Genes and Immunity (2004) 5, 389–397

www.nature.com/gene

FULL PAPER

Local adaptation and population differentiation at the interleukin 13 and interleukin 4 loci

T Sakagami[1,2], DJ Witherspoon[3], T Nakajima[1], N Jinnai[1], S Wooding[3], LB Jorde[3], T Hasegawa[2], E Suzuki[2], F Gejyo[2] and I Inoue[1]

[1]*Division of Genetic Diagnosis, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan;* [2]*Division of Respiratory Medicine, Graduate School of Medical and Dental Sciences, Niigata University, Niigata, Japan;* [3]*Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, USA*

A 25.6 kb region at chromosome 5q31, covering the entire human interleukin 13 (IL-13) and interleukin 4 (IL-4) genes, has been reported to be associated with bronchial asthma. We have examined nucleotide variations at this locus in African, European American, and Japanese populations, using 120 diallelic variants. A block of strong linkage disequilibrium (LD) ($|D'| > 0.7$) spans a 10 kb region containing IL-4 in European American and Japanese populations, and is present but less clear in African samples. Two major haplotypes at IL-4 account for $> 80\%$ of haplotypes in European Americans and Japanese. These haplotypes are common and quite diverged from each other and the ancestral haplotype, resulting in highly significant deviations from neutrality. F_{ST} statistics show that European American and Japanese populations are unusually distinct at the IL-4 locus. The most common haplotype in the European American population is much less common in the Japanese population, and vice versa. This implies that natural selection has acted on IL-4 haplotypes differently in different populations. This selected variation at IL-4 may account for some genetic variance underlying susceptibility to asthma and other allergic diseases. The strong LD observed in the IL-4 region may allow more efficient disease-association studies using this locus.
Genes and Immunity (2004) **5,** 389–397. doi:10.1038/sj.gene.6364109
Published online 24 June 2004

Keywords: *IL-4; IL-13; haplotype; population; natural selection; local adaptation*

Introduction

Chronic airway inflammation associated with bronchial asthma [MIN 600807] is characterized by infiltration of T-lymphocytes that produce T_H2 cytokines and eosinophilic leukocytes.[1,2] The importance of T_H2 cytokines and T_H1/T_H2 balance in asthma is well recognized in patients and animal models.[3,4] Interleukin 4 (IL-4) exerts its major role in T_H2 differentiation, an initial step in IgE mediated immunity. As a downstream effect, IL-4 and interleukin 13 (IL-13) act on B cells to produce IgE. The interleukin 13 gene (*IL-13)* [MIM 147683] and interleukin 4 gene (*IL-4)* [MIM 147780] are located within a 25 kb region on chromosome 5q31.[5] Multiple cytokine genes (*IL-3*, *IL-4*, *IL-5*, *IL-12*, and *IL-13*) are located on chromosome 5q31–33, and are presumably the results of ancient duplication events. *IL-13* and *IL-4* are organized in head-to-tail manner, separated by 12 kb. They share some amino-acid similarity (NCBI Conserved Domain Database entry smart00190), are T_H2 cytokines, and both play roles in IgE-mediated immunity.[5,6] Genome-wide linkage studies have identified evidence of linkage of this region with bronchial asthma[7–9] and other inflammatory disease such as Crohn's disease.[10] Despite numerous genetic-association studies,[11–17] the involvement of the *IL-13*/ *IL-4* locus in susceptibility to allergic disease is unclear. Since the two loci are so closely linked and similar in function, it is not clear as to which might be responsible for the observed association.

With the completion of the Human Genome Project, research interests are shifting from the basic sequence of the human genome to its variation across human populations. Characterizing the nucleotide diversity of single-nucleotide polymorphisms (SNPs) and their haplotypic structure should help in mapping genetic contributions to complex diseases, as well as providing insight into human population history and evolution. A great deal of attention to linkage disequilibrium (LD) is due to the prospect of large-scale-association studies to map susceptibilities to complex diseases. Knowledge of the patterns of LD, haplotype, and population structure is useful for genotype–phenotype association analysis and population genetic analyses.[18–20] Patterns of LD are affected by recombination and gene conversion, as well as all the factors in human evolutionary history, including mutation, genetic drift, natural selection, gene flow, population expansion, and population subdivision.[19] These patterns are not uniform throughout the human genome; on the contrary, SNPs are organized into

Correspondence: Dr I Inoue, Division of Genetic Diagnosis, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.
E-mail: ituro@ims.u-tokyo.ac.jp
Received 27 January 2004; revised 08 April 2004; accepted 15 April 2004; published online 24 June 2004

irregular, partially discrete LD blocks. A significant effort is underway to construct genome-wide LD maps based on haplotype blocks.[21–23]

In the present study, sequence variation at the *IL-13/IL-4* region is examined in order to understand the structure and evolution of the genes. LD and haplotypic structure were compared among three distinct populations having distinct allergy profiles.[24] The patterns of nucleotide diversity and haplotype relationships indicate that these populations are genetically quite distinct at these loci, and that *IL-4* and, to a lesser extent, *IL-13* have been influenced by natural selection.

Results

Sequence variation at the *IL-13/IL-4* region

We determined complete sequences of the 25.6 kb genomic region including the entire *IL-13* and *IL-4* in 51 individuals (21 sub-Saharan populations, 15 European Americans, and 15 Japanese). A total of 114 SNPs and six insertion/deletion polymorphisms in total were identified in the locus (one polymorphism/193 bp). The locations of these variants are shown relative to *IL-13* and *IL-4* in Figure 1, along with the locations of repetitive sequences. The frequencies of these alleles are summarized in Table 1. Among the SNP variants, transition substitutions are more prevalent (73/114, or 64%) than transversion substitutions (41/114, or 36%). The distribution of polymorphisms in the sampled populations is summarized in Table 2. As expected, the African sample contains by far the most population-specific variants.

Tests of neutrality

Nucleotide diversity (π, Table 3) in the *IL-13/IL-4* region is typical for human populations, falling within the central 95% (π between 0.02 and 0.158%) of the empirical distribution determined by Sachidanandam *et al*[25] (Figure 2b). Diversity tends to be higher in the African population sample, as expected.

Tests of neutrality, based on Tajima's D[26] and Fu and Li's D^* and F^*[27] statistics, were conducted for each population sample separately (see Materials and methods). The *IL-13/IL-4* region was tested both in its entirety and subdivided into three functional domains: SNPs 1–19 in the *IL-13* locus, SNPs 20–70 in the intergenic region (IGR) and SNPs 71–120 representing the *IL-4* locus. Results of those tests (observed statistics and two-tailed P-values, not corrected for nonindependence of tests) are shown in Table 3. A simple Bonferroni correction suggests that only results with $P<0.005$ should be considered significant. In all cases, the observed values of the statistics are larger than their expected values, indicating an excess of common polymorphisms, which in turn suggests balancing selection. These excesses are significant in several cases, noted in Table 3. The African population does not show clear evidence of selection in any region, but there is strong evidence of balancing selection within the European American and Japanese populations at the *IL-4* region. This evidence is strongest in the Japanese population.

Under assumptions of constant human population size, little or no evidence of selection would have been detected. However, much independent evidence (reviewed in Harpending and Rogers[28]) shows that human populations have grown in recent history, and that such growth can obscure the effects of certain kinds of natural selection. In particular, population growth produces a pattern of nucleotide variation that mimics that produced by directional (positive) selection, which results in negative values for these statistics. As a result, a majority of the genes in a large survey of 3899 SNPs in 82 individuals of different origins showed a negative Tajima's D-value.[29] This finding is consistent with growth in the human population. Using simulations[30] to account for the effect of a conservative population growth model, the positive values in Table 3 become highly unexpected and thus constitute strong evidence of balancing selection.

Population differentiation at *IL-13/IL-4* locus

Population subdivision was investigated with F_{ST} statistics (see Materials and methods), using the 31 common SNPs and the larger population samples (96 African American, 96 European American, and 96 Japanese subjects) in which these SNPs were typed (Table 4). The genotype frequencies for all of the SNPs met Hardy–Weinberg expectations in all populations (data not shown). The F_{ST} analysis showed that essentially all populations are clearly distinguishable from one another ($P<0.001$ by resampling) within each region, while the African American and African populations showed only marginally significant differentiation in the *IL-4* region ($P=0.04$). Using F_{ST} as a genetic distance, the African and African American populations are most similar and the European American and Japanese populations are remarkably different. F_{ST} values in the *IL-13* region are typical for humans (between 0.11 and 0.16),[31] but F_{ST} values in the IGR and *IL-4* regions are unusually high (compare to Bowcock *et al*[32] and Akey *et al*,[33] who found the typical world-wide F_{ST} at noncoding SNPs to be

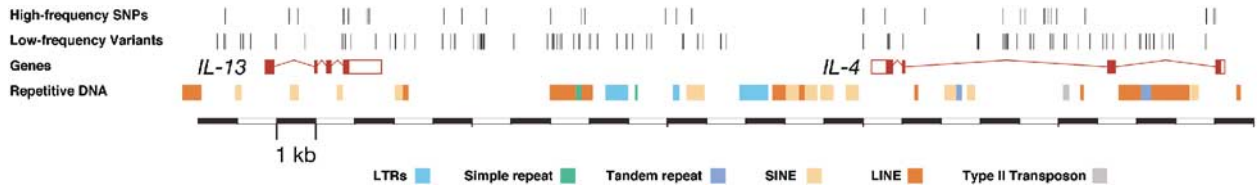

Figure 1 Schematic diagram of genomic structure, repeat elements, and variant sites in the region containing *IL-13* and *IL-4*, based on Ensembl annotation (www.ensembl.org; Hubbard *et al*[61]). Locations of the common SNPs, genotyped in the larger population samples, are shown on the first line ('high-FREQUENCY SNPs'), and the remaining lower-frequency SNPs and other variants are shown on the next line. The mRNA structures of *IL-13* and *IL-4* are shown next (solid regions denote exons, open boxes denote untranslated regions, and introns are shown as connecting lines). Both genes are transcribed from left to right. Various types of repeats are shown last, identified by color according to the legend. All features are shown according to the scale bar, but some variant features are so close together as to be indistinguishable.

Table 1 Frequency of nucleotide variants at SNPs in three population samples

SNP	*Position*[a]	*Variant*[b]	*African* ($n=42$)[c]	*European* ($n=30$)	*Japanese* ($n=30$)
1	49813	G/A	0.33		
2	49643	A/T	0.05		
3[d]	49612	C/T	0.36	0.93	0.80
4	49603	T/A	0.02		
5	49215	G/A	0.05		
6	49146	G/A	0.21		
7	48958	A/C	0.02		
8	48317	C/T	0.17		
9[d]	47986	C/A	0.79	0.93	0.77
10[d]	47959	C/G	0.36	0.07	0.07
11[d]	47752	T/C	0.50		0.10
12	47567	A/C	0.02		
13	46608	A/G	0.05		
14[d]	46578	C/T	0.19	0.83	0.70
15	46538	G/A	0.05		
16[d]	46457	A/G	0.10	0.17	0.30
17	46396	A/G	0.05		
18[d]	45976	A/G	0.02	0.17	0.30
19[d]	45921	C/A	0.36	0.83	0.70
20	45752	C/T	0.95	0.83	0.70
21	45387	G/A	0.10		
22	45263	A/G	0.02		
23	45252	T/C	0.31		
24	45005	G/C	0.02		
25	44752	G/C		0.10	
26	44046	T/C	0.12		
27[d]	44008	T/C	0.02	0.07	0.23
28	43835	G/A	0.05		
29	43729	A/G	0.29		
30[d]	43545	T/C	0.05	0.07	0.13
31	43250	T/A		0.07	
32	43129	G/A	0.07		
33	43070	T/C	0.07		
34	43038	A/G	0.52		
35	42995	ins A[e]	0.02		
36	42984	A/G	0.02		
37	42971	T/C	0.07		
38[d]	42927	G/A	0.17	0.93	0.83
39	42220	T/G	0.02		
40	41356	A/G	0.14		
41[d]	41270	C/T	0.33	0.07	0.13
42	41232	A/G	0.07		
43	41220	G/C	0.02		
44	41195	A/C		0.03	
45	41107	A/G	0.14		
46	41035	C/T	0.67	0.97	0.93
47	40999	T/G	0.10	0.03	0.07
48	40682	T/C	0.12		
49	40542	T/A	0.02		
50[d]	40488	A/T	0.38	0.07	0.10
51[d]	40397	C/T	0.38	0.93	0.90
52	40296	T/C	0.21	0.93	0.90
53	40258	C/G	0.07		
54	40073	G/A	0.07		
55	39577	A/G	0.52	0.07	0.10
56	39334	T/C	0.12		
57	39173	A/G	0.12		
58	38811	T/C	0.79	1.00	1.00
59	38700	del 9 bp[e]	0.71		
60[d]	38335	A/C	0.07	0.10	0.10
61[d]	38058	G/T	0.33	0.10	0.17
62	37890	T/C		0.07	
63	37869	A/G	0.05		
64[d]	37665	G/T	0.93	0.90	0.83
65	37518	G/A			0.03
66	37491	T/C	0.02		
67	37374	T/C	0.02		
68	36920	A/T			0.03
69	36785	G/A			0.03
70	33270	C/T	0.05		
71[d]	33267	T/C	0.67	0.27	0.43
72[d]	32711	C/T	0.69	0.73	0.53
73	32636	T/C			0.03
74	32634	G/A	0.07		
75	32598	T/C			0.03
76	32592	T/C	0.05		
77	32249	A/G			0.07
78[d]	31695	T/C	0.29	0.23	0.47
79	30341	C/G	0.02		
80	30316	A/G			0.03
81	30311	T/C	0.05		
82[d]	29696	G/T		0.70	0.37
83	29681	T/A	0.02		
84	29615	T/G	0.31		
85	29566	T/G	0.02		
86	29328	A/G		0.03	
87[d]	29242	G/C		0.70	0.37
88[d]	29122	C/G	0.29	0.07	0.17
89	28987	del A[e]	0.98	1.00	
90	28903	T/C	0.02		
91[d]	28632	A/G	0.29	0.23	0.47
92[d]	28535	C/G	0.29	0.23	0.47
93	28474	G/C	0.02		
94[d]	28458	C/A	0.26	0.23	0.47
95	28407	del AA[e]	0.74	0.77	0.53
96[d]	28312	A/G	0.29	0.77	0.53
97	28122	A/G	0.05		
98	28067	C/T	0.05		
99	27762	C/G	0.05		
0[d]	27589	G/A	0.29	0.77	0.53
101	27466	A/G	0.02		
102	27128	T/A	0.05		
103	27030	C/T	0.98	1.00	1.00
104	26557	A/G	0.02		
105	26194	A/G	0.02		
106	26162	ins G[e]	0.02		
107	26026	C/T	0.21	0.07	0.17
108	26016	A/G	0.02		
109	25936	del 70 bp[e]	0.69	0.73	0.50
110	26867	A/C	0.29	0.27	0.47
111	25828	G/C	0.21		0.13
112	25615	T/A	0.02		
113	25429	A/G	0.71	0.27	0.47
114	25428	T/G	0.29	0.07	0.17
115	25335	C/T	0.05		
116	24498	G/C	0.05		
117	24492	A/C		0.07	
118[d]	24477	G/A	0.71	0.27	0.47
119[d]	24289	G/A	0.26	0.27	0.47
120[d]	24252	A/C	0.29	0.27	0.47

[a]Position is given according to the numbering in Genbank entry AC004039.
[b]Frequencies of the derived allele (listed first, relative to chimpanzee) are shown.
[c]Sample sizes n in number of chromosomes.
[d]Common SNPs chosen for genotyping in larger population samples.
[e]Insertions (ins) and deletions (del).

approximately 0.12, although values as high as 0.3 are not unusual). These unusually high F_{ST} values suggest that natural selection has favored different haplotypes in different populations, that is, that local adaptation has occurred.

Table 2 Nucleotide variants in three populations

	African	*European*	*Japanese*	*Pooled*
n (chromosomes)	42	30	30	102
SNPs	99	45	48	114
Indels	6	2	2	6
Singletons	24	4	6	30
Doubletons	18	17	4	20
Population-specific variants	64	6	7	0
Heterozygosity of MS1[a]	0.768	0.547	0.759	0.698

[a]MS1 denotes a microsatellite repeat in *IL-4*.

Haplotype structure at *IL-13* and *IL-4*

Haplotypes and their population frequencies were inferred from genotypes at the 31 common SNPs in the larger population samples (96 African American, 96 European American, and 96 Japanese subjects) in which these SNPs were typed. Haplotypes were inferred for each population separately, and for each of the three functional regions assigned above, as well as for the entire region simultaneously. The EM algorithm, and its implementation in PHASE, have been shown to be accurate when alleles are in Hardy–Weinberg equilibrium and large sample sizes (eg >100 chromosomes) are used.[34–36] The SNPs typed here are all in Hardy–

Table 3 Nucleotide diversities and tests of neutrality for each region and population

	IL–13 (4238 bp)			*Intergenic region (11 138 bp)*			*IL–4 (9015 bp)*			*Entire region (24 391 bp)*		
	African	*European*	*Japanese*	*African*	*European*	*Japanese*	*African*	*European*	*Japanese*	*African*	*European*	*Japanese*
n[a]	42	30	30	42	30	30	42	30	30	42	30	30
S[b]	19	7	8	42	18	17	38	20	23	99	45	48
s[c]	4	0	0	9	3	3	11	1	3	24	4	6
θ_T $(\pi)\pm\sigma$, $\times 10^4$[d]	11 ± 5.9	3.6 ± 2.5	6.5 ± 3.9	7.8 ± 4.2	2.3 ± 1.4	3.2 ± 1.9	9.8 ± 5.1	7.2 ± 3.9	10 ± 5.3	9.1 ± 4.5	4.3 ± 2.3	6.3 ± 3.2
$\theta_W\pm\sigma$, $\times 10^4$[e]	10 ± 3.7	4.2 ± 2.0	4.8 ± 2.2	8.8 ± 2.8	4.1 ± 1.6	3.9 ± 1.5	9.8 ± 3.2	5.6 ± 2.1	6.4 ± 2.3	9.4 ± 2.8	4.7 ± 1.6	5.0 ± 1.7
Tajima's D	0.063^{NS}	-0.39^{NS}	$1.1^{\#}$	-0.32^{NS}	-1.5 (0.52)	-0.60^{NS}	$-0.0062^{\#}$	$1.0^{\#\#}$	$2.0^{\#\#\#}$	$-0.13^{\#}$	-0.26^{NS}	$1.0^{\#\#\#}$
Fu and Li's D*	0.1^{NS}	$1.3^{\#}$	$1.3^{\#\#}$	0.19^{NS}	$0.58^{\#}$	$0.51^{\#}$	-0.4^{NS}	$1.3^{\#\#}$	$0.83^{\#}$	$-0.037^{\#}$	$1.2^{\#\#\#}$	$0.95^{\#\#\#}$
Fu and Li's F*	0.16^{NS}	$0.78^{\#}$	$1.3^{\#}$	0.013^{NS}	-0.098^{NS}	$0.17^{\#}$	-0.3^{NS}	$1.3^{\#\#}$	$1.3^{\#\#}$	$-0.081^{\#}$	$0.80^{\#\#\#}$	$1.1^{\#\#\#}$

[a]Number of chromosomes.
[b]Number of segregating sites.
[c]Number of singletons.
[d]Tajima's (1983) estimator of θ, per site; equal to nucleotide diversity π.
[e]Watterson's (1975) estimator of θ, per site.
$^{NS}P>0.01$.
$^{\#}P\leqslant0.01$.
$^{\#\#}P\leqslant0.001$.
$^{\#\#\#}P\leqslant0.0001$.

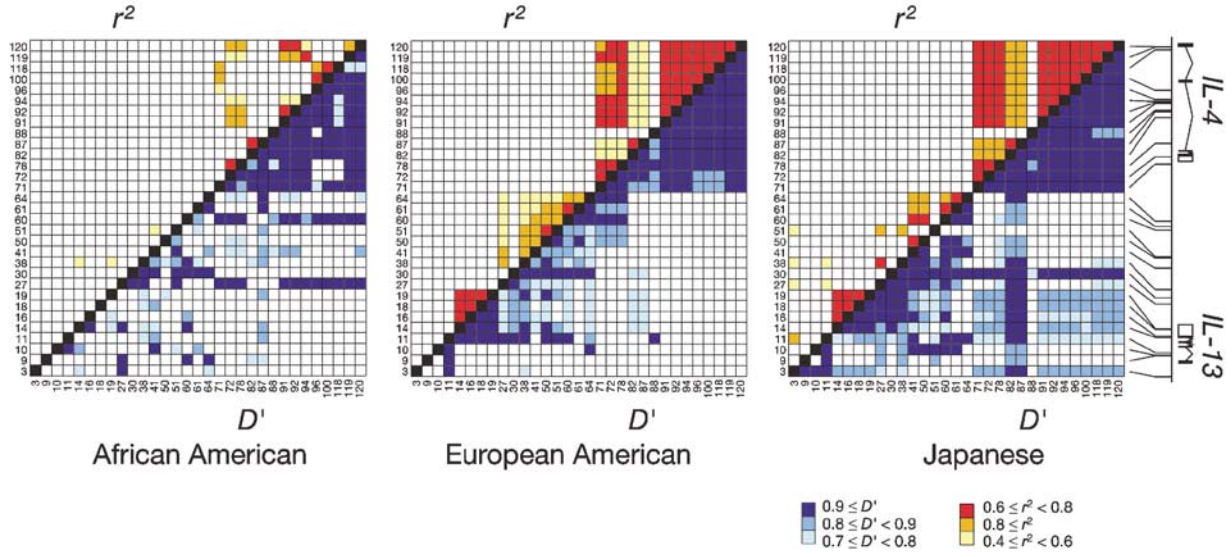

Figure 2 Pairwise LD statistics D' and r^2 for pairs of common SNPs in the region containing *IL-13* and *IL-4* for three populations. $|D'|$ is shown in the lower triangular region and r^2 in the upper triangular region, with their values represented by the same color scale according to the legend. The SNPs being compared are listed along the axes. LD statistics were not estimated for Africans due to the smaller sample size.

Table 4 F_{ST} and s.d. estimates for sets of populations and SNPs

Region	*No. of SNPs*	*African American, European, Japanese*	*African American, African*	*African American, European*	*African American, Japanese*	*European, Japanese*
IL-13	8	0.10±0.0086	0.046±0.0053	0.15±0.018	0.10±0.0097	0.040±0.0029
IGR	9	0.35±0.047	a	0.17±0.037	0.37±0.055	0.47±0.075
IL-4	14	0.27±0.0031	0.017±0.0021	0.25±0.013	0.12±0.0093	0.40±0.0034
All	31	0.25±0.0059	a	0.21±0.0060	0.18±0.0084	0.35±0.0092

[a]The expanded sample of Africans was not typed at the common IGR SNPs.

Table 5 *IL-4* haplotypes and their estimated frequencies in four populations

	IL-4 haplotype	*African American (n=192)*	*African (n=60)*	*European (n=192)*	*Japanese (n=192)*
4H01	TTTTCGACCGAGGA	0.225	0.257	0.15	0.666
4H02	CCCGGGGGAAGAAC	0.145	0.009	0.697	0.25
4H03	TCCTCGGGAGAGAC	0.234	0.273		
4H04	CCCTCCGGAAGAAC	0.111	0.216	0.088	0.062
4H05	TTTTCGACAGAGAA	0.097	0.092		
4H06	TTTTCGGGAAGAAC	0.063		0.005	
4H07	CCCTCCGGAGGAAC	0.029	0.017		
4H08	CCCTCGGGAGAGAC		0.033	0.01	
4H09	TTCTCGGGAGAGAC		0.035		
4H10	TTTTCGACCGAAGA	0.026		0.005	
4H11	CCCGCGGGAAGAAC	0.016	0.003		
4H12	TCTTCGACCGAGGA		0.018		
4H13	CCCTCCGGAAGGAC		0.017		
4H14	CCTTCCGGAGGAAC	0.016			
Chimp	CTCTCGGGAGAAAC				
	IL-13 haplotype				
13H01	CCGCCGGC	0.227	0.002	0.611	0.405
13H02	CAGCCGGC	0.039	0.046	0.096	0.205
13H03	CCGTTGGA	0.119	0.074	0.01	
13H04	TCGCTAAA	0.003		0.105	0.055
13H05	TCGCCGGC	0.043	0.065	0.043	
13H06	TCGTTGGC	0.05	0.092		
13H07	TACCTGGA	0.073	0.051		
13H08	CCGCTAAA			0.033	0.078
13H09	TCGTTAAA				0.102
13H10	TCGTTGGA	0.028	0.07		
13H11	CCGCTGGC	0.009	0.07	0.01	
13H12	TCCCTGGA	0.031	0.056		
13H13	CCCTTGGA	0.012	0.063		
13H14	CCGTTAGA	0.018	0.057		
13H15	TCGCTAGA	0.07	0.001		
13H16	TCGTTAGA		0.071		
13H17	CCCCTAAA		0.001	0.008	0.044
13H18	TCGCTGGA	0.045	0.004		
13H19	CACCTAAA			0.015	0.031
13H20	TCCCTGGC		0.044		
13H21	CACCCGGC	0.003			0.038
13H22	CCGCTGGA	0.035	0.005		
Chimp	TAGCTGGA				

Weinberg equilibrium (not shown). The haplotypes and frequencies inferred for the *IL-13* and *IL-4* regions separately are shown in Table 5. The African sample size was small (60 chromosomes), so these haplotype estimates should be considered preliminary.

In both the *IL-13* and *IL-4* regions, more haplotypes are inferred for the African population than either the European American or Japanese populations, presumably due to the greater age and diversity of the African population. The African American population has even greater diversity; this may be due to the creation of new haplotypes by admixture and recombination with the European population.

In the *IL-13* region, there is a single most common haplotype in European American and Japanese populations (13H01, at 61 and 40% frequencies, respectively). This haplotype is less common in the African American population (23%) and nearly absent in the African population. The most common African haplotype (13H06) reaches only 9.2% frequency in Africans, and is absent in European American and Japanese populations. These frequency gradients are consistent with admixture

between the European American and African populations in the African American population.

At *IL-4*, two haplotypes (4H01 and 4H02) are common in both the European American and Japanese populations and account for over 80% of haplotypes in both those populations. However, 4H01 is far more common in Japanese than 4H02, and while the opposite is true in European Americans. These two haplotypes are among the most divergent from the presumed ancestral haplotype (represented by the chimpanzee sequence). While 4H02 is nearly absent in Africans, the most common haplotype in Africans and African Americans (4H03) is absent in European Americans and Japanese.

Pairwise LD in *IL-13* and *IL-4* region

Measures of LD using D' and r^2 statistics, based on simultaneous haplotype estimates for the full set of 31 common SNPs, were computed for all possible SNP pairs in the larger population samples (Figure 2). These statistics were also computed from haplotype estimates performed on the region-specific subsets of SNPs (above) and no significant difference in the patterns of LD were observed (not shown). LD is lowest in the African American samples, highest in Japanese, and intermediate in European Americans. A clear haplotype block in strong LD spans at least 10 kb in the *IL-4* region. Weaker LD is present between some SNPs in the *IL-13* region, and there is very little LD between the *IL-13* and *IL-4* regions. LD measured by r^2 shows a similar pattern, but drops off more rapidly than D', especially in the African American population.

Haplotype network for *IL-4*

The strong LD observed in the IL-4 region allows construction of an interpretable reduced-median network of the haplotypes,[37] shown in Figure 3. A haplotype constructed from the chimpanzee sequence at the common *IL-4* SNPs is added to suggest an ancestral haplotype. That haplotype lies in the center of the network, while the two most common haplotypes (4H01 and 4H02) lie at opposite extremes of the network. This pattern suggests that these two haplotypes have been maintained for some time by balancing selection.

The numerous, low-frequency haplotypes inferred for the *IL-13* and IGR regions (not shown) and the absence of strong LD in those regions (Figure 2) suggest that recombination has been common in those regions. Consistent with this, reduced-median networks for haplotypes in those regions show a high degree of reticulation (not shown), and are therefore less interpretable than the *IL-4* network.

Discussion

Owing to the difficulty of identifying loci responsible for genetic variation in susceptibility to complex diseases, researchers have begun searching for genes that have been affected by natural selection during the recent evolutionary history of humans.[38,39] Evidence of selection at a locus indicates that its alleles (or functional variants in LD with those alleles) were responsible for significant phenotypic variation, and it is reasonable to suspect that those alleles may now underlie current variation. Thus, genes responsible for genotype–phenotype relationships in malaria resistance, psychiatric disorders, and other diseases affected by recent natural selection have been reported.[38,40–45]

The observations with regard to *IL-13* and *IL-4* are: the two regions are not in LD, while *IL-4* is contained in a block of strong LD; tests of neutrality (eg tests of Tajima's *D*) within populations detect strong deviations from neutrality at the *IL-4* region, consistent with the two common but divergent haplotyes present at that locus, while *IL-13* shows at most weak deviation from neutrality. In addition, the *IGR* and *IL-4* regions show unusually strong population differentiation (high F_{ST}), especially between the European and Japanese populations.

The existence of unexpectedly common yet diverged haplotypes within populations suggests that balancing selection has maintained those haplotypes in the population. Recently, Zhang *et al*[46] have found that a particular pattern of highly diverged haplotypes is common in the human genome and is consistent with

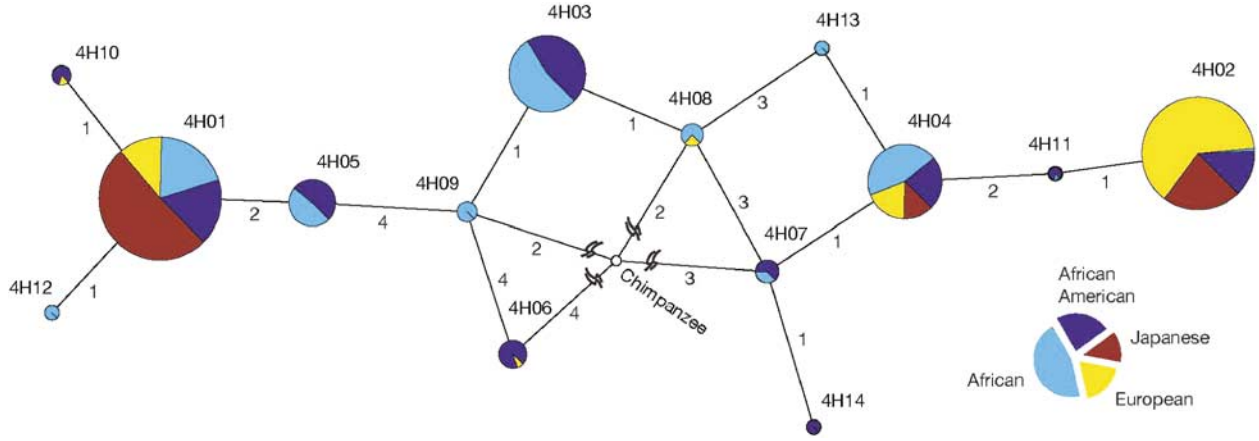

Figure 3 Reduced-median network of the 14 most common *IL-4* haplotypes inferred from the 14 common *IL-4* SNPs. Haplotypes are labeled as in Table 5. The size of each node is proportional to the average frequency of the haplotype across the four sampled populations. The frequency of each haplotype within each population is indicated by color according to the legend. The number of SNP differences between the haplotypes is shown adjacent to each link between nodes. The 'chimpanzee' haplotype represents the nucleotides present in the chimpanzee sequence at these 14 SNPs only; fixed differences from the human sequence are not reflected in the branch lengths.

neutral evolution. Based on these results, it is suggested that high-frequency pairs of highly divergent haplotypes—similar to those observed here in the *IL-4* region—are not evidence of ancient population subdivision or balancing selection. However, the evidence that *IL-4* has not been evolving neutrally does not depend on the pattern shown in Figure 3, but on coalescent simulations and statistical tests specifically designed to detect deviations from neutrality. The measures examined by Zhang *et al*[46] are very different from these statistics, and while they show a pattern worth investigating, their relevance to statistical tests of neutrality is unclear.

Unexpectedly high or low F_{ST} can also be a sign of natural selection.[32,44] F_{ST} is unusually high at the *IL-4* locus compared to other estimates of world-wide F_{ST}, and one of the common haplotypes, 4H02, has a very low frequency in our African samples. It may be that the Japanese population is quite differentiated from others due only to founder effects and drift, and our African sample may be too small to represent accurately the frequency of 4H02.

Alternatively, different mechanisms of selection may be operating in different populations, and so the different haplotype frequencies might be the result of local adaptation. For example, 4H01 may have been selectively favored in Japanese, where it is most common, while 4H02 may have been favored in European Americans, yet nearly lost in Africans. The persistence of both haplotypes at high frequencies in both populations, which suggests balancing selection, may be due to migration between the populations or simply lack of time for the hypothesized selective sweeps to achieve fixation of the selected haplotypes.

Recently, Zhou *et al*[47] reported that Chinese, European, and African populations are unusually differentiated at the *IL-13* locus, according to F_{ST} statistics. No significant excess of common yet divergent haplotypes was noted. To the extent that Chinese and Japanese populations may be similar, these observations are roughly consistent with those reported here for *IL-13*. The data suggest that different selective pressures have been acting on this locus in Asian *vs* European populations. However, the lack of LD between *IL-13* and *IL-4* regions implies that their evolutionary histories are largely independent, and so conclusions with regard to *IL-13* should not be applied to *IL-4*.

Rockman *et al*[48] suggest that a common SNP in the *IL-4* promoter region (−524C/T, our SNP 71) has been under locally adaptive positive selection within the human population. This is based on the observation that the frequency of the derived allele (−524T) is very different between some populations, in particular Asian and European ones. This results in F_{ST} values that could not be explained by models of neutral evolution and was found to be exceptional when compared with a set of 18 neutral markers. A similar pattern is apparent in many SNPs in our data set, especially in the *IGR* and *IL-4* regions between the Japanese and European populations (Table 1). The averaged F_{ST} values in these comparisons (Table 4) are nearly as high as those reported by Rockman *et al*[47] for –524 alone. In the haplotype network of Figure 3, the –524T allele is found in all human haplotypes from 4H09 to the left, so this polymorphism coincides with the two main haplotype groups we have observed. There is only one *IL-4* amino-acid polymorphism in our data set (SNP 77), and it is at a very low frequency. There is evidence that the –524 SNP has functional impacts,[49] so this may be the SNP that is being acted on by natural selection.

The evidence that natural selection has acted differently on two common haplotype groups at *IL-4* suggests that it is more likely to be responsible for the association of this region with asthmatic disease than *IL-13*, which shows weaker evidence of selection. Given the difference in frequency of the two common *IL-4* haplotypes between European Americans and Japanese, one might expect different prevalences of genetic susceptibility to disease in these two populations. The strong LD in the *IL-4* region implies that very few SNPs would need to be typed to test for associations of phenotypes with *IL-4* haplotypes. The evidence of local adaptation at *IL-4* suggests that the two most common *IL-4* haplotypes have functional consequences, and so a marker distinguishing these two haplotypes (perhaps −524C/T itself) might capture most of the functionally significant variation at this locus. The extent of LD on the 3′ side of *IL-4* is yet to be determined.

The increasing prevalence of asthma in Western countries has led to the 'hygiene hypothesis',[2] which suggests that improved hygiene in industrialized societies has reduced the incidence of infections that might stimulate the T_H1-immune system to protect against asthma. It is possible that changes in the incidence and type of infections such as tuberculosis and hepatitis A virus infections that trigger imbalance between T_H1 and T_H2 cells could cause adaptive changes at *IL-13* and *IL-4* loci. Further investigation of the evolutionary implications of *IL-13* and *IL-4* in disease-association studies for asthma and other allergic diseases are required to clarify the consequences of evolutionary pressure on *IL-13* and *IL-4*, and these studies should benefit from the markers of functional variation identified here.

Materials and methods

Subjects

In total, 96 Japanese individuals unselected for disease status were recruited from Niigata University Hospital in Niigata City, Japan. Each subject gave informed consent, and the study was performed under the approval of the Ethical Committees at the School of Medicine, Niigata University and the Institute of Medical Science, The University of Tokyo. Blood samples were collected for isolation of genomic DNA. A total of 21 more subjects were unrelated individuals from sub-Saharan Africa (10 Pygmy, six Alur, and five Nande).[50] In all, 96 European American, 96 African American and nine more subjects from sub-Saharan Africa were from the human variation panel of Northern Europeans, African Americans, and sub-Sahara Africans (Coriell Cell Repository, Camden, NJ, USA). Samples from three chimpanzees (*Pan troglodytes verus*) were also used.

Identification and genotyping of nucleotide variations

Overlapping primer sets covering the genome sequence of the *IL-13*/*IL-4* locus (Genbank accession number, AC004039, from position 23931 to 50151) were designed.

Genomic DNA was subjected to PCR amplification followed by sequencing using BigDye Terminator cycle sequencing and an ABI PRISM 3700 DNA analyzer (Applied Biosystems, Tokyo, Japan). A subset of the samples above was completely sequenced in this region: 42 chromosomes from African, 30 from European American, and 30 from Japanese samples. Polymorphisms were identified by comparing these sequences using the Sequencher™ program (Gene Codes Co, Ann Arbor, MI, USA). Each polymorphism was confirmed by reamplifying and resequencing. All amplifications and sequencing primers were designed using the Primer 3 web site. A selected set of common SNPs (SNPs at which the minor allele was present with frequency of at least 0.10 in at least two of the populations) were genotyped in the full set of Japanese, European, African, and African American samples. The *IL-13/IL-4* locus was sequenced in one chimpanzee (GenBank accession number, AY480012) and the common SNPs were genotyped in two more chimpanzees.

Statistical analyses

The population parameters, θ (per nucleotide site, along with its standard deviation (s.d.)) and π (nucleotide diversity), were estimated according to Watterson and Nei and Li, respectively.[51,52]

To determine if nucleotide variation in the sequenced region is compatible with the null hypothesis of neutral evolution, we compared observed values Tajima's D[26,53] and Fu and Li's D^* and F^*[27] statistics to their expected distributions, derived by simulating the coalescent process under a null model of neutral evolution in a modestly growing population.[30] The human population was assumed to have grown by only 10-fold 100 000 years ago from an initial effective size of 5000 individuals (consistent with Excoffier and references therein),[54] with 20-year generations and a per-generation, per nucleotide mutation rate of 1.8×10^{-8}.[55] The expected distributions of Tajima's D and Fu and Li's statistics shift towards more negative values in models with population growth, and the shift is more pronounced under models with greater population growth.[30] Balancing selection produces positive values of the test statistics, and the values observed for our data are mostly positive; therefore, this choice of demographic parameters results in conservative tests of neutrality. To the extent that recombination has occurred between SNPs in our data, the tests become more conservative.[52,56]

As a measure of population differentiation, Wright's F_{ST} was estimated from the genotypic data for the common SNPs for sets of populations according to the method of Weir and Cockerham,[57] and variances of F_{ST} were estimated using the delete-one jackknife procedure recommended there. To determine whether populations were significantly differentiated, the observed F_{ST} was compared to a null distribution for F_{ST} generated by resampling allelic states for SNPs across the populations being compared.

Haplotypes were inferred from the genotypes at the common SNPs in each of the larger population samples separately, using PHASE 2.0.2[58] with the recommended parameters. The inferred haplotypes and their estimated population frequencies were then used to estimate LD measures. To estimate the degree of LD between SNPs, Lewontin's disequilibrium coefficient (D')[59] and an LD correlation coefficient (r^2)[60] were used.

Acknowledgements

This work was supported by a Research for the Future Program Grant of The Japan Society for the Promotion of Science (II), NIH Grant RO1 GM59290 (LBJ), and NSF Grant BCS-0218370 (LBJ). We are grateful to Miho Kakihara for technical assistances and Dr Atsushi Tajima for critical reading of the manuscript.

References

1 Holt PG, Macaubas C, Stumbles PA, Sly PD. The role of allergy in the development of asthma. *Nature* 1999; **402** (Suppl): B12–B17.
2 Busse WW, Lemanske RF. Asthma. *N Engl J Med* 2001; **344**: 350–362.
3 Mosmann TR, Sad S. The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunol Today* 1996; **17**: 138–146.
4 Umetsu DT, McIntire JJ, Akbari O, Macaubas C, DeKruyff RH. Asthma: An epidemic of dysregulated immunity. *Nat Immunol* 2001; **3**: 715–720.
5 Smirnov DV, Smirnova MG, Korobko VG, Frolova EI. Tandem arrangement of human genes for interleukin-4 and interleukin-13: resemblance in their organization. *Gene* 1995; **155**: 277–281.
6 Zurawski G, de Vries JE. Interleukin 13, an interleukin 4-like cytokine that acts on monocytes and B cells, but not on T cells. *Immunol Today* 1993; **15**: 19–26.
7 The Collaborative Study on the Genetics of Asthma. A genome-wide search for asthma susceptibility loci in ethnically diverse populations. *Nat Genet* 1997; **15**: 389–392.
8 Ober C, Cox NJ, Abney M *et al*. Genome-wide search for asthma susceptibility loci in a founder population. *Hum Mol Genet* 1998; **7**: 1393–1398.
9 Yokouchi Y, Nukaga Y, Shibasaki M *et al*. Significant evidence for linkage of mite-sensitive childhood asthma to chromosome 5q31–q33 near the interleukin 12 B locus by a genome-wide search in Japanese families. *Genomics* 2000; **66**: 152–160.
10 Rioux JD, Daly MJ, Silverberg MS *et al*. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001; **29**: 223–228.
11 Heinzmann A, Mao XQ, Akaiwa M *et al*. Genetic variants of IL-13 signaling and human asthma and atopy. *Hum Mol Genet* 2000; **9**: 549–559.
12 Graves PE, Kabesch M, Halonen M *et al*. A cluster of seven tightly linked polymorphisms in the IL-13 gene is associated with total serum IgE levels in three populations of white children. *J Allergy Clin Immunol* 2000; **105**: 506–513.
13 Howard TD, Whittaker PA, Zaiman AL *et al*. Identification and association of polymorphisms in the interleukin-13 gene with asthma and atopy in a Dutch population. *Am J Respir Cell Mol Biol* 2001; **25**: 377–384.
14 Walley AJ, Cookson WO. Investigation of an interleukin-4 promoter polymorphism for associations with asthma and atopy. *J Med Genet* 1996; **33**: 689–692.
15 Laitinen T, Kauppi P, Ignatius J *et al*. Genetic control of serum IgE levels and asthma: linkage and linkage disequilibrium studies in an isolated population. *Hum Mol Genet* 1997; **6**: 2069–2076.
16 Noguchi E, Shibasaki M, Arinami T *et al*. Association of asthma and the interleukin-4 promoter gene in Japanese. *Clin Exp Allergy* 1998; **28**: 449–453.
17 Elliott K, Fitzpatrick E, Hill D *et al*. The −590C/T and −34C/T interleukin-4 promoter polymorphisms are not associated with atopic eczema in childhood. *J Allergy Clin Immunol* 2001; **108**: 285–287.

18 Jorde LB. Linkage disequilibrium and the search for complex disease genes. *Genome Res* 2000; **10**: 1435–1444.
19 Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002; **3**: 299–309.
20 Nakajima T, Jorde LB, Ishigami T *et al*. Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations. *Am J Hum Genet* 2002; **70**: 108–123.
21 Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**: 229–232.
22 Patil N, Berno AJ, Hinds DA *et al*. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; **294**: 1719–1723.
23 Gabriel SB, Schaffner SF, Nguyen H *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
24 Weiss KB, Gergen PJ. The epidemiology of asthma. In: Blumenthal MN, Bjorksten B (eds) *Genetics of Allergy and Asthma*. Marcel Dekker Inc.: New York, 1997, pp 137–169.
25 Sachidanandam R, Weissman D, Schmidt SC *et al*. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; **409**: 928–933.
26 Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 1983; **105**: 437–460.
27 Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics* 1993; **133**: 693–709.
28 Harpending H, Rogers A. Genetic perspectives on human origins and differentiation. *Ann Rev Genomics Hum Genet* 2000; **1**: 361–385.
29 Stephens JC, Schneider JA, Tanguay DA *et al*. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 2001; **293**: 489–493.
30 Wooding S, Un-kyung K, Bamshad M, Larsen J. Natural selection and molecular evolution in PTC, a bitter taste receptor gene. *Am J Hum Genet* 2004; **74**: 637–646.
31 Tiahkoff SA, Verrelli BC. Patterns of human genetic diversity: Implications for human evolutionary history and disease. *Annu Rev Genom Hum Genet* 2003; **4**: 293–340.
32 Bowcock AM, Kidd JR, Mountain JL *et al*. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci USA* 1991; **88**: 839–843.
33 Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002; **12**: 1805–1814.
34 Fallin D, Schork NJ. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet* 2000; **67**: 947–959.
35 Tishkoff SA, Pakstis AJ, Ruano G, Kidd KK. The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. *Am J Hum Genet* 2000; **67**: 518–522.
36 Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003; **73**: 1162–1169.
37 Bandelt H-J, Forster P, Sykes BC, Richards MB. Mitochondrial portraits of human populations using median networks. *Genetics* 1995; **141**: 743–753.
38 Sabeti PC, Reich DE, Higgins JM *et al*. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; **419**: 832–837.
39 Bamshad M, Wooding SP. Signatures of natural selection in the human genome. *Nat Rev Genet* 2003; **4**: 99–111.
40 Tishkoff SA, Varkonyi R, Cahinhinan N *et al*. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 2001; **293**: 455–462.
41 Verrelli BC, McDonald JH, Argyropoulos G *et al*. Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am J Hum Genet* 2002; **71**: 1112–1128.
42 Ding YC, Chi HC, Grady DL *et al*. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci USA* 2002; **99**: 309–314.
43 Gilad Y, Rosenberg S, Przeworski M, Lancet D, Skorecki K. Evidence for positive selection and population structure at the human MAO-A gene. *Proc Natl Acad Sci USA* 2002; **99**: 862–867.
44 Fullerton SM, Bartoszewicz A, Ybazeta G *et al*. Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *Am J Hum Genet* 2002; **70**: 1096–1106.
45 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.
46 Zhang J, Rowe WL, Clark AG, Buetow KH. Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. *Am J Hum Genet* 2003; **73**: 1073–1081.
47 Zhou G, Zhai Y, Dong X *et al*. Haplotype structure and evidence from positive selection at the human IL13 locus. *Mol Biol Evol* 2004; **21**: 29–35.
48 Rockman MV, Hahn MW, Soranzo N, Goldstein DB, Wray GA. Positive selection on a human-specific transcription factor binding site regulating IL4 expression. *Curr Biol* 2003; **13**: 2118–2123.
49 Rosenwasser LJ, Borish L. Genetics of atopy and asthma: the rationale behind promoter-based candidate gene studies (IL-4 and IL-10). *Am J Respir Crit Care Med* 1997; **156**: S142–S155.
50 Wooding SP, Watkins WS, Bamshad MJ, Dunn DM, Weiss RB, Jorde LB. DNA sequence variation in a 3.7-kb noncoding sequence 5′ of the CYP1A2 gene: Implications for human population history and natural selection. *Am J Hum Genet* 2002; **71**: 528–542.
51 Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 1975; **7**: 256–276.
52 Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 1979; **76**: 5269–5273.
53 Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**: 585–595.
54 Excoffier L. Human demographic history: refining the recent African origin model. *Curr Opin Genet Dev* 2002; **12**: 675–682.
55 Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 2003; **21**: 12–27.
56 Wall JD. Recombination and the power of statistical tests of neutrality. *Genet Res Camb* 1999; **74**: 65–79.
57 Weir BS, Cockerham CC. Estimating F statistics for the analysis of population structure. *Evolution* 1984; **38**: 1358–1370.
58 Na L, Stephens M. Modeling linkage disequilibrium, and identifying recombination hotspots using SNP data. *Genetics* 2003; **165**: 2213–2233.
59 Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 1964; **49**: 49–67.
60 Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet* 1968; **38**: 226–231.
61 Hubbard T, Barker D, Birney E *et al*. The Ensembl genome database project. *Nucleic Acids Res* 2002; **30**: 38–41.