```
                                    α1
A*2301 GGC TCC CAC TCC ATG AGG TAT TTC TCC ACA TCC GTG TCC CGG CCC GGC CGC GGG GAG CCC  20
A*2501 ... ... ... ... ... ... ... ... .A. ..C ... ... ... ... ... ... ... ... ... ...
A*3301 ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ... ... ... ...

A*2301 CGC TTC ATC GCC GTG GGC TAC GTG GAC GAC ACG CAG TTC GTG CGG TTC GAC AGC GAC GCC  40
A*2501 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
A*3301 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

A*2301 GCG AGC CAG AGG ATG GAG CCG CGG GCG CCG TGG ATA GAG CAG GAG GGG CCG GAG TAT TGG  60
A*2501 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
A*3301 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

A*2301 GAC GAG GAG ACA GGG AAA GTG AAG GCC CAC TCA CAG ACT GAC CGA GAG AAC CTG CGG ATC  80
A*2501 ... CG. A.C ... C.. ..T ... ... ... ... ... ... ... ... ... .G. ... ... ...
A*3301 ... CG. A.C ... C.. ..T ... ... ... ... ... ... .T. ... ... .T. G.. ... G.. .C.

                                                α2
A*2301 GCG CTC CGC TAC TAC AAC CAG AGC GAG GCC|GGT TCT CAC ACC CTC CAG ATG ATG TTT GGC 100
A*2501 ... ... ... ... ... ... ... ... ... .A.|... ... ... ... ... A.. ... .G. ... .A. ...
A*3301 CT. .G. G.. ... ... ... ... ... ... ...|... ... ... ... ... A.. ... ... ... .A. ...

A*2301 TGC GAC GTG GGG TCG GAC GGG CGC TTC CTC CGC GGG TAC CAC CAG TAC GCC TAC GAC GGC 120
A*2501 ... ... ... ... C.. ... ... ... ... ... ... ... ... ... ..G ... G.. ..T ... ...
A*3301 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... G.. ... ... ...

A*2301 AAG GAT TAC ATC GCC CTG AAA GAG GAC CTG CGC TCT TGG ACC GCG GCG GAC ATG GCG GCT 140
A*2501 ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ...
A*3301 ... ... ... ... ... ... ... T.. ..C ... ... ... ... ... ... ... ... ... ... ...

A*2301 CAG ATC ACC CAG CGC AAG TGG GAG GCG GCC CGT GTG GCG GAG CAG TTG AGA GCC TAC CTG 160
A*2501 ... ... ... ... ... ... ... ... A.. ... .A. .A. ... ... ... .G. ... ... ... ...
A*3301 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

A*2301 GAG GGC ACG TGC GTG GAC GGG CTC CGC AGA TAC CTG GAG AAC GGG AAG GAG ACG CTG CAG 180
A*2501 ... ... CG. ... ... ..G T.. ... ... ... ... ... ... ... ... ... ... ... ... ...
A*3301 ... ... ... ... ... ..G T.. ... ... ... C.. ... ... ... ... ... ... ... ... ...

                    α3
A*2301 CGC ACG|GAC CCC CCC AAG ACA CAT ATG ACC CAC CAC CCC ATC TCT GAC CAT GAG GCC ACT 200
A*2501 ... ...|... G.. ... ... ..G ... ... ..T ... ... G.T G.. ... ... ... ... ..C
A*3301 ... ...|... ... ... .G. ..G ... ... ..T ... ... G.T G.. ... ... ... ... ..C

A*2301 CTG AGA TGC TGG GCC CTG GGC TTC TAC CCT GCG GAG ATC ACA CTG ACC TGG CAG CGG GAT 220
A*2501 ... ..G ... ... ... ... A.. ... ... ... ... ... ... ... ... ... ... ... ... ...
A*3301 ... ..G ... ... ... ... A.. ... ... ... ... ... ... ... ... ... ... ... ... ...

A*2301 GGG GAG GAC CAG ACC CAG GAC ACG GAG CTT GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC 240
A*2501 ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ..G ...
A*3301 ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ...

A*2301 TTC CAG AAG TGG GCA GCT GTG GTG GTA CCT TCT GGA GAG GAG CAG AGA TAC ACC TGC CAT 260
A*2501 ... ... ... ... ... ..G T.. ... ... ..G ... ... ... C.. ... ... ... ... ... ...
A*3301 ... ... ... ... ... ..G T.. ... ... ..G ... ... ... C.. ... ... ... ... ... ...

A*2301 GTG CAG CAT GAG GGT CTG CCC AAG CCC CTC ACC CTG AGA TGG 274
A*2501 ... ... ... ... ... ... ... ... ... ... ... ... ... ...
A*3301 ... ... ... ... ... ..C ... ... ... ... ... ... ... ...
```

FIGURE 4.1. Nucleotide sequences of three human class I *HLA-A* alleles for the three extracellular domains α_1, α_2, and α_3. A dot (.) shows identity with the first sequence. Exons boundaries are marked with vertical lines. The nucleotides at the antigen recognition site (ARS) are in boldface.

values, we obtain $\hat{d}_S = 0.0566 \pm 0.0169$ and $\hat{d}_N = 0.0499 \pm 0.0093$. Therefore, $\hat{d}_S$ has decreased and $\hat{d}_N$ has increased slightly.

Adaptive Evolution

X-ray diffraction studies have shown that class I MHC molecules form a groove in which a foreign peptide is bound (Bjorkman et al. 1987a,

Table 4.2 Codons that are different between the HLA *A*2301* and *A*2501* alleles.

Codon		s_d	n_d	$s_d + n_d$	Codon		s_d	n_d	$s_d + n_d$
*9	TCC-TAC		1	1	*156	TTG-TGG		1	1
10	ACA-ACC	1		1	*163	ACG-CGG	0.5	1.5	2
*62	GAG-CGG		2	2	*166	GAC-GAG		1	1
*63	GAG-AAC		2	2	*167	GGG-TGG		1	1
*65	GGG-CGG		1	1	184	CCC-GCC		1	1
*66	AAA-AAT		1	1	187	ACA-ACG	1		1
*77	AAC-AGC		1	1	190	ACC-ACT	1		1
90	GCC-GAC		1	1	193	CCC-GCT	1	1	2
*95	CTC-ATC		1	1	194	ATC-GTC		1	1
*97	ATG-AGG		1	1	200	ACT-ACC	1		1
*99	TTT-TAT		1	1	202	AGA-AGG	1		1
105	TCG-CCG		1	1	207	GGC-AGC		1	1
*114	CAC-CAG		1	1	230	CTT-CTC	1		1
*116	TAC-GAC		1	1	239	GGA-GGG	1		1
117	GCC-GCT	1		1	245	GCA-GCG	1		1
127	AAA-AAC		1	1	246	GCT-TCT		1	1
*149	GCG-ACG		1	1	249	GTA-GTG	1		1
*151	CGT-CAT		1	1	253	GAG-CAG		1	1
*152	GTG-GAG		1	1		Total	11.5	29.5	41

Note: Antigen recognition sites are indicated with an asterisk.

1987b). This groove is called the antigen recognition site (ARS) and consists of 57 amino acid sites (boldfaced letters in Figure 4.1). If we apply the Nei-Gojobori method to the 57 amino acid sites of the ARS for alleles *A*2301* and *A*2501*, we obtain $S_d = 0.5$, $N_d = 20.5$, $S = 40.5$, and N

Table 4.3 Numbers of synonymous ($\hat{d}_S$) and nonsynonymous ($\hat{d}_N$) substitutions between the HLA *A*2301* and *A*2501* alleles for the extracellular region and the antigen recognition sites (ARS).

	Extracellular Region ($C = 274$)		ARS ($C = 57$)	
Method	$\hat{d}_S$	$\hat{d}_N$	$\hat{d}_S$	$\hat{d}_N$
$R = 0.5$				
NG[a]	6.08 ± 1.81	4.87 ± 0.91	1.24 ± 1.76	17.63 ± 4.03
LWL[b]	6.52 ± 2.02	4.82 ± 0.89	0.03 ± 1.87	17.25 ± 3.99
$R = 0.85$[c]				
Modified-NG	5.66 ± 1.69	4.99 ± 0.93	1.17 ± 1.66	18.03 ± 4.13
PBL[d]	4.59 ± 1.46	4.80 ± 0.90	0.02 ± 1.14	17.04 ± 3.96
Kumar	4.55 ± 1.46	4.74 ± 0.91	0.36 ± 1.96	16.79 ± 4.03
Ina II	4.87 ± 1.47	5.31 ± 0.99	1.50 ± 2.13	16.67 ± 3.81
GY[e]	12.17	4.25	0.02	16.98

Note: $\hat{d}_S$ and $\hat{d}_N$ are multiplied by 100.

[a]NG: Nei-Gojobori.
[b]LWL: Li-Wu-Luo.
[c]$R = 0.85$ was used only for the Modified-NG method. In the other methods, R was computed automatically.
[d]PBL: Pamilo-Bianchi-Li.
[e]GY: Goldman-Yang.