Genes and Immunity (2005) 6, 398–406

www.nature.com/gene

FULL PAPER

Evidence for natural selection in the HAVCR1 gene: high degree of amino-acid variability in the mucin domain of human HAVCR1 protein

T Nakajima[1,2], S Wooding[3], Y Satta[4], N Jinnai[1], S Goto[1], I Hayasaka[5], N Saitou[6], J Guan-jun[7], K Tokunaga[8], LB Jorde[3], M Emi[2] and I Inoue[1]

[1]*Division of Genetic Diagnosis, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan;* [2]*Department of Molecular Biology, Institute of Gerontology, Nippon Medical School, Kawasaki, Japan;* [3]*Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT, USA;* [4]*Department of Biosystems Science, Graduate University for Advanced Studies, Hayama, Japan;* [5]*Kumamoto Primates Park, Sanwa Kagaku Kenkyusho Co. Ltd, Kumamoto, Japan;* [6]*Division of Population Genetics, National Institute of Genetics, Mishima, Japan;* [7]*Red Cross Blood Center of Harbin, Harbin, China;* [8]*Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*

The family of genes encoding T-cell immunoglobulin and mucin-domain containing proteins (Tim), which are cell-surface molecules expressed in CD4$^+$ T helper cells, has important roles in the immune system. Here, we report three unusual patterns of genetic variation in the human hepatitis A virus cellular receptor 1 gene (HAVCR1) that are similar to patterns observed in major histocompatibility complex loci. First, levels of polymorphism in exon 4 of HAVCR1 were exceptionally high in humans (nucleotide diversity (π) = 45.45 $\times$ 10^{-4}). Second, nonsynonymous substitutions and insertion/deletion variants were more frequent than synonymous substitutions in that exon (10 out of 12 variants). The rate of the mean number of nucleotide substitutions at nonsynonymous sites to synonymous sites at HAVCR1-exon 4 is >1 ($P_A/P_S = 1.92$ and $\pi_A/\pi_S = 2.23$). Third, levels of divergence among human, chimp, and gorilla sequences were unusually high in HAVCR1-exon 4 sequences. These features suggest that patterns of variation in HAVCR1 have been shaped by both positive and balancing natural selection in the course of primate evolution. Evidence that the effects of natural selection are largely restricted to the mucin domain of HAVCR1 suggests that this region may be of particular evolutionary and epidemiological interest.
Genes and Immunity (2005) **6,** 398–406. doi:10.1038/sj.gene.6364215; published online 12 May 2005

Keywords: *HAVCR1; Tim gene family; natural selection*

Introduction

The TIM gene family, *HAVCR1* (human hepatitis A virus cellular receptor 1), *HAVCR2*, and *TIMD4*, encodes a small group of type I transmembrane glycoproteins composed of an immunoglobulin-like domain positioned atop a mucin-like domain. This structure is similar to that of several known adhesion proteins, and points to a possible role of TIM proteins as adhesion molecules.[1,2] The expression of these genes in T helper (Th) cells suggests that these proteins might have particularly important roles in regulating the immune response. For example, *HAVCR2*, which is highly expressed by Th1 cells, appears to be involved in the differentiation of CD4$^+$ Th cells and the activation of macrophages.[3] The HAVCR2 pathway also provides an important mechanism for downregulating Th1-dependent immune responses and facilitates the development of immunological tolerance.[4,5]

HAVCR1, the first TIM family member to be identified, was initially recognized as a receptor for the hepatitis A virus (HAV)[6] and was later implicated in susceptibility to allergic asthma in mice[7] and humans.[8] This evidence, along with the participation of other TIM genes in the human immune system, suggests that functional variation in *HAVCR1* could be a particularly important determinant of the immune response. Yet, little is known about patterns of functional variation in this gene.

Information about the presence of functional variation in genes can be obtained in a variety of ways. In humans, linkage and association studies are often used to identify genotype–phenotype correlations; in mice, targeted gene knockouts are often useful. While these approaches have been highly successful, they can be prohibitively expensive. A major goal of this study was to generate new hypotheses about the presence of functional polymorphisms by testing for signatures of natural selection in the *HAVCR1* gene. As natural selection acts on variable phenotypes, the effects of natural selection will be

Correspondence: Dr T Nakajima, Department of Molecular Biology, Institute of Gerontology, Nippon Medical School, 1-396 Kosugi-cho, Nakahara-ku, Kawasaki 211-8533, Japan. E-mail: tnakajim@nms.ac.jp
Received 18 January 2005; revised 15 March 2005; accepted 22 March 2005; published online 12 May 2005

strongest on those variants that confer such phenotypes. Thus, evidence of natural selection in a genomic region suggests that functionally important variants are present in that region. By examining patterns of nucleotide diversity both within humans and between humans, chimpanzees, and gorillas, we hoped to identify those sites most likely to confer phenotypic variation. We found that signatures of balancing natural selection are present in *HAVCR1*; however, these signatures are localized to a region encoding an extracellular mucin domain that is likely involved in the recognition of molecules outside the cell. We hypothesize that these variants confer phenotypic variation and are thus good candidates for use in future linkage and association analyses.

Results

Expression of Tim-family genes in human CD4$^+$ Th cells

CD4$^+$ Th cells were isolated from human peripheral blood using monoclonal antibodies to CD4 coupled to magnetic beads (MACS, Miltenyi Biotechnology, Auburn, CA, USA), and stimulated *in vitro* under Th1- or Th2-differentiating conditions by plate-bound anti-CD3ε along with IL-12 + anti-IL-4 or by anti-CD28 with IL-4 + anti-IFN-γ + anti-IL-12.[9] As shown in Figure 1, human *HAVCR1, HAVCR2,* and *TIMD4* were highly expressed in Th1-differentiating conditions 5 days after stimulation. *HAVCR1* in the mouse is specifically expressed by differentiating and differentiated murine Th2 cells,[7] and the expression of Tim-family genes in human CD4$^+$ Th cells implies that Tim proteins are likely to participate in immune regulation in the human.

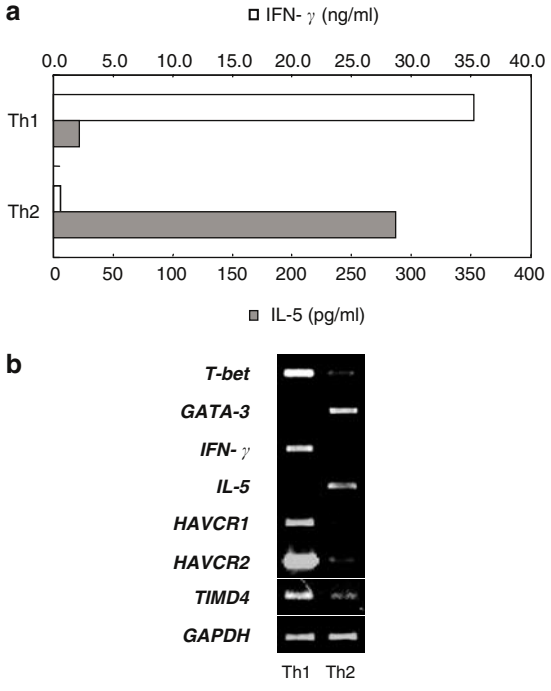

Figure 1 Expression of *HAVCR1* in Th1-differentiated human CD4$^+$ Th cells. (**a**) Production of IFN-γ and IL-5 in CD4$^+$ Th cells stimulated under Th1- or Th2-differentiating conditions *in vitro*. (**b**) Expression of various human genes, including *HAVCR1,* under Th1-differentiating conditions after 5 days of stimulation.

Nucleotide polymorphisms in human HAVCR1, HAVCR2, and TIMD4 genes

Patterns of DNA sequence variation were evaluated in all exons of *HAVCR1*, *HAVCR2*, and *TIMD4* in 281 human samples (94 African-American, 92 Caucasian, and 95 Japanese). A total of 28 sequence variants were identified: 17 in *HAVCR1*, three in *HAVCR2*, and eight in *TIMD4* (Figure 2 and Table 1). Many of the nucleotide variants detected in coding regions were observed in exon 4 of *HAVCR1* (12 sequence variations), a region encoding the N-terminal half of the mucin domain. Six of the 12 variants observed in exon 4 of *HAVCR1* were nonsynonymous nucleotide substitutions. Four were in-frame insertion/deletion polymorphisms (one 18 bp/6aa deletion and three 3 bp/1aa deletions). The abundance of nonsynonymous nucleotide substitutions and insertion/deletion polymorphisms within one exon suggested an evolutionary history of natural selection, as in the case of antigen recognition sites in major histocompatibility complex (MHC)-HLA genes.[10]

To determine whether natural selection has occurred at *HAVCR1*-exon 4, we determined the complete sequences of all *HAVCR1* exons and ~1.5 kb of sequence in its promoter region among 624 individuals from 10 different populations (73 native African, 94 African-American, 49 Palestinian, 46 Druze, 46 Ashkenazi, 92 Caucasian, 51 Indian, 95 Japanese, 23 Korean, and 55 Northern Chinese (Heilongjian)). TA cloning was performed using TOPO™ TA cloning kits (Invitrogen, Carlsbad, CA, USA) to determine the gametic phase for individuals who were heterozygous in *HAVCR1*-exon 4. The level of nucleotide diversity (π) for this locus (45.45×10^{-4} in all human samples) was unusually high, about three times greater than the value for noncoding promoter regions (15.41×10^{-4} in all human samples; Table 2).

The patterns of nucleotide substitutions at nonsynonymous sites and/or synonymous sites in *HAVCR1*-exon 4 were evaluated. The mean number of nucleotide substitutions at nonsynonymous sites (P_A) and synonymous sites (P_S) were evaluated based on 11 *HAVCR1*-exon 4 sequences identified in human population samples. The value of P_A at *HAVCR1*-exon 4 (0.0085) is much higher than the value for other coding regions of HAVCR1 (0.0011) and the ratio of P_A to P_S at *HAVCR1*-exon 4 is >1 ($P_A/P_S = 1.92$).

A similar pattern was found in the mean number of nucleotide substitutions at nonsynonymous sites (π_A) and the mean number of nucleotide substitutions at synonymous sites (π_S) in human population samples.[11,12] As shown in Table 2, π_A (54.09×10^{-4} in all human samples) is much higher than π_S (24.31×10^{-4} in all human samples) and the ratio of π_A to π_S at *HAVCR1*-exon 4 is >1 ($\pi_A/\pi_S = 2.23$). This π_A/π_S value is much higher than the value based on 13 582 bp of coding sequences from 106 genes (2.23 *vs* 0.64), described by Cargill *et al*.[13]

Haplotype network of HAVCR1-exon 4 in human populations

I all, 11 sequences were identified at *HAVCR1*-exon 4 in population samples (Figure 3 and Table 3). African and African-American samples showed more diverse allelic

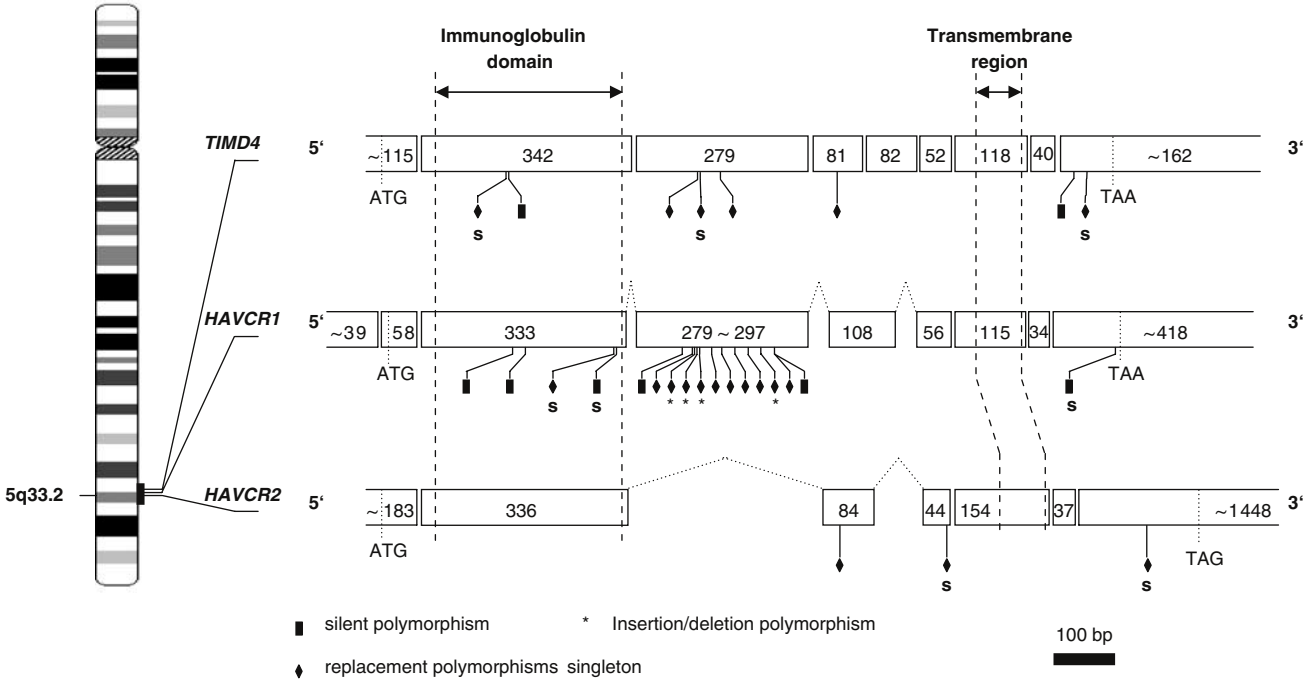

Figure 2 Sequence variations observed in coding regions of *HAVCR1*, *HAVCR2*, and *TIMD4*.

Table 1 Sequence variations in *HAVCR1*, *HAVCR2*, and *TIMD4*

Gene symbol	*Location*	*Nucleotide variation*	*Amino-acid variation*	*Comment*	*Frequency*		
					African-American	*Caucasian*	*Japanese*
HAVCR1	Exon 3	A/C	Ser 44 Ser		0.165	0.156	0.016
HAVCR1	Exon 3	C/T	Gly 50 Gly		0.074	0.000	0.000
HAVCR1	Exon 3	G/C	Asp 99 His	Singleton	0.000	0.005	0.000
HAVCR1	Exon 3	C/T	Gly 101 Gly	Singleton	0.000	0.005	0.000
HAVCR1	Exon 4	G/A	Thr 152 Thr		0.023	0.000	0.000
HAVCR1	Exon 4	T/C	Thr 158 Met		0.256	0.360	0.142
HAVCR1	Exon 4	3bp del	Thr 160 del		0.097	0.156	0.016
HAVCR1	Exon 4	3bp del	Thr 161 del		0.074	0.000	0.000
HAVCR1	Exon 4	18bp del	6aa 161-166 del	Ref[8]	0.142	0.360	0.142
HAVCR1	Exon 4	C/T	Thr 175 Met		0.068	0.000	0.000
HAVCR1	Exon 4	C/T	Pro 180 Leu		0.182	0.167	0.016
HAVCR1	Exon 4	C/T	Thr 187 Met		0.011	0.000	0.000
HAVCR1	Exon 4	C/T	Thr 195 Met		0.034	0.000	0.000
HAVCR1	Exon 4	3bp del	Thr 201 del		0.392	0.161	0.126
HAVCR1	Exon 4	A/G	Thr 208 Ala		0.170	0.371	0.079
HAVCR1	Exon 4	G/T	Thr 208 Thr		0.170	0.156	0.016
HAVCR1	Exon 9	C/T	Leu 361 Phe	Singleton	0.000	0.005	0.000
HAVCR2	Exon 3	T/G	Leu 140 Arg	rs1036199[a]	0.156	0.156	0.016
HAVCR2	Exon 4	T/C	Ile 171 Thr	Singleton	0.016	0.000	0.000
HAVCR2	Exon 7	C/T	Pro 277 Leu	Singleton	0.016	0.000	0.000
TIMD4	Exon 2	G/A	Gly 63 Ser	Singleton	0.000	0.000	0.016
TIMD4	Exon 2	G/A	Lys 65 Lys		0.016	0.016	0.031
TIMD4	Exon 3	C/A	Thr 170 Lys		0.047	0.000	0.000
TIMD4	Exon 3	G/A	Val 172 Met	Singleton	0.016	0.000	0.000
TIMD4	Exon 3	C/T	Thr 179 Ile		0.141	0.000	0.000
TIMD4	Exon 4	C/T	Ala 240 Val	rs6873053[a]	0.000	0.047	0.000
TIMD4	Exon 9	A/G	Lys 359 Lys		0.078	0.000	0.000
TIMD4	Exon 9	A/G	Val 365 Met	rs7731575[a], singleton	0.016	0.000	0.000

Sequence variations in HAVCR2 and TIMD4 were identified in 32 African-American, 32 Caucasian, and 32 Japanese.
[a]Number from dbSNP database of NCBI (http://www.ncbi.nlm.nih.gov/SNP/).

Table 2 Nucleotide diversities ($\pi \times 10^{-4}$) in *HAVCR1* among 10 populations

	Exon 4				*Other coding regions*				*Promoter*
	π_{total}	π_S	π_A	π_A/π_S	π_{total}	π_S	π_A	π_A/π_S	π
Native African	52.00	35.46	58.80	1.66	4.47	18.11	0	—	12.47
African-American	50.76	40.88	54.82	1.34	5.20	21.05	0	—	15.11
Druze	55.87	36.70	63.74	1.74	4.89	17.64	0.72	0.04	13.37
Palestinian	42.99	21.09	51.99	2.47	2.19	8.88	0	—	11.76
Ashkenazi	52.40	29.86	61.66	2.06	3.11	12.58	0	—	16.22
Caucasian	51.41	32.17	59.32	1.84	3.89	14.66	0.36	0.02	16.74
Indian	31.73	3.12	43.54	14.45	0.31	1.27	0	—	14.72
Japanese	15.95	4.41	20.69	4.69	0.46	1.86	0	—	12.85
Korean	29.46	10.24	37.36	11.71	0.55	2.22	0	—	—
North Chinese	19.17	2.23	26.12	3.65	0.23	0.94	0	—	—
All	45.45	24.31	54.09	2.23	5.12	20.40	0.12	0.006	15.41[a]

[a]Nucleotide diversity of *HAVCR1* promoter region for all human populations is based on the data from eight human population samples.

frequencies and the highest level of heterozygosity, 0.823 and 0.815, respectively. In East Asian samples D3-A was the most frequent haplotype, and these populations showed lower heterozygosity than any other group (Figure 3 and Table 3). Ewens–Watterson neutrality tests[14] did not deviate significantly from expectations under neutrality, except in native African, African–American, Ashkenazi populations, and in all combined human samples (Table 3).

A haplotype network for 11 *HAVCR1*-exon 4 haplotypes shows that three main clusters of high-frequency haplotypes are separated by relatively long branches in human populations (Figure 3).

Nucleotide diversity among human, chimpanzee, and gorilla HAVCR1-exon 4 sequences

Comparisons of *HAVCR1*-exon 4 in human, chimpanzee, and gorilla sequences revealed considerable nucleotide divergence among the three primate species (Figure 4). Figure 5 illustrates neighbor-joining phylogenic trees derived from 1505-bp promoter sequences and 700-bp sequences around exon 4 of human, chimpanzee, and gorilla *HAVCR1* genes. These trees showed that levels of divergence among exon-4 sequences are large in comparison with those in promoter sequences. Further, some human sequences were more similar to chimpanzee sequences than they were to other human sequences. This pattern implies that chimpanzee and human sequences diverged before the human–chimpanzee split more than 5 000 000 years ago.[15] Levels of divergence between gorilla and human were also relatively large. For example, gorilla and human sequences with haplotype D3-A differed at 6.3% of synonymous sites in exon 4 (five nucleotide differences in 79.33 synonymous sites).

It is notable that high nucleotide diversity among primate sequences was also observed at the boundary of exon 4 and intron 4 (Figure 6). Comparisons of *HAVCR1*-intron 4 sequences from human and gorilla showed that the high degree of divergence observed in exon 4 was also present in a 100-bp segment of intron 4 adjoining exon 4. We hypothesize that this pattern of variation is due to hitchhiking effects.

Discussion

Two patterns of variation in HAVCR1-exon 4 suggest that evolutionary processes in this region have not been neutral, but have been driven by natural selection.

First, under neutral conditions, genetic drift is expected to result in the coalescence of gene genealogies over a relatively short time period.[16] The high levels of divergence within humans, including the divergence of some exon 4 alleles prior to human–chimpazee speciation, are most consistent with the hypothesis that balancing natural selection has acted to maintain different *HAVCR1*-exon 4 alleles for an extended period of time.

Under balancing selection, alleles are maintained at intermediate frequencies, often as a result of rare-allele advantage. Two major models of balancing selection, overdominant and frequency-dependent selection, are suggested. Frequency-dependent selection will result in a constant turnover of alleles in a population. While this model is expected to generate a higher number of substitutions per synonymous site than per nonsynonymous site, it can explain neither the high degree of polymorphism seen here nor the long-term persistence of advantageous alleles.[17] However, the overdominant selection hypothesis can adequately explain the large extent of polymorphism we have seen at *HAVCR1*-exon 4 within human populations. Overdominant selection, where the fitness of alleles in the heterozygous state is greater than those that are homozygous, is the model used to explain the large degree of polymorphism at the MHC loci.[10] Some evidence now exists to support an advantage for MHC heterozygosity, which has been associated with protection against several viral diseases including delayed progression of HIV infection[18] and resistance to hepatitis B virus.[19]

The high levels of amino-acid substitution in *HAVCR1*-exon 4 suggest that positive natural selection has been active along with balancing natural selection in this region. Both P_A/P_S and π_A/π_S in this region are >1. The π_A/π_S ratio is much higher than a published value based on 13 582 bp of coding sequences from 106 genes (2.23 *vs* 0.64).[13] These values are difficult to evaluate statistically

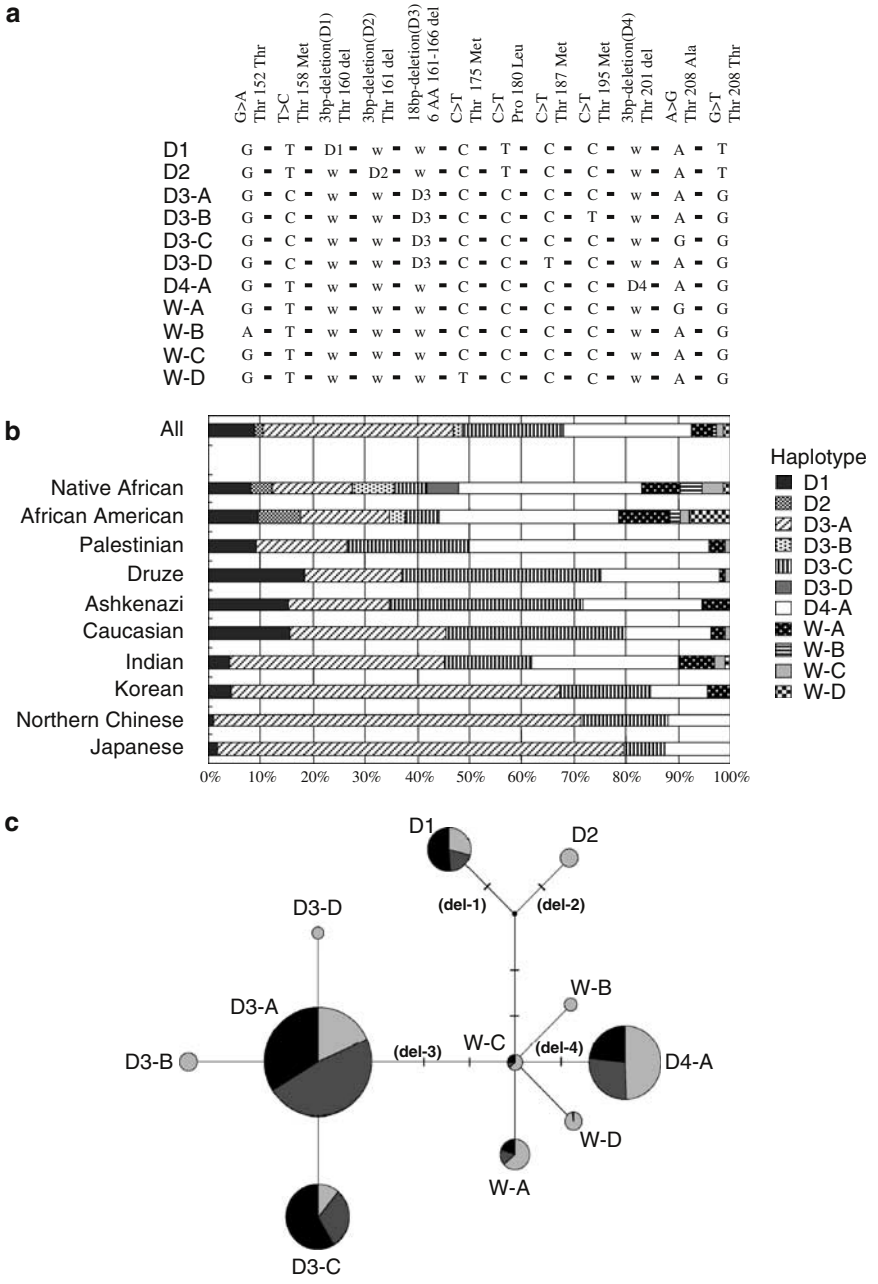

Figure 3 (**a**) Haplotypes for human *HAVCR1*-exon 4. (**b**) Haplotype frequencies of *HAVCR1*-exon 4 in 10 human populations. (**c**) Minimum spanning network relating haplotypes for human *HAVCR1*-exon 4. Each circle in the network represents a haplotype, and the area of each circle indicates its frequency relative to the other haplotypes. Each line connecting haplotypes represents a single nucleotide substitution. Connections with more than one nucleotide substitution are indicated with slashes, each of which represents one substitution. Deletions are indicated with slashes and labeled in parentheses. Within each circle, the relative frequency of the haplotype in each continental population is indicated by shading: light = Africa; medium = Asia; and dark = Europe.

because the effects of natural selection on the human genome as a whole, and thus the underlying distributions of the ratios, are not known. Nonetheless, our results suggest that the rate of amino-acid change in the mucin domain encoded by exon 4 is elevated substantially over that in most other genes. Neither a high mutation rate in *HAVCR1*-exon 4 nor gene conversion can explain this high rate of nonsynonymous substitution. The simplest explanation is that positive natural selection has actively preserved amino-acid changes.

The effects of positive natural selection are also suggested by the relatively high nucleotide divergence among primate species, shown in the phylogenic trees for *HAVCR1*-exon 4 sequences in Figure 5. These trees show that *HAVCR1*-exon 4 sequences are highly divergent relative to alleles in the *HAVCR1* promoter region. One explanation for this finding is that natural selection has constrained the promoter region, preventing it from diverging. However, the high levels of amino-acid substitution in exon 4 suggest

Table 3 *HAVCR1* haplotype frequencies and Ewens–Watterson neutrality test in 10 human populations

Haplotype	*Native African (n = 146)*	*African-American (n = 188)*	*Palestinian (n = 98)*	*Druze (n = 92)*	*Ashkenazi (n = 92)*	*Caucasian (n = 184)*	*Indian (n = 102)*	*Japanese (n = 190)*	*Korean (n = 46)*	*Northern Chinese (n = 110)*	*All human samples (n = 1248)*
D1	0.082	0.096	0.092	0.185	0.152	0.158	0.039	0.016	0.044	0.009	0.087
D2	0.041	0.080	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017
D3-A	0.151	0.170	0.173	0.185	0.196	0.299	0.412	0.779	0.630	0.700	0.366
D3-B	0.082	0.032	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014
D3-C	0.062	0.064	0.235	0.380	0.370	0.342	0.167	0.079	0.174	0.164	0.189
D3-D	0.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
D4	0.349	0.346	0.459	0.228	0.228	0.163	0.284	0.126	0.109	0.118	0.244
W-A	0.075	0.096	0.031	0.011	0.054	0.027	0.069	0.000	0.044	0.009	0.042
W-B	0.041	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008
W-C	0.041	0.016	0.010	0.011	0.000	0.011	0.020	0.000	0.000	0.000	0.012
W-D	0.014	0.080	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.014
Heterozygosity	0.823	0.815	0.695	0.735	0.747	0.741	0.715	0.371	0.557	0.469	0.761
Ewens–Watterson neutrality test (Slatkin's exact *P*-value)	0.005*	0.007*	0.246	0.301	0.014*	0.086	0.310	0.334	0.372	0.708	0.031*

*Significantly different at the 5% level.

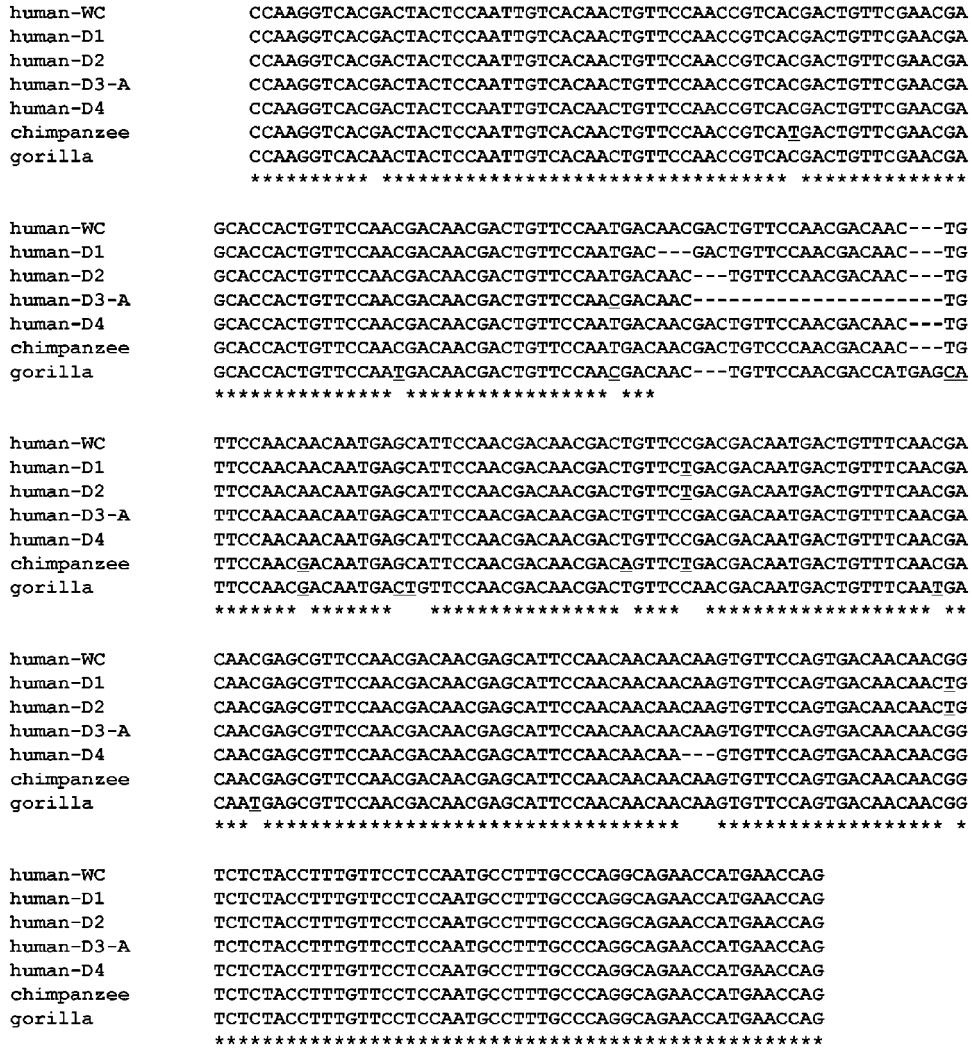

Figure 4 Alignment of *HAVCR1*-exon 4 sequences from human, chimpanzee, and gorilla. Asterisks indicate identical nucleotides among the three primate species.

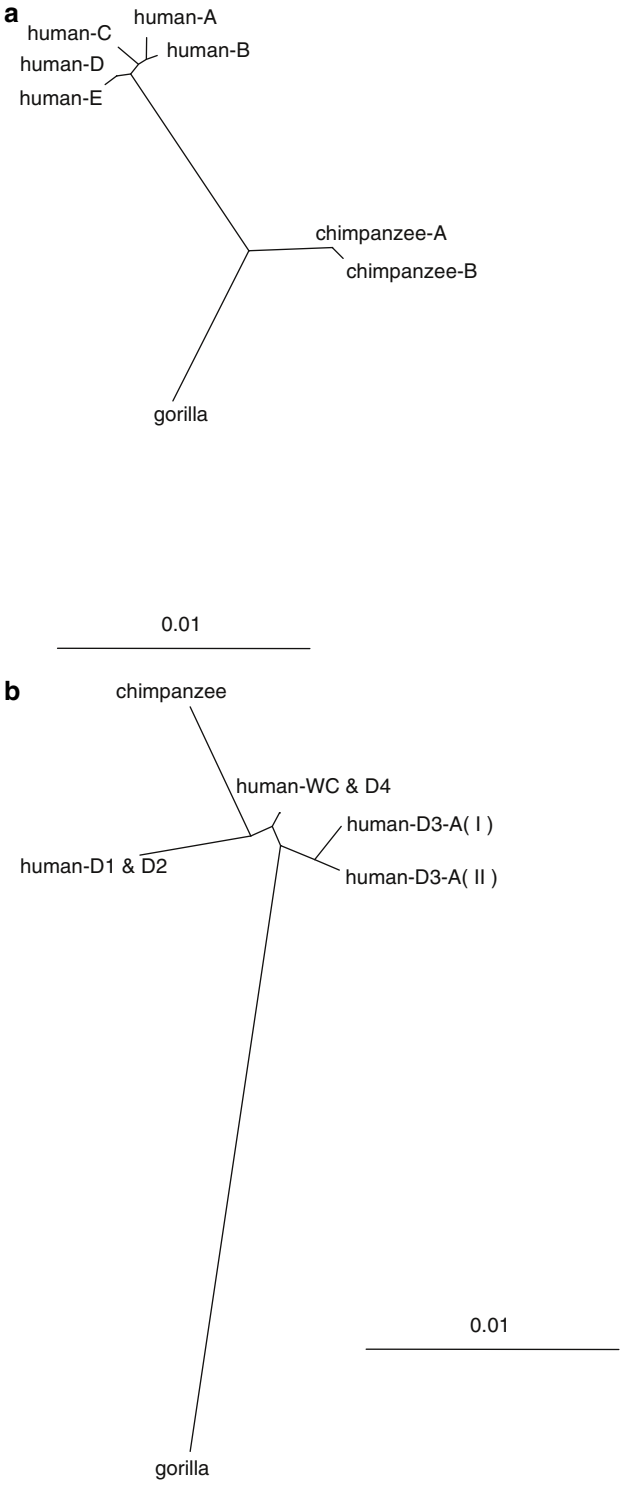

Figure 5 Neighbor-joining phylogenic trees for 1505-bp promoter sequences (**a**) and 700-bp sequences including exon 4 (**b**) in three primate species.

an alternative hypothesis, that positive natural selection has caused amino-acid substitutions at an unusually high rate, allowing long branches to emerge in the tree.

Taken together, these patterns of variation suggest that both positive and balancing natural selection have been acted on *HAVCR1*. These patterns are similar in many respects to those observed in MHC loci, which also seem to have been affected by both balancing and positive natural selection.[10] While deep divergence among alleles within human populations suggests that these alleles have been maintained for an extended period of time, the excess of nonsynonymous nucleotide substitutions suggests that the molecule has been under ongoing selective pressure.

What might have provided the selective pressure for *HAVCR1*-exon 4? The most obvious hypothesis is that HAVCR1 has been under pressure to avoid viral infection. This hypothesis is motivated by evidence that *HAVCR1* functions as a cellular receptor for the HAV and that the mucin domain of HAVCR1 is required for efficient HAV uncoating.[6,20] Variation in genes encoding cell-surface viral receptor molecules often alters susceptibility to viral infection.[21] For example, a mutation in the chemokine receptor CCR5 provides resistance to HIV infection.[22,23] Thus, sequence variation in the HAVCR1 protein might alter susceptibility to viral infection, with *HAVCR1* heterozygosity conferring an evolutionary advantage. This hypothesis is supported by an etiological study by McIntire *et al*,[8] which indicated that *HAVCR1*-exon 4 genotypes might contribute to the effects of HAV on cytokine production. An alternative hypothesis is that allergic and/or autoimmune disease may have provided the selective pressure at this locus during human evolution. Linkage analyses in mice and humans have shown that *TIM* genes are linked to susceptibility to allergic and/or autoimmune disease.[24,25] An association of *HAVCR1*-exon 4 sequence variants with susceptibility to asthma in humans[8] has also been reported.

Although the role of the HAVCR1 protein in the immune system remains poorly understood, patterns of variation in the gene suggest that important functional variants are present. As natural selection acts on variable phenotypes, the effects of natural selection will be strongest on those genes that confer such phenotypes. Thus, evidence of natural selection in the mucin domain of HAVCR1 suggests that functionally important variants are present in that region. Evidence that natural selection may have preserved functional variation in the mucin domain of HAVCR1 is consistent with the suggestion that this protein is involved in the adhesion of extracellular molecules and suggests that *HAVCR1* might have been under long-term pressure to adapt to a constantly changing environment.

Evidence that natural selection has acted to maintain functional variation in exon 4 of *HAVCR1*, in particular, suggests that linkage and association studies in which *HAVCR1* is a candidate gene may benefit from a focus on the mucin domain. As natural selection seems to have acted to preserve functional variation in this region, it is a reasonable *a priori* candidate in efforts to identify genotype–phenotype correlations. We propose that variants distinguishing the most divergent human haplotypes, such as those that distinguish the haplotype clusters in Figure 3c, will be particularly informative in such studies.

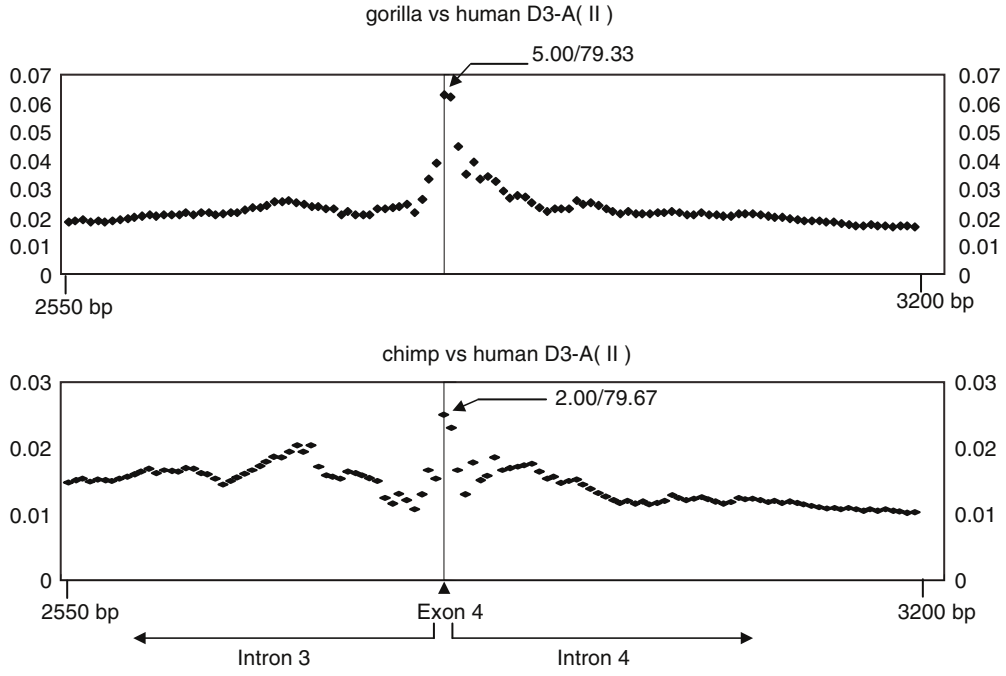

Figure 6 Nucleotide diversities among primate sequences around *HAVCR1-exon 4*, as evaluated in genomic segments with exon 4 extending every 50 bp to each side of exon 4, in introns 3 or 4.

Methods

Cell isolation and stimulation

Human $CD4^+$ Th cells were isolated from peripheral blood cells by positive selection using MACS (Miltenyi Biotech., Auburn, CA, USA) in a MACS separation column. $CD4^+$ Th cells were stimulated *in vitro* under Th1- or Th2-differentiating conditions, either with plate-bound anti-CD3 (2 μg/ml) along with IL-12 (10 ng/ml) + anti–IL-4, or anti-CD28 (2 μg/ml) along with IL-4 (10 ng/ml) + anti-IFN-γ + anti-IL-12. After 5 days of stimulation, cells were collected for preparation of mRNA and supernatants were stored at −70°C for IL-5 and IFN-γ assays. RT-PCR reactions were performed using the ThermoScript™ RT-PCR system (Invitrogen, Carlsbad, CA, USA). Cytokine concentrations were measured by ELISA using commercially available kits for IL-5 and IFN-γ (R&D Systems, Minneapolis, MN, USA).

Population samples

DNA sequences were obtained from 73 native Africans (31 Pygmy, 7 Alur, 18 Nande, and 17 Hema), 85 African-Americans, 49 Palestinians, 46 Druze, 46 Ashkenazi, 92 Caucasians, 51 Indians, 95 Japanese, 23 Korean, and 55 Northern Chinese (Heilongjian)). DNA samples from the Druze, Ashkenazi, and Palestinian populations were obtained from the National Laboratory for the Genetics of Israeli Populations (Tel Aviv University, Israel). African-American and Caucasian DNA samples were obtained from the Coriell Institute for Medical Research (Camden, NJ, USA). DNA samples from two ape species, including 16 western African chimpanzees (*Pan troglodytes verus*) and one gorilla (*Gorilla gorilla*), were also analyzed.

Identification and genotyping of nucleotide variations

Overlapping primer sets covering all exons of *HAVCR1*, ~1.5 kb of promoter sequences upstream of the transcription initiation site of *HAVCR1*, and all exons of *HAVCR2 and TIMD4*, were designed on the basis of size and overlap of PCR amplicons. Genomic DNA was subjected to PCR amplification followed by sequencing, using the BigDye Terminator cycling system. Sequencing analysis was performed in an ABI Prism 3700 automated DNA sequencer (Applied Biosystems). Sequence variations were identified by comparing sequences with the Sequencher™ program (Gene Code Co., Ann Arbor, MI, USA). Each polymorphism was confirmed by reamplifying and resequencing from the same or the opposite strand. TA cloning was performed using TOPO™ TA-cloning kits (Invitrogen, Carlsbad, CA, USA), to determine the gametic phase for individuals who were heterozygous at the *HAVCR1*-exon 4 locus.

Statistical analysis

The mean numbers of nucleotide substitutions per nonsynonymous (π_A) or synonymous site (π_S)[26] were calculated using DnaSP version 3.50[27] (available at www.ub.es/dnasp/). The Ewens–Watterson neutrality test was performed with the Arlequin computer program,[28] which is available at the ARLEQUIN web site. Alignment and neighbor-joining trees for primate sequences were inferred using the Clustal X program.[29]

Acknowledgements

We thank Yoshiko Miwa for technical assistance. This work was supported by a Future Program Grant of The Japan Society for the Promotion of Science (II, ME); by a grant from the Japanese Ministry of Public Health and

Welfare for Research on the Human Genome; by a Grant-in-Aid for Scientific Research on Medical Genome Science from the Ministry of Education, Culture, Sports, Science and Technology of Japan; and by NIH Grants GM 59290 and HL070048.

References

1 Kuchroo VK, Umetsu DT, DeKruyff RH, Freeman GJ. The TIM gene family: emerging roles in immunity and disease. *Nat Rev Immunol* 2003; **3**: 454–462.
2 Bailly V, Zhang Z, Meier W, Cate R, Sanicola M, Bonventre JV. Shedding of kidney injury molecule-1, a putative adhesion protein involved in renal regeneration. *J Biol Chem* 2002; **277**: 39739–39748.
3 Monney L, Sabatos CA, Gaglia JL *et al*. Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature* 2002; **415**: 536–541.
4 Sabatos CA, Chakravarti S, Cha E *et al*. Interaction of Tim-3 and Tim-3 ligand regulates T helper type 1 responses and induction of peripheral tolerance. *Nat Immunol* 2003; **4**: 1102–1110.
5 Sanchez-Fueyo A, Tian J, Picarella D *et al*. Tim-3 inhibits T helper type 1-mediated auto- and alloimmune responses and promotes immunological tolerance. *Nat Immunol* 2003; **4**: 1093–1101.
6 Feigelstock D, Thompson P, Mattoo P, Zhang Y, Kaplan GG. The human homolog of HAVcr-1 codes for a hepatitis A virus cellular receptor. *J Virol* 1998; **72**: 6621–6628.
7 McIntire JJ, Umetsu SE, Akbari O *et al*. Identification of Tapr (an airway hyperreactivity regulatory locus) and the linked Tim gene family. *Nat Immunol* 2001; **2**: 1109–1116.
8 McIntire JJ, Umetsu SE, Macaubas C *et al*. Hepatitis A virus link to atopic disease. *Nature* 2003; **425**: 576.
9 Das J, Chen CH, Yang L, Cohn L, Ray P, Ray A. A critical role for NF-kappa B in GATA3 expression and TH2 differentiation in allergic airway inflammation. *Nat Immunol* 2001; **2**: 45–50.
10 Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 1988; **335**: 167–170.
11 Fay JC, Wyckoff GJ, Wu CI. Positive and negative selection on the human genome. *Genetics* 2001; **158**: 1227–1234.
12 Leabman MK, Huang CC, DeYoung J *et al*. Pharmacogenetics of membrane transporters investigators. Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc Natl Acad Sci USA* 2003; **100**: 5896–5901.
13 Cargill M, Altshuler D, Ireland J *et al*. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; **22**: 231–238.
14 Slatkin M. An exact test for neutrality based on the Ewens sampling distribution. *Genet Res* 1994; **64**: 71–74.
15 Klein RG. *The Human Career: Human Biological and Cultural Origins*. The University of Chicago Press: Chicago and London, 1999.
16 Bamshad M, Wooding SP. Signatures of natural selection in the human genome. *Nat Rev Genet* 2003; **4**: 99–111.
17 Takahata N, Nei M. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 1990; **124**: 967–978.
18 Carrington M, Nelson GW, Martin MP *et al*. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 1999; **283**: 1748–1752.
19 Thursz MR, Thomas HC, Greenwood BM, Hill AV. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat Genet* 1997; **17**: 11–12.
20 Silberstein E, Xing L, van de Beek W, Lu J, Cheng H, Kaplan GG. Alteration of hepatitis A virus (HAV) particles by a soluble form of HAV cellular receptor 1 containing the immunoglobin-and mucin-like regions. *J Virol* 2003; **77**: 8765–8774.
21 Schneider-Schaulies J. Cellular receptors for viruses: links to tropism and pathogenesis. *J Gen Virol* 2000; **81**: 1413–1429.
22 Dean M, Carrington M, Winkler C *et al*. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 1996; **273**: 1856–1862.
23 Bamshad MJ, Mummidi S, Gonzalez E *et al*. A strong signature of balancing selection in the 5′ cis-regulatory region of CCR5. *Proc Natl Acad Sci USA* 2002; **99**: 10539–10544.
24 Daniels SE, Bhattacharrya S, James A *et al*. A genome-wide search for quantitative trait loci underlying asthma. *Nature* 1996; **383**: 247–250.
25 Encinas JA, Kuchroo VK. Mapping and identification of autoimmunity genes. *Curr Opin Immunol* 2000; **12**: 691–697.
26 Nei M. *Molecular Evolution Genetics*. Columbia University Press: New York, 1987.
27 Rozas J, Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 1999; **15**: 174–175.
28 Schneider S, Kueffer JM, Roesslie D, Excoffier L. *Arlequin: A Software for Population Genetic Data Analysis*. University of Geneva: Geneva, Switzerland, 2000.
29 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997; **25**: 4876–4882.