nielsen, 1998). However, termination sites were included in the computation of the numbers of sites (s_S, s_C, s_R, and s_N) in the computer simulation described in Tables 1–4, 7, 8, 11, and 12. To examine the effect of including termination sites on results, a computer simulation was conducted in a similar manner as in Table 3, but by excluding termination sites. False-positive rates decreased dramatically, and were > 5% only for TGG, CAA, and CAG (Table 5). Similar results were obtained when r_I/r_V = 10 (Table 6); false-positive rate was > 5% only for TGG. It should be noted that, when the codons that were different from the termination codon only at one nucleotide position were used for generating the ancestral sequence, estimates of r_I/r_V were often highly biased (Tables 3–6). To

Table 1. Proportions of significant ($p < 0.05$) results in the computer simulation under the assumptions that the equilibrium frequencies for 61 sense codons were the same for generating the ancestral sequence, and $r_I/r_V = 1$

n	r_C/r_S	r_R/r_S	Ratio of transitional/ transversional nucleotide diversity	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
64	0.2	0.2	1.36 ± 0.20	0.00	0.03	0.00	0.02	0.00	0.07	0.00	0.00
	0.2	1	1.04 ± 0.12	0.00	0.03	0.00	0.01	0.00	0.05	0.02	0.00
	0.2	5	0.80 ± 0.06	0.00	0.03	0.15	0.00	0.02	0.02	0.40	0.00
	1	0.2	1.20 ± 0.14	0.00	0.01	0.00	0.01	0.00	0.03	0.00	0.01
	1	1	1.03 ± 0.09	0.00	0.01	0.00	0.00	0.00	0.02	0.01	0.00
	1	5	0.82 ± 0.06	0.00	0.01	0.14	0.00	0.03	0.01	0.21	0.00
	5	0.2	1.12 ± 0.08	0.22	0.00	0.00	0.01	0.11	0.00	0.00	0.25
	5	1	1.04 ± 0.06	0.21	0.00	0.00	0.00	0.12	0.00	0.00	0.16
	5	5	0.88 ± 0.05	0.19	0.00	0.14	0.00	0.20	0.00	0.01	0.01
128	0.2	0.2	1.34 ± 0.19	0.00	0.11	0.00	0.05	0.00	0.21	0.00	0.00
	0.2	1	1.05 ± 0.12	0.00	0.12	0.01	0.01	0.00	0.12	0.08	0.00
	0.2	5	0.81 ± 0.06	0.00	0.11	0.37	0.00	0.09	0.05	0.74	0.00
	1	0.2	1.19 ± 0.11	0.01	0.01	0.00	0.05	0.00	0.06	0.00	0.04
	1	1	1.03 ± 0.09	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01
	1	5	0.83 ± 0.05	0.01	0.02	0.36	0.00	0.12	0.01	0.47	0.00
	5	0.2	1.14 ± 0.06	0.54	0.00	0.00	0.05	0.33	0.01	0.00	0.62
	5	1	1.05 ± 0.06	0.53	0.00	0.01	0.01	0.35	0.00	0.00	0.42
	5	5	0.88 ± 0.04	0.51	0.00	0.38	0.00	0.56	0.00	0.01	0.01
256	0.2	0.2	1.35 ± 0.14	0.00	0.28	0.00	0.13	0.00	0.42	0.01	0.00
	0.2	1	1.03 ± 0.09	0.00	0.29	0.02	0.01	0.00	0.22	0.20	0.00
	0.2	5	0.81 ± 0.05	0.00	0.25	0.64	0.00	0.20	0.09	0.88	0.00
	1	0.2	1.18 ± 0.10	0.02	0.02	0.00	0.13	0.01	0.09	0.00	0.14
	1	1	1.03 ± 0.07	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01
	1	5	0.83 ± 0.05	0.02	0.02	0.65	0.00	0.29	0.01	0.75	0.00
	5	0.2	1.15 ± 0.06	0.84	0.00	0.00	0.13	0.58	0.02	0.00	0.87
	5	1	1.05 ± 0.06	0.84	0.00	0.01	0.01	0.61	0.00	0.00	0.74
	5	5	0.88 ± 0.04	0.84	0.00	0.71	0.00	0.88	0.00	0.02	0.02

examine the effect of using highly biased estimates of r_I/r_V on results, a computer simulation was also conducted in a similar manner as in Tables 3–6, but by using the correct r_I/r_V in the analysis. When termination sites were included in the computation of the numbers of sites, false-positive rates did not decrease significantly (Tables 7 and 8) compared with those in Tables 3 and 4. However, when termination sites were excluded, false-positive rates were always < 5% whether $r_I/r_V = 1$ or 10 (Tables 9 and 10). These results indicated that the effect of including termination sites in the computation of the numbers of sites was greater than that of using highly biased estimates of r_I/r_V for inflating the false-positive rate.

Table 11 summarizes the results of the computer simulation conducted under the assumptions that recombination occurred during evolution, and $r_I/r_V = 1$. False-positive rates increased with d_T, and were occasionally > 5% when $d_T > 128$, which corresponded to the situations that more than about 50% and 25% of all interior branches were incorrect in the phylogenetic trees with $n = 128$ and 256, respectively. Interestingly, however, true-positive rates also always increased with d_T. These results were obtained because c_S, c_C, c_R, and c_N were proportionately overestimated with d_T, while keeping the relationships between them (data not shown). The difference between c_S, c_C, c_R, and c_N became statistically significant as d_T increased, simply due to an increase in the sample size (total number of nucleotide substitutions). Since the degree of overestimation for c_S, c_C, c_R, and c_N is positively correlated with the branch lengths of the phy-

Table 2. Proportions of significant ($p < 0.05$) results in the computer simulation under the assumptions that the equilibrium frequencies for 61 sense codons were the same for generating the ancestral sequence, and $r_I/r_V = 10$

n	r_C/r_S	r_R/r_S	Ratio of transitional/ transversional nucleotide diversity	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
64	0.2	0.2	13.49 ± 2.55	0.00	0.04	0.00	0.02	0.00	0.07	0.00	0.00
	0.2	1	10.25 ± 1.65	0.00	0.04	0.00	0.00	0.00	0.05	0.02	0.00
	0.2	5	6.41 ± 0.60	0.00	0.04	0.15	0.00	0.03	0.03	0.25	0.00
	1	0.2	11.81 ± 1.76	0.00	0.01	0.00	0.01	0.00	0.03	0.00	0.01
	1	1	10.32 ± 1.29	0.00	0.01	0.00	0.00	0.00	0.02	0.01	0.00
	1	5	7.20 ± 0.67	0.00	0.01	0.14	0.00	0.04	0.01	0.14	0.00
	5	0.2	9.33 ± 0.79	0.27	0.00	0.00	0.01	0.17	0.01	0.00	0.19
	5	1	9.18 ± 0.74	0.27	0.00	0.00	0.00	0.18	0.00	0.00	0.11
	5	5	8.17 ± 0.54	0.26	0.00	0.14	0.00	0.28	0.00	0.02	0.01
128	0.2	0.2	13.41 ± 2.41	0.00	0.14	0.00	0.06	0.00	0.23	0.00	0.00
	0.2	1	9.96 ± 1.21	0.00	0.13	0.01	0.01	0.00	0.14	0.05	0.00
	0.2	5	6.21 ± 0.50	0.00	0.12	0.35	0.00	0.11	0.08	0.53	0.00
	1	0.2	11.84 ± 1.49	0.01	0.01	0.00	0.06	0.00	0.06	0.00	0.03
	1	1	10.25 ± 1.02	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01
	1	5	7.06 ± 0.53	0.01	0.01	0.34	0.00	0.15	0.01	0.31	0.00
	5	0.2	9.16 ± 0.67	0.61	0.00	0.00	0.05	0.41	0.01	0.00	0.36
	5	1	8.95 ± 0.58	0.60	0.00	0.01	0.01	0.43	0.00	0.00	0.25
	5	5	7.89 ± 0.46	0.59	0.00	0.35	0.00	0.66	0.00	0.02	0.01
256	0.2	0.2	13.10 ± 1.97	0.00	0.32	0.00	0.16	0.00	0.47	0.01	0.00
	0.2	1	10.09 ± 1.20	0.00	0.31	0.01	0.01	0.00	0.28	0.14	0.00
	0.2	5	6.08 ± 0.42	0.00	0.29	0.56	0.00	0.23	0.16	0.73	0.00
	1	0.2	11.52 ± 1.16	0.02	0.01	0.00	0.15	0.01	0.11	0.00	0.10
	1	1	10.22 ± 0.88	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
	1	5	7.00 ± 0.47	0.01	0.02	0.55	0.00	0.30	0.01	0.53	0.00
	5	0.2	8.84 ± 0.57	0.86	0.00	0.00	0.14	0.62	0.03	0.00	0.51
	5	1	8.79 ± 0.53	0.86	0.00	0.01	0.01	0.67	0.00	0.00	0.42
	5	5	7.79 ± 0.35	0.86	0.00	0.59	0.00	0.93	0.00	0.03	0.01

logenetic tree, the false-positive rate may be < 5% for d_T > 128 and < 128 when $b < 0.01$ and > 0.01, respectively (data not shown). Similar results were obtained when $r_I/r_V = 10$ (Table 12).

Results of data analysis In the analysis of class I HLA, termination sites were excluded from the computation of the numbers of sites. When HLA-A, -B, and -C were analyzed separately, the results obtained at each of the homologous codon sites did not appear to be inconsistent (data not shown). Therefore, the data for HLA-A, -B, and -C were combined to increase the sensitivity of the methods. For each of the homologous codon sites of HLA-A, -B, and -C, c_S was summed to obtain c_S', and s_S was averaged with weights proportional to the total branch length of the phylogenetic tree to obtain s_S'. The total (average) branch lengths of the phylogenetic trees were 0.30 (0.0016), 0.61 (0.0019), and 0.17 (0.0016) for HLA-A, -B, and -C, respectively. c_C' and s_C', c_R' and s_R', and c_N' and s_N' were obtained in a similar manner. Codon positions 363–365 of HLA-A and positions 319 and 364–366 of HLA-C were excluded, so that only the positions that were shared among HLA-A, -B, and -C were analyzed. Positive and negative selection operating on conservative and radical substitution and biases toward these substitutions were assessed as indicated above, by replacing c_S, s_S, c_C, s_C, c_R, s_R, c_N, and s_N with c_S', s_S', c_C', s_C', c_R', s_R', c_N', and s_N'.

Table 13 lists the amino acid sites where significant results were obtained in the class I HLA. Positive selec-

Table 3. Proportions of significant ($p < 0.05$) results in the computer simulation under the assumptions that the ancestral sequence was generated as a repeat of each of 61 sense codons, $n = 256$, $r_C/r_S = r_R/r_S = 1$, and $r_I/r_V = 1$

Codon	Proportion of ancestral codon	Ratio of transitional/ transversional nucleotide diversity	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.78	1.02 ± 0.07	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTC	0.78	1.03 ± 0.07	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTA	0.82	1.48 ± 0.11	0.01	0.02	0.03	0.00	0.01	0.02	0.03	0.00
TTG	0.80	1.19 ± 0.10	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCT	0.78	1.03 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCC	0.78	1.03 ± 0.07	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
TCA	0.83	1.49 ± 0.11	0.00	0.07	0.02	0.00	0.00	0.07	0.03	0.00
TCG	0.80	1.19 ± 0.08	0.01	0.04	0.02	0.00	0.01	0.04	0.02	0.00
TAT	0.82	1.48 ± 0.11	0.17	0.00	0.13	0.00	0.18	0.00	0.01	0.01
TAC	0.82	1.48 ± 0.11	0.16	0.00	0.12	0.00	0.17	0.00	0.01	0.01
TGT	0.80	1.19 ± 0.08	0.01	0.01	0.02	0.00	0.01	0.00	0.02	0.00
TGC	0.80	1.19 ± 0.08	0.01	0.01	0.02	0.00	0.01	0.01	0.02	0.00
TGG	0.82	0.36 ± 0.04	0.00	0.05	0.00	0.02	0.00	0.04	0.11	0.00
CTT	0.78	1.00 ± 0.07	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
CTC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
CTA	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTG	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01
CCC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01
CCA	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCG	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CAT	0.78	0.99 ± 0.08	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.02
CAC	0.78	1.01 ± 0.08	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.03
CAA	0.80	0.66 ± 0.05	0.00	0.05	0.00	0.08	0.00	0.08	0.01	0.03
CAG	0.80	0.67 ± 0.05	0.00	0.05	0.00	0.08	0.00	0.08	0.01	0.03
CGT	0.78	1.01 ± 0.07	0.02	0.00	0.01	0.02	0.01	0.02	0.01	0.02
CGC	0.78	1.01 ± 0.07	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGA	0.80	0.67 ± 0.06	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGG	0.78	0.99 ± 0.06	0.02	0.00	0.02	0.02	0.01	0.02	0.00	0.02
ATT	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATA	0.78	1.00 ± 0.07	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.01
ATG	0.78	1.00 ± 0.07	0.00	0.02	0.00	0.02	0.00	0.02	0.02	0.01
ACT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACC	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACA	0.78	1.00 ± 0.08	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
ACG	0.78	1.00 ± 0.07	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
AAT	0.78	1.00 ± 0.07	0.00	0.02	0.00	0.02	0.00	0.02	0.01	0.02
AAC	0.78	1.01 ± 0.08	0.00	0.02	0.00	0.02	0.00	0.03	0.01	0.02
AAA	0.80	1.18 ± 0.08	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.03
AAG	0.80	1.17 ± 0.09	0.00	0.01	0.00	0.03	0.00	0.03	0.00	0.03
AGT	0.78	0.99 ± 0.08	0.00	0.02	0.00	0.02	0.00	0.03	0.01	0.01
AGC	0.78	1.01 ± 0.08	0.00	0.02	0.00	0.02	0.00	0.03	0.02	0.01
AGA	0.80	1.17 ± 0.09	0.03	0.00	0.02	0.01	0.02	0.01	0.00	0.03
AGG	0.78	0.99 ± 0.07	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.03
GTT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTA	0.78	1.00 ± 0.07	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GTG	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCC	0.78	1.01 ± 0.07	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GCA	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.78	1.02 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GAT	0.78	1.01 ± 0.07	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAC	0.78	1.01 ± 0.07	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAA	0.80	1.19 ± 0.09	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.05
GAG	0.80	1.19 ± 0.08	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.05
GGT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
GGC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
GGA	0.80	1.17 ± 0.09	0.02	0.01	0.00	0.05	0.01	0.04	0.00	0.06
GGG	0.78	0.99 ± 0.06	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01

Table 4. Proportions of significant ($p < 0.05$) results in the computer simulation under the assumptions that the ancestral sequence was generated as a repeat of each of 61 sense codons, $n = 256$, $r_C/r_S = r_R/r_S = 1$, and $r_I/r_V = 10$

Codon	Proportion of ancestral codon	Ratio of transitional/ transversional nucleotide diversity	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.79	10.03 ± 0.88	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTC	0.79	10.04 ± 0.91	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTA	0.80	14.42 ± 1.38	0.02	0.01	0.02	0.00	0.02	0.01	0.02	0.00
TTG	0.79	11.95 ± 1.05	0.02	0.01	0.01	0.00	0.02	0.02	0.01	0.00
TCT	0.79	10.19 ± 0.99	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
TCC	0.79	10.23 ± 0.88	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCA	0.80	14.70 ± 1.47	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCG	0.79	12.03 ± 1.16	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
TAT	0.80	14.38 ± 1.34	0.03	0.01	0.03	0.01	0.03	0.01	0.01	0.02
TAC	0.80	14.75 ± 1.50	0.03	0.01	0.02	0.01	0.03	0.01	0.01	0.02
TGT	0.79	11.96 ± 1.14	0.02	0.01	0.02	0.01	0.02	0.01	0.01	0.01
TGC	0.79	12.13 ± 1.20	0.02	0.01	0.02	0.01	0.02	0.01	0.01	0.02
TGG	0.90	3.52 ± 0.35	0.00	0.12	0.00	0.02	0.00	0.06	0.53	0.00
CTT	0.79	10.06 ± 0.91	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTC	0.79	10.14 ± 0.89	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTA	0.79	10.12 ± 0.85	0.02	0.01	0.01	0.00	0.02	0.01	0.01	0.00
CTG	0.79	10.09 ± 0.95	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
CCT	0.79	10.12 ± 0.84	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCC	0.79	10.06 ± 0.86	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
CCA	0.79	10.25 ± 0.91	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
CCG	0.79	10.25 ± 0.87	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CAT	0.79	10.02 ± 0.92	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
CAC	0.79	10.21 ± 0.98	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
CAA	0.84	6.45 ± 0.58	0.01	0.00	0.00	0.14	0.00	0.13	0.00	0.04
CAG	0.84	6.65 ± 0.64	0.01	0.00	0.00	0.14	0.00	0.13	0.00	0.04
CGT	0.79	10.18 ± 0.83	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGC	0.79	10.24 ± 0.92	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGA	0.84	6.67 ± 0.53	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGG	0.79	9.65 ± 0.87	0.00	0.01	0.01	0.02	0.01	0.02	0.01	0.01
ATT	0.79	10.00 ± 0.85	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATC	0.79	10.02 ± 0.88	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATA	0.79	10.01 ± 0.83	0.00	0.03	0.00	0.00	0.00	0.03	0.02	0.00
ATG	0.79	10.08 ± 0.95	0.00	0.02	0.00	0.00	0.00	0.02	0.02	0.00
ACT	0.79	10.03 ± 0.84	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACC	0.79	9.99 ± 0.90	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACA	0.79	10.16 ± 0.86	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACG	0.79	10.18 ± 0.89	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AAT	0.79	10.09 ± 0.88	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01
AAC	0.79	10.08 ± 0.91	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02
AAA	0.79	11.83 ± 1.07	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AAG	0.80	11.89 ± 1.15	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AGT	0.79	10.09 ± 0.89	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGC	0.79	10.14 ± 0.77	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGA	0.79	11.82 ± 1.14	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AGG	0.79	10.19 ± 0.92	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
GTT	0.79	10.00 ± 0.87	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GTC	0.79	10.08 ± 0.93	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GTA	0.79	10.16 ± 0.91	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTG	0.79	10.17 ± 0.83	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.79	10.13 ± 0.84	0.01	0.02	0.01	0.00	0.02	0.02	0.02	0.00
GCC	0.79	10.12 ± 0.85	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GCA	0.79	10.23 ± 0.86	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.79	10.29 ± 0.97	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GAT	0.79	10.06 ± 0.94	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAC	0.79	10.16 ± 0.91	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAA	0.79	12.03 ± 1.12	0.03	0.00	0.01	0.02	0.01	0.02	0.00	0.04
GAG	0.79	11.99 ± 1.13	0.03	0.00	0.01	0.02	0.01	0.02	0.00	0.04
GGT	0.79	10.20 ± 0.85	0.02	0.01	0.02	0.01	0.02	0.02	0.01	0.01
GGC	0.79	10.11 ± 0.81	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
GGA	0.79	12.04 ± 1.18	0.03	0.00	0.01	0.02	0.01	0.02	0.00	0.03
GGG	0.79	10.30 ± 0.79	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02

Table 5. Proportions of significant ($p < 0.05$) results in the computer simulation under the same assumptions as in Table 3, except that termination sites were excluded from the computation of s_S, s_C, s_R, and s_N

Codon	Proportion of ancestral codon	Ratio of transitional/ transversional nucleotide diversity	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.78	1.02 ± 0.07	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTC	0.78	1.03 ± 0.07	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTA	0.82	1.48 ± 0.11	0.03	0.01	0.03	0.00	0.03	0.01	0.02	0.00
TTG	0.80	1.19 ± 0.10	0.01	0.01	0.02	0.00	0.01	0.01	0.02	0.00
TCT	0.78	1.03 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCC	0.78	1.03 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCA	0.83	1.49 ± 0.11	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCG	0.80	1.19 ± 0.08	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TAT	0.82	1.48 ± 0.11	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01
TAC	0.82	1.48 ± 0.11	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01
TGT	0.80	1.19 ± 0.08	0.00	0.02	0.00	0.01	0.00	0.02	0.01	0.00
TGC	0.80	1.19 ± 0.08	0.00	0.02	0.00	0.01	0.00	0.02	0.01	0.00
TGG	0.82	0.36 ± 0.04	0.00	0.04	0.00	0.02	0.00	0.04	0.07	0.00
CTT	0.78	1.00 ± 0.07	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
CTC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
CTA	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTG	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01
CCC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
CCA	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCG	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CAT	0.78	0.99 ± 0.08	0.00	0.01	0.00	0.03	0.00	0.03	0.00	0.02
CAC	0.78	1.01 ± 0.08	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.02
CAA	0.80	0.66 ± 0.05	0.00	0.05	0.00	0.06	0.00	0.07	0.01	0.02
CAG	0.80	0.67 ± 0.05	0.00	0.05	0.00	0.06	0.00	0.07	0.01	0.02
CGT	0.78	1.01 ± 0.07	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGC	0.78	1.01 ± 0.07	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGA	0.80	0.67 ± 0.06	0.02	0.00	0.02	0.02	0.01	0.02	0.00	0.02
CGG	0.78	0.99 ± 0.06	0.02	0.00	0.02	0.02	0.02	0.02	0.00	0.02
ATT	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATA	0.78	1.00 ± 0.07	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
ATG	0.78	1.00 ± 0.07	0.00	0.02	0.00	0.02	0.00	0.02	0.02	0.01
ACT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACC	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACA	0.78	1.00 ± 0.08	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
ACG	0.78	1.00 ± 0.07	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
AAT	0.78	1.00 ± 0.07	0.00	0.02	0.00	0.02	0.00	0.03	0.01	0.02
AAC	0.78	1.01 ± 0.08	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
AAA	0.80	1.18 ± 0.08	0.00	0.01	0.00	0.02	0.00	0.02	0.00	0.02
AAG	0.80	1.17 ± 0.09	0.00	0.01	0.00	0.02	0.00	0.02	0.01	0.02
AGT	0.78	0.99 ± 0.08	0.00	0.02	0.00	0.02	0.00	0.03	0.02	0.01
AGC	0.78	1.01 ± 0.08	0.00	0.03	0.00	0.02	0.00	0.03	0.02	0.01
AGA	0.80	1.17 ± 0.09	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AGG	0.78	0.99 ± 0.07	0.02	0.01	0.01	0.02	0.01	0.02	0.00	0.03
GTT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTA	0.78	1.00 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTG	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCC	0.78	1.01 ± 0.07	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GCA	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.78	1.02 ± 0.07	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GAT	0.78	1.01 ± 0.07	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAC	0.78	1.01 ± 0.07	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAA	0.80	1.19 ± 0.09	0.01	0.01	0.00	0.02	0.00	0.02	0.01	0.03
GAG	0.80	1.19 ± 0.08	0.01	0.01	0.00	0.02	0.00	0.02	0.01	0.03
GGT	0.78	1.01 ± 0.07	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
GGC	0.78	1.01 ± 0.07	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
GGA	0.80	1.17 ± 0.09	0.02	0.01	0.01	0.02	0.01	0.02	0.01	0.02
GGG	0.78	0.99 ± 0.06	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01

Table 6. Proportions of significant ($p < 0.05$) results in the computer simulation under the same assumptions as in Table 4, except that termination sites were excluded from the computation of s_S, s_C, s_R, and s_N

Codon	Proportion of ancestral codon	Ratio of transitional/ transversional nucleotide diversity	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.79	10.03 ± 0.88	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTC	0.79	10.04 ± 0.91	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTA	0.80	14.42 ± 1.38	0.03	0.01	0.02	0.00	0.03	0.01	0.02	0.00
TTG	0.79	11.95 ± 1.05	0.02	0.01	0.01	0.00	0.02	0.01	0.01	0.00
TCT	0.79	10.19 ± 0.99	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
TCC	0.79	10.23 ± 0.88	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCA	0.80	14.70 ± 1.47	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCG	0.79	12.03 ± 1.16	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
TAT	0.80	14.38 ± 1.34	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02
TAC	0.80	14.75 ± 1.50	0.02	0.01	0.01	0.01	0.02	0.02	0.01	0.02
TGT	0.79	11.96 ± 1.14	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.02
TGC	0.79	12.13 ± 1.20	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02
TGG	0.90	3.52 ± 0.35	0.00	0.05	0.00	0.02	0.00	0.03	0.11	0.00
CTT	0.79	10.06 ± 0.91	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTC	0.79	10.14 ± 0.89	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTA	0.79	10.12 ± 0.85	0.02	0.01	0.02	0.00	0.02	0.01	0.02	0.00
CTG	0.79	10.09 ± 0.95	0.02	0.02	0.02	0.00	0.02	0.01	0.02	0.00
CCT	0.79	10.12 ± 0.84	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCC	0.79	10.06 ± 0.86	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
CCA	0.79	10.25 ± 0.91	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCG	0.79	10.25 ± 0.87	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CAT	0.79	10.02 ± 0.92	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01
CAC	0.79	10.21 ± 0.98	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
CAA	0.84	6.45 ± 0.58	0.01	0.00	0.01	0.03	0.01	0.04	0.00	0.01
CAG	0.84	6.65 ± 0.64	0.01	0.00	0.01	0.03	0.01	0.04	0.00	0.01
CGT	0.79	10.18 ± 0.83	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGC	0.79	10.24 ± 0.92	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGA	0.84	6.67 ± 0.53	0.01	0.00	0.01	0.02	0.01	0.02	0.00	0.01
CGG	0.79	9.65 ± 0.87	0.01	0.00	0.01	0.02	0.01	0.02	0.00	0.02
ATT	0.79	10.00 ± 0.85	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATC	0.79	10.02 ± 0.88	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATA	0.79	10.01 ± 0.83	0.00	0.03	0.00	0.00	0.00	0.03	0.02	0.00
ATG	0.79	10.08 ± 0.95	0.00	0.02	0.00	0.00	0.00	0.02	0.02	0.00
ACT	0.79	10.03 ± 0.84	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACC	0.79	9.99 ± 0.90	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACA	0.79	10.16 ± 0.86	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACG	0.79	10.18 ± 0.89	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AAT	0.79	10.09 ± 0.88	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01
AAC	0.79	10.08 ± 0.91	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01
AAA	0.79	11.83 ± 1.07	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01
AAG	0.80	11.89 ± 1.15	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
AGT	0.79	10.09 ± 0.89	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGC	0.79	10.14 ± 0.77	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGA	0.79	11.82 ± 1.14	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01
AGG	0.79	10.19 ± 0.92	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
GTT	0.79	10.00 ± 0.87	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GTC	0.79	10.08 ± 0.93	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GTA	0.79	10.16 ± 0.91	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTG	0.79	10.17 ± 0.83	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.79	10.13 ± 0.84	0.01	0.02	0.01	0.00	0.02	0.02	0.02	0.00
GCC	0.79	10.12 ± 0.85	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GCA	0.79	10.23 ± 0.86	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.79	10.29 ± 0.97	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GAT	0.79	10.06 ± 0.94	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAC	0.79	10.16 ± 0.91	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAA	0.79	12.03 ± 1.12	0.03	0.00	0.01	0.02	0.01	0.01	0.00	0.03
GAG	0.79	11.99 ± 1.13	0.03	0.00	0.01	0.02	0.01	0.02	0.00	0.03
GGT	0.79	10.20 ± 0.85	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.01
GGC	0.79	10.11 ± 0.81	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
GGA	0.79	12.04 ± 1.18	0.03	0.00	0.01	0.02	0.01	0.02	0.00	0.03
GGG	0.79	10.30 ± 0.79	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02

Table 7. Proportions of significant ($p < 0.05$) results in the computer simulation under the same assumptions as in Table 3, except that the correct r_I/r_V was used in the analysis

Codon	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTC	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTA	0.00	0.07	0.02	0.00	0.00	0.07	0.03	0.00
TTG	0.00	0.04	0.02	0.00	0.00	0.04	0.02	0.00
TCT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCC	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
TCA	0.00	0.08	0.02	0.00	0.00	0.08	0.03	0.00
TCG	0.01	0.04	0.02	0.00	0.01	0.04	0.02	0.00
TAT	0.15	0.00	0.14	0.00	0.17	0.00	0.02	0.01
TAC	0.15	0.00	0.14	0.00	0.17	0.00	0.02	0.01
TGT	0.01	0.01	0.02	0.00	0.01	0.01	0.03	0.00
TGC	0.01	0.01	0.02	0.00	0.01	0.01	0.03	0.00
TGG	0.00	0.03	0.00	0.02	0.00	0.03	0.06	0.00
CTT	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
CTC	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
CTA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCT	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01
CCC	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01
CCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CAT	0.01	0.01	0.00	0.02	0.00	0.03	0.00	0.02
CAC	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.03
CAA	0.00	0.02	0.00	0.04	0.00	0.04	0.00	0.04
CAG	0.00	0.02	0.00	0.05	0.00	0.04	0.00	0.04
CGT	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGC	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGA	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGG	0.02	0.00	0.02	0.02	0.01	0.02	0.00	0.02
ATT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATA	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.01
ATG	0.00	0.02	0.00	0.02	0.00	0.02	0.02	0.01
ACT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACA	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
ACG	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
AAT	0.00	0.02	0.00	0.02	0.00	0.02	0.01	0.02
AAC	0.00	0.02	0.00	0.02	0.00	0.03	0.01	0.02
AAA	0.01	0.01	0.00	0.04	0.00	0.04	0.00	0.04
AAG	0.00	0.01	0.00	0.04	0.00	0.04	0.00	0.04
AGT	0.00	0.02	0.00	0.02	0.00	0.02	0.02	0.01
AGC	0.00	0.02	0.00	0.02	0.00	0.03	0.02	0.01
AGA	0.03	0.00	0.01	0.01	0.02	0.01	0.00	0.03
AGG	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
GTT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTA	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GTG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCC	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GAT	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAC	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAA	0.00	0.02	0.00	0.04	0.00	0.04	0.00	0.04
GAG	0.00	0.02	0.00	0.04	0.00	0.04	0.00	0.04
GGT	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
GGC	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
GGA	0.02	0.01	0.00	0.05	0.01	0.04	0.00	0.04
GGG	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01

Table 8. Proportions of significant ($p < 0.05$) results in the computer simulation under the same assumptions as in Table 4, except that the correct r_I/r_V was used in the analysis

Codon	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTA	0.01	0.03	0.01	0.00	0.01	0.03	0.01	0.00
TTG	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
TCT	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
TCC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCA	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
TCG	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
TAT	0.03	0.01	0.03	0.01	0.03	0.01	0.01	0.02
TAC	0.03	0.01	0.02	0.01	0.03	0.01	0.01	0.02
TGT	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
TGC	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
TGG	0.00	0.12	0.00	0.02	0.00	0.06	0.47	0.00
CTT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTA	0.02	0.01	0.01	0.00	0.02	0.01	0.01	0.00
CTG	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
CCT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCC	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
CCA	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
CCG	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
CAT	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
CAC	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
CAA	0.02	0.00	0.00	0.10	0.00	0.10	0.00	0.06
CAG	0.02	0.00	0.00	0.10	0.00	0.10	0.00	0.06
CGT	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGC	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGA	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGG	0.00	0.01	0.01	0.02	0.01	0.02	0.01	0.01
ATT	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATC	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATA	0.00	0.03	0.00	0.00	0.00	0.03	0.02	0.00
ATG	0.00	0.02	0.00	0.00	0.00	0.02	0.02	0.00
ACT	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACC	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AAT	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01
AAC	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01
AAA	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AAG	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AGT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGA	0.02	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AGG	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
GTT	0.01	0.02	0.01	0.00	0.02	0.02	0.02	0.00
GTC	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GTA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.02	0.02	0.01	0.00	0.02	0.02	0.01	0.00
GCC	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GAT	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAC	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAA	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.03
GAG	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.03
GGT	0.02	0.01	0.02	0.01	0.02	0.02	0.01	0.01
GGC	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
GGA	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GGG	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02

Table 9. Proportions of significant ($p < 0.05$) results in the computer simulation under the same assumptions as in Table 5, except that the correct r_I/r_V was used in the analysis

Codon	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTC	0.00	0.03	0.01	0.00	0.00	0.03	0.02	0.00
TTA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TAT	0.00	0.02	0.00	0.02	0.00	0.02	0.02	0.01
TAC	0.00	0.02	0.00	0.02	0.00	0.03	0.02	0.01
TGT	0.00	0.03	0.00	0.01	0.00	0.03	0.02	0.00
TGC	0.00	0.03	0.00	0.01	0.00	0.03	0.02	0.00
TGG	0.00	0.02	0.00	0.02	0.00	0.02	0.02	0.01
CTT	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
CTC	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
CTA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCT	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01
CCC	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
CCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CAT	0.00	0.01	0.00	0.03	0.00	0.03	0.00	0.02
CAC	0.01	0.01	0.00	0.03	0.00	0.03	0.00	0.03
CAA	0.00	0.02	0.00	0.02	0.00	0.02	0.01	0.02
CAG	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
CGT	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGC	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGA	0.02	0.00	0.02	0.02	0.01	0.02	0.00	0.02
CGG	0.02	0.00	0.02	0.02	0.02	0.02	0.00	0.02
ATT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ATA	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
ATG	0.00	0.02	0.00	0.02	0.00	0.02	0.02	0.01
ACT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACA	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
ACG	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
AAT	0.00	0.02	0.00	0.02	0.00	0.03	0.01	0.02
AAC	0.00	0.02	0.00	0.02	0.00	0.03	0.01	0.02
AAA	0.00	0.01	0.00	0.03	0.00	0.02	0.00	0.02
AAG	0.00	0.01	0.00	0.03	0.00	0.03	0.00	0.02
AGT	0.00	0.02	0.00	0.02	0.00	0.03	0.02	0.01
AGC	0.00	0.03	0.00	0.02	0.00	0.03	0.02	0.01
AGA	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.03
AGG	0.02	0.01	0.01	0.02	0.01	0.02	0.00	0.02
GTT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCC	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
GAT	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAC	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GAA	0.00	0.02	0.00	0.02	0.00	0.03	0.01	0.02
GAG	0.00	0.02	0.00	0.03	0.00	0.03	0.01	0.02
GGT	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
GGC	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01
GGA	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.02
GGG	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01

Table 10. Proportions of significant ($p < 0.05$) results in the computer simulation under the same assumptions as in Table 6, except that the correct r_I/r_V was used in the analysis

Codon	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
TTT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TTA	0.02	0.01	0.01	0.00	0.02	0.01	0.01	0.00
TTG	0.02	0.01	0.01	0.00	0.02	0.01	0.01	0.00
TCT	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
TCC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
TCA	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
TCG	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
TAT	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.02
TAC	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02
TGT	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
TGC	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
TGG	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.02
CTT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CTA	0.02	0.01	0.02	0.00	0.02	0.01	0.02	0.00
CTG	0.02	0.02	0.02	0.00	0.02	0.02	0.02	0.00
CCT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCC	0.01	0.02	0.02	0.00	0.02	0.02	0.02	0.00
CCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CCG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
CAT	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
CAC	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
CAA	0.02	0.00	0.01	0.02	0.01	0.01	0.00	0.02
CAG	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
CGT	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGC	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
CGA	0.01	0.00	0.01	0.02	0.01	0.02	0.00	0.01
CGG	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
ATT	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATC	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ATA	0.00	0.03	0.00	0.00	0.00	0.03	0.02	0.00
ATG	0.00	0.02	0.00	0.00	0.00	0.02	0.02	0.00
ACT	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACC	0.01	0.02	0.01	0.00	0.01	0.02	0.01	0.00
ACA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
ACG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AAT	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01
AAC	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.01
AAA	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AAG	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AGT	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGC	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
AGA	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
AGG	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.02
GTT	0.01	0.02	0.01	0.00	0.02	0.02	0.02	0.00
GTC	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GTA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GTG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCT	0.02	0.02	0.01	0.00	0.02	0.02	0.01	0.00
GCC	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00
GCA	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GCG	0.01	0.02	0.02	0.00	0.01	0.02	0.02	0.00
GAT	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAC	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GAA	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.03
GAG	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.03
GGT	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.01
GGC	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
GGA	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02
GGG	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.02

Table 11. Proportions of significant ($p < 0.05$) results in the computer simulation under the assumptions that recombination occurred during evolution, and $r_I/r_V = 1$

n	r_C/r_S r_R/r_S	g at which recombination occurred	d_T	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
128	0.2	None	0	0.00	0.11	0.00	0.05	0.00	0.21	0.00	0.00
		2	2	0.00	0.12	0.00	0.05	0.00	0.21	0.00	0.00
		3	8	0.00	0.12	0.00	0.05	0.00	0.22	0.00	0.00
		4	16	0.00	0.12	0.00	0.05	0.00	0.22	0.00	0.00
		5	32	0.00	0.13	0.00	0.06	0.00	0.23	0.00	0.00
		6	64	0.00	0.15	0.00	0.07	0.00	0.25	0.01	0.00
		7	128	0.00	0.18	0.00	0.08	0.00	0.29	0.01	0.01
		All	250	0.00	0.23	0.00	0.11	0.00	0.34	0.02	0.01
	1	None	0	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01
		2	2	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01
		3	8	0.01	0.02	0.01	0.01	0.00	0.02	0.01	0.01
		4	16	0.01	0.02	0.01	0.01	0.00	0.03	0.01	0.01
		5	32	0.01	0.02	0.01	0.01	0.01	0.03	0.01	0.01
		6	64	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.01
		7	128	0.02	0.03	0.02	0.02	0.01	0.04	0.03	0.02
		All	250	0.03	0.05	0.03	0.03	0.03	0.06	0.03	0.03
	5	None	0	0.51	0.00	0.38	0.00	0.55	0.00	0.01	0.01
		2	2	0.51	0.00	0.38	0.00	0.56	0.00	0.01	0.01
		3	8	0.51	0.00	0.38	0.00	0.56	0.00	0.01	0.01
		4	16	0.51	0.00	0.39	0.00	0.56	0.00	0.01	0.02
		5	32	0.52	0.00	0.40	0.00	0.57	0.00	0.02	0.02
		6	64	0.54	0.00	0.41	0.00	0.58	0.00	0.02	0.02
		7	128	0.58	0.00	0.46	0.00	0.62	0.00	0.03	0.03
		All	250	0.64	0.00	0.52	0.00	0.68	0.00	0.04	0.04
256	0.2	None	0	0.00	0.28	0.00	0.13	0.00	0.41	0.01	0.00
		2	2	0.00	0.28	0.00	0.13	0.00	0.42	0.01	0.00
		3	8	0.00	0.28	0.00	0.13	0.00	0.42	0.01	0.00
		4	16	0.00	0.28	0.00	0.13	0.00	0.42	0.01	0.00
		5	32	0.00	0.29	0.00	0.14	0.00	0.42	0.01	0.00
		6	64	0.00	0.30	0.00	0.14	0.00	0.43	0.01	0.01
		7	128	0.00	0.32	0.00	0.16	0.00	0.45	0.01	0.01
		8	256	0.00	0.35	0.00	0.19	0.00	0.48	0.02	0.01
		All	506	0.00	0.41	0.00	0.23	0.00	0.53	0.03	0.02
	1	None	0	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01
		2	2	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01
		3	8	0.02	0.02	0.02	0.01	0.01	0.02	0.01	0.01
		4	16	0.02	0.02	0.02	0.01	0.01	0.03	0.02	0.01
		5	32	0.02	0.02	0.02	0.01	0.02	0.03	0.02	0.01
		6	64	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02
		7	128	0.02	0.03	0.02	0.02	0.02	0.04	0.02	0.02
		8	256	0.03	0.04	0.04	0.03	0.03	0.05	0.04	0.03
		All	506	0.05	0.05	0.05	0.04	0.05	0.06	0.05	0.04
	5	None	0	0.83	0.00	0.71	0.00	0.87	0.00	0.02	0.02
		2	2	0.83	0.00	0.71	0.00	0.87	0.00	0.02	0.02
		3	8	0.83	0.00	0.71	0.00	0.87	0.00	0.02	0.02
		4	16	0.83	0.00	0.71	0.00	0.87	0.00	0.02	0.02
		5	32	0.83	0.00	0.71	0.00	0.87	0.00	0.02	0.02
		6	64	0.83	0.00	0.71	0.00	0.87	0.00	0.02	0.02
		7	128	0.84	0.00	0.72	0.00	0.87	0.00	0.02	0.03
		8	256	0.86	0.00	0.75	0.00	0.88	0.00	0.03	0.04
		All	506	0.88	0.00	0.78	0.00	0.90	0.00	0.04	0.05

Table 12. Proportions of significant ($p < 0.05$) results in the computer simulation under the assumptions that recombination occurred during evolution, and $r_I/r_V = 10$

n	r_C/r_S r_R/r_S	g at which recombination occurred	d_T	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
128	0.2	None	0	0.00	0.14	0.00	0.06	0.00	0.23	0.00	0.00
		2	2	0.00	0.14	0.00	0.06	0.00	0.23	0.00	0.00
		3	8	0.00	0.14	0.00	0.06	0.00	0.23	0.00	0.00
		4	16	0.00	0.15	0.00	0.07	0.00	0.24	0.00	0.00
		5	32	0.00	0.16	0.00	0.07	0.00	0.25	0.01	0.00
		6	64	0.00	0.17	0.00	0.08	0.00	0.27	0.01	0.00
		7	128	0.00	0.21	0.00	0.10	0.00	0.31	0.01	0.00
		All	250	0.00	0.26	0.00	0.14	0.00	0.38	0.02	0.01
	1	None	0	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01
		2	2	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01
		3	8	0.01	0.01	0.01	0.01	0.00	0.02	0.01	0.01
		4	16	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01
		5	32	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.01
		6	64	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.01
		7	128	0.02	0.03	0.02	0.02	0.02	0.04	0.03	0.02
		All	250	0.04	0.04	0.03	0.02	0.03	0.06	0.04	0.02
	5	None	0	0.59	0.00	0.35	0.00	0.66	0.00	0.02	0.01
		2	2	0.59	0.00	0.35	0.00	0.66	0.00	0.02	0.01
		3	8	0.59	0.00	0.35	0.00	0.66	0.00	0.02	0.01
		4	16	0.60	0.00	0.35	0.00	0.66	0.00	0.03	0.01
		5	32	0.60	0.00	0.36	0.00	0.67	0.00	0.03	0.01
		6	64	0.61	0.00	0.38	0.00	0.68	0.00	0.03	0.01
		7	128	0.66	0.00	0.42	0.00	0.72	0.00	0.04	0.02
		All	250	0.71	0.00	0.47	0.00	0.78	0.00	0.05	0.02
256	0.2	None	0	0.00	0.31	0.00	0.16	0.00	0.47	0.01	0.00
		2	2	0.00	0.32	0.00	0.16	0.00	0.47	0.01	0.00
		3	8	0.00	0.32	0.00	0.16	0.00	0.47	0.01	0.00
		4	16	0.00	0.32	0.00	0.16	0.00	0.47	0.01	0.00
		5	32	0.00	0.33	0.00	0.17	0.00	0.48	0.01	0.00
		6	64	0.00	0.33	0.00	0.17	0.00	0.49	0.01	0.00
		7	128	0.00	0.35	0.00	0.19	0.00	0.50	0.01	0.01
		8	256	0.00	0.39	0.00	0.21	0.00	0.54	0.02	0.01
		All	506	0.00	0.44	0.00	0.25	0.00	0.59	0.03	0.02
	1	None	0	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
		2	2	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
		3	8	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.01
		4	16	0.02	0.02	0.01	0.01	0.01	0.02	0.02	0.01
		5	32	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.01
		6	64	0.02	0.02	0.02	0.01	0.02	0.03	0.02	0.01
		7	128	0.02	0.02	0.02	0.01	0.02	0.03	0.03	0.02
		8	256	0.03	0.03	0.03	0.02	0.03	0.04	0.04	0.02
		All	506	0.05	0.04	0.04	0.03	0.04	0.06	0.05	0.03
	5	None	0	0.86	0.00	0.59	0.00	0.93	0.00	0.03	0.02
		2	2	0.86	0.00	0.59	0.00	0.93	0.00	0.03	0.02
		3	8	0.86	0.00	0.59	0.00	0.93	0.00	0.03	0.02
		4	16	0.86	0.00	0.59	0.00	0.93	0.00	0.03	0.02
		5	32	0.86	0.00	0.59	0.00	0.93	0.00	0.03	0.02
		6	64	0.87	0.00	0.59	0.00	0.93	0.00	0.03	0.02
		7	128	0.87	0.00	0.60	0.00	0.93	0.00	0.04	0.02
		8	256	0.88	0.00	0.63	0.00	0.94	0.00	0.05	0.03
		All	506	0.89	0.00	0.66	0.00	0.95	0.00	0.06	0.04

Table 13. Amino acid sites where significant ($p < 0.05$) results were obtained in the class I HLA

Position	Major codon	Group	$r_C > r_S$	$r_C < r_S$	$r_R > r_S$	$r_R < r_S$	$r_N > r_S$	$r_N < r_S$	$r_R > r_C$	$r_R < r_C$
–20	GCG	Non-ARS		✓				✓		
–10	GGG	Non-ARS				✓				✓
–9	GCC	Non-ARS		✓				✓		
–8	CTG	Non-ARS	✓				✓			
–1	GCC	Non-ARS		✓				✓		
2	TCC	Non-ARS		✓				✓		
9	TAC	Peptide-binding	✓		✓		✓			
10	ACC	Non-ARS		✓				✓		
18	GGG	Non-ARS				✓		✓		
21	CGC	Non-ARS	✓							✓
23	ATC	Non-ARS		✓				✓		
31	ACG	Non-ARS		✓				✓		
32	CAG	Non-ARS	✓							✓
43	CCG	Non-ARS		✓				✓		
44	AGG	Non-ARS				✓		✓		
45	ATG	Non-ARS			✓				✓	
47	CCG	Non-ARS		✓				✓		
50	CCG	Non-ARS		✓				✓		
58	GAG	TCR-binding				✓		✓		
62	CGG	TCR-binding			✓		✓			
63	GAG	Peptide-binding	✓		✓		✓			
66	ATC	Peptide-binding			✓		✓		✓	
67	GTG	Peptide-binding	✓				✓			✓
69	GCC	TCR-binding	✓		✓		✓			
70	AAC	Peptide-binding			✓		✓		✓	
73	ACT	Peptide-binding	✓				✓			
74	GAC	Peptide-binding			✓		✓		✓	
76	GAG	TCR-binding			✓		✓			
77	AGC	Peptide-binding	✓		✓		✓			
80	AAC	Peptide-binding	✓		✓		✓			
81	CTG	Peptide-binding	✓				✓			
82	CGC	TCR-binding			✓		✓		✓	
83	GGC	Non-ARS			✓		✓		✓	
90	GCC	Non-ARS			✓				✓	
95	CTC	Peptide-binding	✓		✓		✓			
97	AGG	Peptide-binding			✓		✓		✓	
99	TAT	Peptide-binding		✓		✓		✓		✓
103	GTG	Non-ARS	✓				✓			
107	GGG	Non-ARS	✓							✓
113	CAT	Non-ARS		✓					✓	
114	GAC	Peptide-binding			✓		✓		✓	
116	TAC	Peptide-binding	✓		✓		✓			
131	CGC	Non-ARS							✓	
135	GCG	Non-ARS		✓				✓		
138	ACC	Non-ARS		✓				✓	✓	
147	TGG	Peptide-binding								✓
151	CGT	TCR-binding	✓							✓
152	GTG	Peptide-binding			✓		✓		✓	
156	CTG	Peptide-binding	✓		✓		✓		✓	
163	ACG	TCR-binding	✓		✓		✓		✓	
166	GAG	TCR-binding	✓							✓
167	TGG	Peptide-binding								✓
171	TAC	Peptide-binding			✓				✓	
177	GAG	Non-ARS	✓							✓
178	ACG	Non-ARS			✓				✓	
182	GCG	Non-ARS		✓				✓		
188	CAT	Non-ARS		✓		✓		✓		
228	ACC	Non-ARS		✓				✓		
230	CTT	Non-ARS		✓				✓		
267	CCG	Non-ARS		✓				✓		
276	CCA	Non-ARS		✓				✓		
312	AGC	Non-ARS		✓				✓		

tion was inferred to operate on conservative, radical, and nonsynonymous substitution at 20, 23, and 24 sites, respectively, and negative selection at 21, 6, and 23 sites, respectively. Whenever positive or negative selection was inferred to operate on nonsynonymous substitution, the same selection was inferred to operate on conservative or radical substitution. In contrast, positive and negative selection were inferred to operate on conservative and radical substitution even when no selection was inferred to operate on nonsynonymous substitution.

Positive selection was inferred to operate on any of conservative, radical, and nonsynonymous substitution at 17 peptide-binding sites, 7 TCR-binding sites, 0 non-binding site, and 10 non-ARSs. The proportions of positively selected sites among the peptide-binding sites and TCR-binding sites were significantly greater than those among the non-binding sites and non-ARSs (Fisher's exact test; $p = 0.0019$ and $p = 2.03 \times 10^{-14}$ in comparisons of peptide-binding sites with non-binding sites and non-ARSs, and $p = 0.030$ and $p = 6.64 \times 10^{-6}$ in comparisons of TCR-binding sites with non-binding sites and non-ARSs). The proportions among the peptide-binding sites and TCR-binding sites and those among the non-binding sites and non-ARSs were not significantly different. Negative selection was inferred to operate on any of conservative, radical, and nonsynonymous substitution at 1 peptide-binding site, 1 TCR-binding site, 0 non-binding site, and 23 non-ARSs. The proportions of negatively selected sites were not significantly different among these groups.

Bias toward conservative substitution was inferred at 11 sites. Among these sites, positive selection operating on conservative substitution was inferred at 7 sites, whereas negative selection operating on radical substitution at 2 sites. Bias toward radical substitution was inferred at 17 sites. Among these sites, positive selection operating on radical substitution was inferred at 14 sites, whereas negative selection operating on conservative substitution at 2 sites.

The same results were obtained when termination sites were included in the computation of the numbers of sites, except that $r_R < r_C$ and $r_R > r_C$ were not observed at positions 147 and 163, respectively.

DISCUSSION

In the present study, methods for detecting natural selection operating on conservative and radical substitution and for detecting biases toward these substitutions at single amino acid sites were developed. Natural selection operating specifically on conservative and radical substitution was detected more efficiently by comparing r_S with r_C and r_S with r_R than by comparing r_S with r_N both in the computer simulation and real data analysis, although natural selection operating on both conservative and radical substitution was detected more efficiently by the latter than by the former in the computer simulation. These results were obtained because c_C and c_R are always $\leq c_N$ and s_C and s_R are always $\leq s_N$.

In the computer simulation, false-positive rates of these methods were always < 5%, unless termination sites were included in the computation of the numbers of sites and estimates of r_I/r_V were highly biased. Therefore, termination sites should be excluded from the computation of the numbers of sites. When the codons that were different from the termination codon only at one nucleotide position were used for generating the ancestral sequence, estimates of r_I/r_V were often highly biased because the transition or transversion that could produce the termination codon was not allowed to occur. For example, the second and third positions of CAT, CAC, CAA, and CAG are all non-degenerate and two-fold degenerate sites, respectively. However, a transition at the first position of CAA and CAG produces a termination codon, whereas that of CAT and CAC is a nonsynonymous substitution. Since the former substitution was not allowed to occur, r_I for CAA and CAG was suppressed compared with that for CAT and CAC. Therefore, the ratio of transitional/transversional nucleotide diversity was 0.99 and 1.01 for CAT and CAC but 0.66 and 0.67 for CAA and CAG when $r_I/r_V = 1$, and was 10.02 and 10.21 for CAT and CAC but 6.45 and 6.65 for CAA and CAG when $r_I/r_V = 10$. In real data analysis, however, heterogeneous codons are likely to be included in the ancestral sequence, so that estimates of r_I/r_V may not be highly biased. It should be noted that estimates of r_I/r_V may also be biased when positive or negative selection operated on conservative or radical substitution. To reduce this bias, only four-fold degenerate sites may be used to estimate r_I/r_V, as in the case for the analysis of class I HLA.

When incorrect phylogenetic trees were used in the analysis, reliable results were obtained when $d_T < 128$ for $b = 0.01$. These results were consistent with those observed in previous studies, where reliable results were obtained with the parsimony method when incorrect phylogenetic trees with $d_T < 25$ for $b = 0.01$, 0.02, and 0.03 were used in the analysis (Suzuki and Gojobori, 1999; Suzuki, 2004). In the analysis of class I HLA, average branch lengths of the phylogenetic trees were much smaller than 0.01. In addition, the effect of gene conversion on diversification of HLA-A, -B, and -C is known to be small (Parham et al., 1988; Gu and Nei, 1999; Nei and Rooney, 2005), suggesting that it was unlikely that $d_T > 128$ at each codon site. These observations indicated that the results obtained from the analysis of class I HLA were not affected by gene conversion to any large extent.

The proportions of positively selected sites among the peptide-binding sites and TCR-binding sites were significantly greater than those among the non-binding sites and non-ARSs, suggesting that positive selection operated on binding to peptides and TCRs. Positive selection was

inferred at positions –8, 21, 32, 45, 83, 90, 103, 107, 177, and 178 in the non-ARSs. However, all of these sites, except for position –8, were located in the α1 and α2 domains, where all of the ARSs lie. Therefore, these sites may also be involved in the binding to a peptide or TCR. Position –8 is located in the signal peptide. Interestingly, however, the motif of type I signal peptidase (Nielsen et al., 1996, 1997), by which the signal peptide of class I HLA is cleaved, is also located at positions –12 and –10, such that cleavage can occur between positions –10 and –9. If this is the case, it is possible that position –8 is also located in the α1 domain and participate in the binding to a peptide or TCR.

It has been reported that $r_N > r_S$ and $r_R > r_C$ for the entire region of the peptide-binding sites, suggesting that positive selection operated on radical substitution (Hughes et al., 1990). In the present study, the number of sites with $r_R > r_C$ was greater than that with $r_R < r_C$, and $r_R > r_C$ for the entire region of the peptide-binding sites (binomial probability; $p < 0.0000005$), which were consistent with Hughes et al. (1990). However, positive selection appeared to operate on both conservative and radical substitution, suggesting that not only amino acid charge but also other properties changed binding specificity (Wong et al., 2006). $r_R > r_C$ has been considered as an indicator of positive selection operating on radical substitution under the assumption that radical substitution is more likely to be positively selected than conservative substitution (Zhang, 2000). However, in an analysis of the envelope glycoproteins of hepatitis C virus, Hanada et al. (2006) failed to find a significant excess of radical substitutions for the entire region of positively selected sites, which were considered to be involved in binding to CTLs and antibodies, compared with the entire region of negatively selected sites. This may be because antigenic peptides are not necessarily exposed outside of a protein but may be buried inside, where radical substitutions may be constrained to maintain the three-dimensional structure (Hastings et al., 1993). Therefore, relative frequencies of sites with $r_R > r_C$ and $r_R < r_C$ may be different among antigenic and immune system proteins, and the relationship between r_C and r_R for a region of positively selected sites may not be predictable in general. In the present study, negative selection operating on conservative substitution, but not positive selection operating on radical substitution, was observed at some of the codon sites with $r_R > r_C$ in the class I HLA, suggesting that $r_R > r_C$ may not necessarily be an indicator of positive selection operating on radical substitution.

Hughes and Friedman (2005) claimed that methods for detecting positively selected amino acid sites may be unreliable, based on the observations that d_S and d_N were negatively correlated and $d_N > d_S$ can occur by chance in the comparison of 1993,217 codon sites in 4133 protein-coding genes between a pair of yeast species. However, it was suggested that the correlation was observed due to arbitrary data filtering and was of no biological significance (Yang, 2006). In addition, the parsimony and likelihood methods are intended to test whether the observation could be obtained by chance under the null hypothesis at a significance level. It should be noted that the significance level should be modified when multiple tests are conducted. However, the number of effective tests that should be corrected may be relatively small compared with the total number of sites in the sequences analyzed, because the number of nucleotide substitutions observed at a codon site is often insufficient for detecting statistically significant differences between r_S, r_C, r_R, and r_N (Fitch et al., 1997). This problem will be treated in a separate paper.

In the present study, charge was used as the property of the amino acids. However, most of the implications obtained from the computer simulation are likely to apply when other properties such as polarity, hydrophobicity, and volume are used, because they are mainly related to the problems of sample size and the features of codons.

In conclusion, methods were developed for detecting natural selection operating on conservative and radical substitution and for detecting biases toward these substitutions at single amino acid sites. Computer simulation and real data analysis indicated that these methods were reliable, unless termination sites were included in the computation of the numbers of sites, estimates of r_I/r_V were highly biased, and recombination occurred extensively. These methods may be useful for understanding the molecular mechanisms of natural selection.

The author thanks Masatoshi Nei, Robin M. Bush, Shozo Yokoyama, and two anonymous reviewers for valuable suggestions and comments. This work was supported by the National Institutes of Health grant GM020293 to Masatoshi Nei and KAKENHI 17770007 to Y.S.

REFERENCES

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., and Wiley, D. C. (1987a) Structure of the human class I histocompatibility antigen, HLA-A2. Nature **329**, 506–512.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., and Wiley, D. C. (1987b) The foreign antigen binding site and T cell recognition regions. Nature **329**, 512–518.

Dagan, T., Talmor, Y., and Graur, D. (2002) Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. Mol. Biol. Evol. **19**, 1022–1025.

Fitch, W. M. (1971) Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**, 406–416.

Fitch, W. M., Bush, R. M., Bender, C. A., and Cox, N. J. (1997) Long term trends in the evolution of H(3) HA1 human influenza type A. Proc. Natl. Acad. Sci. USA **94**, 7712–7718.

Gu, X., and Nei, M. (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. Mol. Biol. Evol. **16**, 147–156.

Hanada, K., Gojobori, T., and Li, W.-H. (2006) Radical amino acid change versus positive selection in the evolution of viral envelope proteins. Gene **385**, 83–88.

Hastings, G. Z., Francis, M. J., Rowlands, D. J., and Chain, B. M. (1993) Epitope analysis of the T cell response to a complex antigen: proliferative responses to human rhinovirus capsids. Eur. J. Immunol. **23**, 2300–2305.

Huelsenbeck, J. P., and Dyer, K. A. (2004) Bayesian estimation of positively selected sites. J. Mol. Evol. **58**, 661–672.

Hughes, A. L., and Friedman, R. (2005) Variation in the pattern of synonymous and nonsynonymous difference between two fungal genomes. Mol. Biol. Evol. **22**, 1320–1324.

Hughes, A. L., Green, J. A., Garbayo, J. M., and Roberts, R. M. (2000) Adaptive diversification within a large family of recently duplicated, placentally expressed genes. Proc. Natl. Acad. Sci. USA **97**, 3319–3323.

Hughes, A. L., and Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**, 167–170.

Hughes, A. L., Ota, T., and Nei, M. (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. **7**, 515–524.

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.

Kosakovsky Pond, S. L., and Frost, S. D. W. (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. **22**, 1208–1222.

Massingham, T., and Goldman, N. (2005) Detecting amino acid sites under positive selection and purifying selection. Genetics **169**, 1753–1762.

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

Nei, M., and Kumar, S. (2000) Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York.

Nei, M., and Rooney, A. P. (2005) Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. **39**, 121–152.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10**, 1–6.

Nielsen, H., Engelbrecht, J., von Heijne, G., and Brunak, S. (1996) Defining a similarity threshold for a functional protein sequence pattern: the signal peptide cleavage site. Proteins **24**, 165–177.

Parham, P., Lomen, C. E., Lawlor, D. A., Ways, J. P., Holmes, N., Coppin, H. L., Salter, R. D., Wan, A. M., and Ennis, P. D. (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. Proc. Natl. Acad. Sci. USA **85**, 4005–4009.

Penny, D., and Hendy, M. D. (1985) The use of tree comparison metrics. Syst. Bool. **34**, 75–82.

Robinson, D. F., and Foulds, L. R. (1981) Comparison of phylogenetic trees. Math. Biosci. **53**, 131–147.

Robinson, J., Waller, M. J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L. J., Stoehr, P., and Marsh, S. G. E. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res. **31**, 311–314.

Rzhetsky, A., and Nei, M. (1992) A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**, 945–967.

Saitou, N. (1989) A theoretical study of the underestimation of branch lengths by the maximum parsimony principle. Sys. Zool. **38**, 1–6.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**, 406–425.

Smith, N. G. C. (2003) Are radical and conservative substitution rates useful statistics in molecular evolution? J. Mol. Evol. **57**, 467–478.

Suzuki, Y. (1999) Molecular Evolution of Pathogenic Viruses. Ph.D. dissertation, Department of Genetics, School of Life Science, The Graduate University for Advanced Studies, Hayama.

Suzuki, Y. (2004) New methods for detecting positive selection at single amino acid sites. J. Mol. Evol. **59**, 11–19.

Suzuki, Y., and Gojobori, T. (1999) A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **16**, 1315–1328.

Suzuki, Y., Gojobori, T., and Nei, M. (2001) ADAPTSITE: detecting natural selection at single amino acid sites. Bioinformatics **17**, 660–661.

Suzuki, Y., and Nei, M. (2002) Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **19**, 1865–1869.

Takahashi, K., and Nei, M. (2000) Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol. Biol. Evol. **17**, 1251–1258.

Taylor, W. R. (1986) The classification of amino acid conservation. J. Theor. Biol. **119**, 205–218.

Wong, W. S. W., Sainudiin, R., and Nielsen, R. (2006) Identification of physicochemical selective pressure on protein encoding nucleotide sequences. BMC Bioinformatics **7**, 148.

Wong, W. S. W., Yang, Z., Goldman, N., and Nielsen, R. (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics **168**, 1041–1051.

Yang, Z. (2006) On the varied pattern of evolution of 2 fungal genomes: a critique of Hughes and Friedman. Mol. Biol. Evol. **23**, 2279–2282.

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

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**, 431–449.

Zhang, J. (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. **50**, 56–68.