= 130.5. We therefore have $\hat{d}_S = 0.0124 \pm 0.0176$ and $\hat{d}_N = 0.1763 \pm 0.0403$ (Table 4.3). A Z test shows that $Z = 3.7$ and $\hat{d}_N$ is significantly greater than $\hat{d}_S$ at the 0.1% level when a one-tail test is used. In contrast, the modified Nei-Gojobori method gives $S_R = 43.11$ and $N_R = 127.89$ (with $R = 0.85$), so that we have $\hat{d}_S = 0.0117 \pm 0.0166$ and $\hat{d}_N = 0.1806 \pm 0.0414$. The Z test again shows that $\hat{d}_N$ is significantly greater than $\hat{d}_S$ at the 0.1% level. Therefore, both methods show that $\hat{d}_N$ is greater than $\hat{d}_S$, and this strongly suggests that the ARS of class I MHC molecules is the target of positive Darwinian selection.

Small-Sample Tests

In the above tests, we used a large-sample test, which is not really valid because S_d was only 0.5. Let us now use Fisher's exact test. If we use the conservative Nei-Gojobori method, we obtain $S = 41$ and $N = 130$ approximately. We also assume $S_d = 1$ and $N_d = 20$ to make the test even more conservative. We then have the 2×2 contingency table given in parentheses in Table 4.1. Fisher's exact test gives a P value of 0.018. This indicates that $\hat{d}_N$ is significantly greater than $\hat{d}_S$. If we use the modified Nei-Gojobori method, we obtain $S_R = 44$ and $N_R = 127$ (with $R = 0.85$), and Fisher's test gives a P value of 0.012. Therefore, the P values for the small-sample test are higher than those for the large-sample test.

Tests of $\bar{d}_N - \bar{d}_S$ or $\bar{p}_N - \bar{p}_S$

Since the power of detecting positive selection is low in this case because of the small number of codons involved, let us consider the averages of $\hat{d}_N$ and $\hat{d}_S$. In the present case, there are three sequences, so $\hat{d}_N$ and $\hat{d}_S$ can be computed for three pairs of alleles. We can then obtain the averages ($\bar{d}_N$ and $\bar{d}_S$) of these values. If we use the Nei-Gojobori method, they become $\bar{d}_N = (0.1763 + 0.1822 + 0.1479)/3 = 0.1688$ and $\bar{d}_S = (0.0124 + 0.0000 + 0.0124)/3 = 0.0083$. Therefore, the difference $\hat{D} = \bar{d}_N - \bar{d}_S$ is 0.1605. If we use the bootstrap method, the standard error of $\hat{D} = \bar{d}_N - \bar{d}_S$ becomes 0.0322. (This was computed by a program in MEGA2 using 1000 bootstrap replications.) Therefore, we have $Z = 4.98$. If we use Nei and Jin's method, we obtain $Z = 4.80$, which is again highly significant. These results reinforce the conclusion reached by comparison of two sequences.

4.2. Methods Based on Kimura's 2-Parameter Model

Li-Wu-Luo Method

Li et al. (1985) developed another method, based on Kimura's 2-parameter model. They first noted that when the degeneracy of the genetic code is considered, the nucleotide sites of codons can be classified into 4-fold degenerate, 2-fold degenerate, and 0-fold degenerate (nondegenerate) sites with a few exceptions (e.g., isoleucine codons). A site is called 4-

fold degenerate if all possible changes at the site are synonymous, 2-fold degenerate if one of the three possible changes is synonymous, and 0-fold degenerate if all changes are nonsynonymous or nonsense mutations. For example, the third nucleotide positions of the valine codons are 4-fold degenerate sites, and the second positions of all codons are 0-fold degenerate sites. The third positions of the three isoleucine codons are actually 3-fold degenerate sites, but they are regarded as 2-fold degenerate sites to simplify the computation.

Using the above rule, we can compute the numbers of three types of sites for each of the two sequences and denote by L_0, L_2, and L_4 the average numbers of 0-fold, 2-fold, and 4-fold degenerate sites for the two sequences compared, respectively. We then compare the two sequences, codon by codon, and classify each nucleotide difference as either a transition or a transversion. We denote by P_i and Q_i the proportions of transitional and transversional nucleotide differences at the i-th class of nucleotide sites ($i = 0$, 2, or 4). (Actually, they considered all possible evolutionary pathways between each pair of codons as in the case of the Nei-Gojobori method and computed P_i and Q_i taking into account the likelihood of occurrence of each amino acid substitution. See Li et al. [1985] for the detail.) We can then estimate the numbers of transitional (A_i) and transversional (B_i) substitutions per site for each of the three classes of nucleotide sites. That is,

$$A_i = \frac{1}{2}\ln(a_i) - \frac{1}{4}\ln(b_i) \tag{4.9a}$$

$$B_i = \frac{1}{2}\ln(b_i) \tag{4.9b}$$

where $a_i = 1/(1 - 2P_i - Q_i)$ and $b_i = 1/(1 - 2Q_i)$.

We note that all substitutions at 4-fold sites are synonymous and all substitutions at 0-fold sites are nonsynonymous. At 2-fold sites, transitional changes (A_2) are mostly synonymous, whereas transversional changes are mostly nonsynonymous. Assuming that nucleotide substitution occurs with equal frequency among the four nucleotides A, T, C, and G, Li et al. (1985) suggested that one third of 2-fold degenerate sites are potentially synonymous sites and two thirds are potentially nonsynonymous sites. With this assumption, they proposed that d_S and d_N be estimated by the following formulas.

$$\hat{d}_S = \frac{3[L_2A_2 + L_4(A_4 + B_4)]}{L_2 + 3L_4} \tag{4.10a}$$

$$\hat{d}_N = \frac{3[L_0(A_0 + B_0) + L_2B_2]}{3L_0 + 2L_2} \tag{4.10b}$$

These formulas depend on a number of assumptions, which are not always satisfied with actual data. First, the type of a given nucleotide site in one sequence may not be the same as that of the homologous site in the other sequence. For example, the type of a given position in one se-