Am. J. Hum. Genet. 51:592–608, 1992

HLA-DR Polymorphism in a Senegalese Mandenka Population: DNA Oligotyping and Population Genetics of DRB1 Specificities

Jean-Marie Tiercy,*'‡ Alicia Sanchez-Mazas,†'§ Laurent Excoffier,†'§ Xiaowen Shi-Isaac,* Michel Jeannet,* Bernard Mach,‡ and André Langaney†'§

*Transplantation Immunology Unit, Hôpital Cantonal, †Laboratory of Genetics and Biometry, Department of Anthropology and Ecology, and ‡Department of Genetics and Microbiology, University of Geneva, Geneva; and §URA 49, CNRS Museum, Paris

Summary

HLA class II loci are useful markers in human population genetics, because they are extremely variable and because new molecular techniques allow large-scale analysis of DNA allele frequencies. Direct DNA typing by hybridization with sequence-specific oligonucleotide probes (HLA oligotyping) after enzymatic in vitro PCR amplification detects HLA allelic polymorphisms for all class II loci. A detailed HLA-DR oligotyping analysis of 191 individuals from a geographically, culturally, and genetically well-defined western African population, the Mandenkalu, reveals a high degree of polymorphism, with at least 24 alleles and a heterozygosity level of .884 for the DRB1 locus. The allele DRB1*1304, defined by DNA sequencing of the DRB1 first-domain exon, is the most frequent allele (27.1%). It accounts for an unusually high DR13 frequency, which is nevertheless within the neutral frequency range. The next most frequent specificities are DR11, DR3, and DR8. Among DRB3-encoded alleles, DR52b (DRB3*02) represents as much as 80.7% of all DR52 haplotypes. A survey of HLA-DR specificities in populations from different continents shows a significant positive correlation between genetic and geographic differentiation patterns. A homozygosity test for selective neutrality of DR specificities is not significant for the Mandenka population but is rejected for 20 of 24 populations. Observed high heterozygosity levels in tested populations are compatible with an overdominant model with a small selective advantage for heterozygotes.

Introduction

Human major histocompatibility complex (MHC) polymorphism plays a key role in the control of the immune response, organ transplantation, and the susceptibility to autoimmune diseases. It is also an important tool in human population genetics, since numerous population surveys have used HLA serological data (Langaney 1979; Piazza et al. 1980; Nei and Roychoudhury 1982; Cavalli-Sforza et al. 1988; Sanchez-Mazas and Langaney 1988; Sokal 1988; Sanchez-Mazas 1990; Barbujani and Sokal 1991). During the past 10 years, molecular studies of HLA class II genes have revealed a high diversity at both DNA-polymorphism and gene-organization levels (Kappes and Strominger 1988; Mach et al. 1990; Trowsdale et al. 1991).

The HLA-DR region of the human MHC is centromeric to class I genes and is located between the HLA-DQ subregion and the HLA class III genes. It contains one monomorphic DRA locus and, depending on the haplotype, one or two expressed DRβ-chain loci (fig. 1). While all DR haplotypes possess a functional DRB1 gene, encoding the DR1–DR18 serological allospecificities, most haplotypes express a second DRβ-chain encoded by DRB3 for DR3, DR5, and DR6 haplotypes; by DRB4 for DR4, DR7, and DR9 haplotypes; and by DRB5 for DR2 haplotypes (Gorski et al. 1987; Kappes and Strominger 1988; Kawai et

Received October 23, 1991; final revision received May 4, 1992.

Address for correspondence and reprints: Dr. Jean-Marie Tiercy, Transplantation Immunology Unit, Lab 8252, Centre Médical Universitaire, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland.

0002-9297/92/5103-0017$02.00

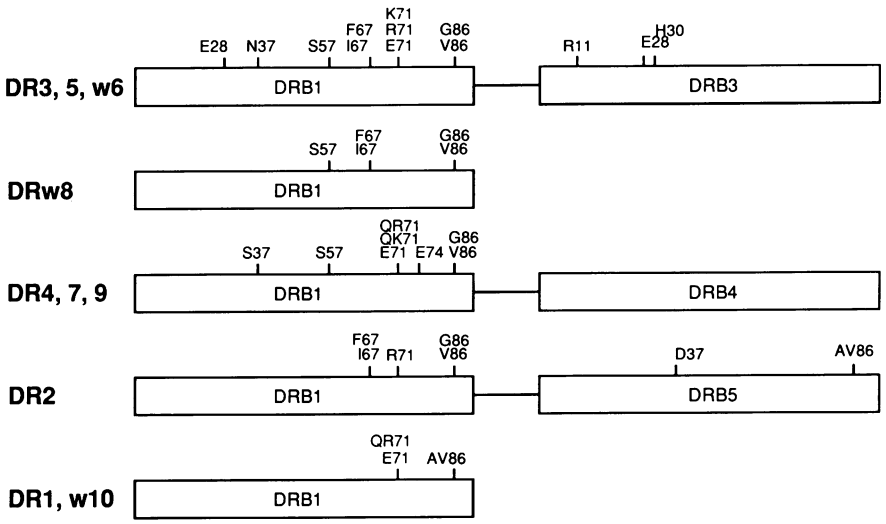

Figure 1 Schematic representation of the DRB loci within the different DR1-DR10 specificities. DRB3, DRB4, and DRB5 are not truly allelic (Gorski et al. 1987; Mach et al. 1990; Tiercy et al. 1990*a*). The position of the oligonucleotide probes used in this study for the analysis of DRB1, DRB3, and DRB5 allelic polymorphisms is indicated. Each probe is named according to the corresponding amino acid sequence (one-letter code) and the position on the DRB first-domain exons (aa 5–94).

al. 1989; Mach et al. 1990; Tiercy et al. 1990*a*). DRB3 and DRB4 encode, respectively, the DR52 and DR53 supertypic specificities. DRB1 is by far the most polymorphic HLA class II locus, with more than 40 alleles defined at the DNA sequence level (Bodmer et al. 1991). DRB3 and DRB5 are less polymorphic (four alleles), and DRB4 is apparently monomorphic. A large number of class II alleles are not detected by routine serological typing, essentially because of the lack of monospecific antisera and because of cross-reactions between closely related alleles. However, recent technical developments have provided tools to analyze HLA class II polymorphism at the DNA level by hybridization with sequence-specific oligonucleotide probes ("oligotyping") (Angelini et al. 1986), which have been applied, in the context of population studies, both to HLA-DR loci (Ronningen et al. 1990; Fernandez-Vina et al. 1991; Gao et al. 1991; Obata et al. 1991) and to HLA-DQ loci (Helmuth et al. 1990). Prior amplification of the DNA by PCR (Mullis and Faloona 1987) drastically simplifies this procedure (Tiercy et al. 1990*b;* Mach and Tiercy 1991). This novel HLA-oligotyping method therefore becomes suitable for large-scale epidemiological and anthropological analyses.

In this paper we present a detailed HLA-DR oligotyping analysis of 191 individuals from a geographically, culturally, and genetically well-defined population from western Africa, the Niokholo Mandenkalu (singular: Mandenka) from eastern Senegal. HLA-DR frequencies are estimated, with sample and population pedigree information being taken into account. The selective neutrality of the DR locus is tested using the Ewens-Watterson homozygosity test (Watterson 1978) for 24 populations. As an overdominant model has been invoked by several authors to account for both the large number of alleles and the high level of heterozygosity observed at the MHC loci (Doherty and Zinkernagel 1975; Black and Salzano 1981; Hedrick and Thomson 1983; Takahata and Nei 1990), homozygosity levels obtained under this alternative model are compared with observed values, to give estimates of heterozygotes selective advantage. HLA-DR frequency variation on a worldwide scale is studied and then interpreted in the context of a nonneutral model.

Material and Methods

Population Sampling

The Mandenka community belongs to the Mande linguistic group, which represents several hundred thousand individuals distributed over Senegal, Guinea, and Mali. The Niokholonke subgroup located in eastern Senegal, 30 km northeast of Kedougou, numbers today some 3,000 individuals. Although basically endogamous, it maintains a certain amount of gene flow

with other ethnic groups (mainly Bedik and Fulani) of the Niokholo region (Langaney and Gomila 1973; Lalouel and Langaney 1976; Blanc et al. 1990).

Twenty milliliters of peripheral blood were taken from each of 205 Mandenka volunteers from five different Niokholonke villages. Blood samples were collected in Vacutainer sterile tubes, and a half volume of RPMI medium supplemented with 10% FCS and heparin (1% final concentration) was added. Samples were kept at 15–20°C and were delivered to Geneva within 48 h. They were subjected to Ficoll gradients, and the granulocyte pellets were used for DNA extraction (Tiercy et al. 1989).

PCR and HLA-DR Oligotyping

Oligotyping for the major DR1–DR14 specificities (DR generic typing) was done as described elsewhere (Tiercy et al. 1990*b*, 1991*a*; Mach and Tiercy 1991), after both generic PCR amplification with primers DRBP1 (5′ CCGGATCCTTCGTGTCCCCACAGCACG 3′) and DRBP2 (5′ TCGCCGCTGCACTGTGAAG 3′) and hybridization with 14 oligonucleotide probes. In a second step, DR3, DR11, and DR13 subtypes were determined after DRB1 group-specific PCR, using primers DRBP5 (5′ GGAGTACTCTACGTCTGAG 3′) and DRBP2 (see above) and probes S57, R71, QR71, F67, I67, E71, G86, V86, and E28 (fig. 1) (Tiercy et al. 1989; Mach and Tiercy 1991). DR8 subtypes were determined after generic PCR using oligos S57, F67, I67, V86, and G86 (fig. 1) (Morel et al. 1990). DR4 subtypes were determined after group-specific PCR using primers DRBP4 (5′ GAGCAGGTTAAACATGAGTG 3′) and DRBP2 (see above) and probes S37, S57, E71, QK71, QR71, E74, V86, and G86 (fig. 1). DR2 allelic polymorphism at DRB1 and DRB5 loci was defined according to the description given by Tiercy et al. (1991*b*). DRB3 allelic polymorphism (DR52a/52b/52c) was determined after generic PCR using probes R11, E28, and H30 (fig. 1) (Tiercy et al. 1988). The sequences of all DR oligoprobes shown in figure 1 are given in table 1. The sequences of probes for DR generic typing have been given elsewhere (Mach and Tiercy 1991). DNA sequence of the first domain exons of DRB1 and DRB3 loci was obtained according to a method described by Tiercy et al. (1990*a*).

Estimation of Gene Frequencies in a Sample of Related Individuals

Pedigree information for more than 3,200 Mandenkalu from the Niokholo region was collected by a demographic team of "Musée de l'Homme" (Paris). The 191 individuals of our sample were not unrelated, and their genealogical relationships were established. The problem of estimating gene frequencies in a sample of related individuals has recently received much attention (Boehnke 1991). Several procedures exist to take family structures into account when gene frequency estimates are derived (Chakraborty 1991). We propose here to extend Chakraborty's (1978) methodology for arbitrary pedigree structure and size. This procedure amounts to attributing differential weights to individuals, depending on their tested relatives, before gene frequency estimations.

Let us first consider a family structure with a sibship of size s, where the untested parents are assumed to be unrelated. Note that the number of untested parents (p) may be greater than 2 in polygamous societies. We compute independently, for each untested parent, the expected number of genes (g) transmitted to its own progeny. These numbers (g_i; $i = 0, 1, 2, \ldots, p$) are random variables which can take values of 1 or 2 with probability $P(g_i = j), j = 1, 2$. As at least one gene of each parent is transmitted with a probability of 1, we are only interested in computing the probability of transmitting the second gene to the sibship. The probability that a second gene is not transmitted fol-

Table 1

Oligonucleotide Probes Used for Identification of DR1, DR2, DR3, DR4, DR8, DR11, and DR13 Subtypes

Oligonucleotide	Sequence (5′ to 3′)	Amino Acids
E71	GGCCCGCTCGTCTTCCAGG	68–73
K71	CGGCCCGCTTGTCTTCCAG	68–73
R71	GGCCCGCCTGTCTTCCAGG	68–73
F67	CTTCCAGGAAGTCCTTCTG	64–69
I67	CTTCCAGGATGTCCTTCTG	64–69
G86	GAAGCTCTCACCAACCCCG	85–89
V86	GAAGCTCTCCACAACCCCG	85–89
N37	AGCGCACGTTCTCCTCCTG	34–39
S57	GTACTCGGCGCTAGGCCGC	55–60
QK71	CGGCCCGCTTCTGCTCCAG	68–73
QR71	CCGCGGCCCGCCTCTGCTC	69–74
E74	GTGTCCACCTCGGCCCGCC	71–77
S37	AGCGCACGGACTCCTCTTG	34–39
QR71/2	CGGCCCGCCTCTGCTCCAG	68–73
AV86	GCTCACCACAGCCCCGTAG	83–88
D37	GAAGCGCAAGTCCTCCTCT	35–40
R71/2	CCGCGGCGCGCCTGTCTTC	69–74
R11	CTCAGACTTACGCAGCTCC	9–14
E28	GGAAGTATCTCTCCAGGAAC	26–31
H30	GGAAGTGTCTCTCCAGGAAC	26–31

lows a simple binomial distribution depending on the sibship size (s_i) of the considered parent and is given as $(\frac{1}{2})^{s_i-1}$. It follows that $P(g_i=1) = 2^{(1-s_i)}$ and $P(g_i=2) = 1-2^{(1-s_i)}$. Therefore, for each parent, the expected number of independent genes found among the sibship is

$$E(g_i) = 2-2^{(1-s_i)}. \quad (1)$$

Following Chakraborty (1978), we can attribute weight components (w_i) to each gene of the ith sibship, for individuals whose parents have not been tested, as

$$w_i = E(g_i)/s_i = 2\left(\frac{1-2^{-s_i}}{s_i}\right). \quad (2)$$

The computational weight ($\hat{w}$) of an individual and of each of his or her genes, the origin of which is unknown a priori, is taken as the mean of the paternal and maternal weight components, as both parents may have different sibships in the sample, a situation which is very common in polygamous societies. When this procedure is generalized to an arbitrary pedigree structure which has been incompletely tested, a weighting scheme for each individual is found to be

$$\hat{w} = \hat{w}_m\delta_m\left(\frac{1-2^{-s_m}}{s_m}\right)+\hat{w}_f\delta_f\left(\frac{1-2^{-s_f}}{s_f}\right), \quad (3)$$

with the m and f subscripts referring to the mother and the father, respectively, and with the δ's being indicator variables equal to 0 if a parental contributor has been tested, and equal to 1 if it has not. Note that the sibship size (s_m or s_f) refers here to the number of tested sibs plus the number of untested sibs who have transmitted genes to sampled individuals.

The computational weights of all 191 sampled individuals were iteratively found by scanning oldest generations first. We also computed the expected number of independent individuals (n) in the sample, as the sum of all individual weights.

Maximum-likelihood gene frequency estimates were computed with an EM algorithm procedure (Dempster et al. 1977) using weighted and unweighted gene counts. The χ^2 test for departure from Hardy-Weinberg equilibrium was first performed on all phenotypic classes. In case of significant χ^2, the test was carried out after clustering phenotypic classes with expected numbers that were fewer than five individuals into larger ones. The frequencies of HLA-DR specificities DR1–DR10 of 26 populations, including our sample, were used in a principal component analysis performed with the statistical package NTSYS (Rohlf 1990).

Genetic distances between arbitrary populations j and k were computed, from gene frequencies (p_i), as

$$d_{jk} = \frac{1}{2}\sum_i|p_{ij}-p_{ik}|, \quad (4)$$

(Powell et al. 1972) with the sum taken over all detected alleles. Geographic distances between populations were computed as arc-chord distances from latitude and longitude coordinates. When no precise location was available from the literature, coordinates of the closest city were taken. Coordinates of the capital city were used for samples designated by the name of the country. The correlation between geographic and genetic distance matrices, as well as its significance obtained by using a permutation procedure (Mantel 1967), were computed with the statistical package NTSYS (Rohlf 1990).

Expected Homozygosity under Neutrality and Overdominance

The Ewens-Watterson test (Watterson 1978) for selective neutrality was carried out for 24 samples of figure 3, using both the simulation procedure described by Fuerst et al. (1977) and the algorithm given by Excoffier (1988). Neutral expectations of allelic frequency distributions were obtained by simulating 1,000 random samples with fixed sizes and given numbers of alleles and drawn from populations at equilibrium. The F-statistics (here equivalent to the sample homozygosity in a random mating population, obtained as $\Sigma_i p_i^2$) were computed for each simulated sample. The lower 5% limit of the F neutral distribution ($F_{5\%}$), which approximates the lower limit of a unilateral 95% confidence interval around F, was determined.

The expected sample homozygosity under overdominance has been derived by Watterson (1978). Under this selection model, all heterozygotes are supposed to have a selective advantage s over the homozygotes, which leads, in the long term, to both an increase in the number of alleles maintained in the population (k) and a decrease in the homozygosity (F), compared with neutral expectations. Under overdominance, Watterson (1978) derived the expected number of alleles in a sample of r genes as

$$E(k) = \sum_{i=0}^{r-1} \frac{\theta}{\theta + i} + \frac{\alpha\theta r(r-1)}{(r+1)^2(r+\theta)(r+1+\theta)} + O(\alpha^2) \quad (5)$$

and derived the sample expected homozygosity as

$$E(F|k) = \frac{1}{1+\theta} - \frac{2\alpha r^2\theta}{(3+\theta)(2+\theta)(1+\theta)^2(r+\theta)(r+1+\theta)} + O(\alpha^2), \quad (6)$$

where $\alpha = 2Ns$, $\theta = 4Nu$, N is the population size, and u is the mutation rate.

For a given value of α, the observed number of alleles (k_{obs}) in the sample, taken as an estimate of $E(k)$, is used to extract θ from equation (5), with the term in α^2 being neglected because it does not outweigh the first two terms unless α is very large (Watterson 1977). This parameter is then substituted in equation (6), to give the expected homozygosity by using the same value of α as is used in equation (5). Equations (5) and (6) are computed for different values of α until $E(F|k)$ becomes equal to the observed homozygosity (F_{obs}).

Results

Mandenkalu Genetic Diversity

HLA-DR oligotyping analysis was performed using a two-step strategy (Mach and Tiercy 1991), as described in Material and Methods. Under the conditions used in this study, a total of 42 DRB1, 3 DRB3, and 3 DRB5 alleles can be discriminated. Throughout the oligotyping analysis, a majority of DR13-positive individuals showed a new oligotyping hybridization pattern (oligo E71-positive, N37-negative, S57-positive, and V86-positive). DRB first-domain exons of one of these individuals were amplified, subcloned into M13mp18, and sequenced. The DRB1 sequence is identical to the recently described DRB1*1304 allele (Lee et al. 1990*a*). The DRB3 sequence corresponds to the DRB3*02 allele.

The 82 HLA-DRB1 genotypes, occurrence counts, and total corresponding weights, with pedigree data considered, are shown in table 2. Some genotypes present untyped DRB1 specificities (08**, 0802/08**, 11**/13**, 1101/13**, 13**/08** and 1304/08**) which may be similar to already detected genotypes, so that at least 76 different genotypes may be defined. Note that the occurrence of 7 heterozygote genotypes (0102/0802, 0301/1301, 0801/1302, 1101/1303, 1102/1601, and 1303/1601) is only due to the segregation of genes from one generation to the next, and thus they are not used when gene frequencies are estimated from weighted gene counts. HLA-DR gene frequencies have been estimated independently using weighted and unweighted gene counts, for

Table 2

HLA-DRB1 Genotype Frequencies

Genotype (*n*)	Total wt	Genotype (*n*)	Total wt
0101 (1)	1.0000	1101 (4)	3.4714
0102/07[a] (1)	.8750	1101/07[a] (1)	.8750
0102/0802 (1)	.0000	1101/0901 (1)	.5000
0301 (4)	2.0000	1101/1102 (2)	1.3333
0301/0405 (1)	.6035	1101/13[a] (4)	2.6667
0301/07[a] (1)	.3333	1101/1301 (1)	.3750
0301/08[a]	.2917	1101/1303 (1)	.0000
0301/0802 (1)	.7500	1101/1304 (8)	4.9193
0301/1001 (2)	1.7500	1101/1601 (1)	.5000
0301/1101 (3)	1.5147	1102 (2)	1.0000
0301/1102 (3)	1.8708	1102/07[a] (2)	1.7500
0301/1301 (1)	.0000	1102/0901 (1)	1.0000
0301/1302 (3)	2.2188	1102/1104 (1)	.0000
0301/1304 (12)	7.7373	1102/1301 (2)	1.2198
0301/1601 (2)	1.1219	1102/1302 (1)	.8333
0302/0801 (1)	1.0000	1102/1304 (6)	3.1094
0302/1001 (1)	.2500	1102/1305 (1)	.3750
0302/1101 (2)	.7083	1102/1601 (1)	.0000
0302/1201 (1)	.8750	1103/08[a] (1)	.5833
0302/1301 (1)	.6667	1104/01[a] (1)	.8750
0302/1304 (5)	3.1989	1104/1301 (1)	.7917
0403/0801 (1)	1.0000	13[a]/08[a] (1)	.3750
0405/1101 (1)	.8750	1301/07[a] (2)	.6667
07[a]/1001 (2)	2.0000	1301/1001 (1)	.3750
08[a] (2)	1.3958	1301/1302 (4)	2.8333
0801 (1)	.5000	1301/1304 (4)	2.7500
0801/1101 (1)	1.0000	1302/1303 (1)	1.0000
0801/1302 (1)	.0000	1302/1304 (3)	1.6667
0801/1304 (1)	.8750	1302/1601 (2)	.0000
0802 (3)	1.9063	1303/07[a] (2)	1.3750
0802/07[a] (4)	2.1719	1303/1302 (1)	.5000
0802/0901 (1)	1.0000	1303/1304 (1)	.4626
0802/1101 (3)	1.9844	1303/1601 (1)	.0000
0802/1102 (1)	.3750	1304 (15)	9.4979
0802/1302 (2)	1.1250	1304/07[a] (9)	6.6094
0802/1601 (2)	1.2188	1304/08[a] (8)	5.6354
0803/1102 (1)	.6667	1304/1001 (7)	4.5417
0803/13[a] (1)	.6719	1304/1305 (2)	.8750
0803/1304 (1)	.7500	1304/1601 (2)	1.0000
1001 (1)	1.0000	1401/07[a] (2)	1.1250
11[a]/13[a] (1)	.5833	1401/0901 (1)	.7500

[a] Has not been subtyped for HLA-DRB1 alleles.

HLA-DR broad specificities (equivalent to serological DR1–DR14 specificities), HLA-DRB1, and HLA-DRB3 alleles (table 3). A few DRB1 specificities were not subtyped in some individuals. This may explain the small discrepancies found between the sum of subtype frequencies within each allospecificity and that of the specificity itself. Weighted gene frequencies are generally very close to unweighted gene frequencies, as previous results suggest (Chakraborty 1978; Boehnke 1991). Note, however, that the HLA-DR2 specificity (allele DRB1*1601) is observed in three zero-weight heterozygote genotypes which are discarded when weighted gene frequencies are computed. This results in a much lower weighted frequency estimate. When pedigree information is taken into account, the effective sample size (118.1 effectively independent individuals) represents only 68.1% of the original sample size (191 related individuals). The main effect observed when there is correction for identity by descent is thus the increase in the variance of gene frequencies—$V(p_i) = p_i(1-p_i)/2n$—by a factor of 1.62 (191/118.1). This may be of considerable importance when differences are tested with the frequencies observed in other samples. The χ^2 tests for goodness of fit between observed and expected phenotypic frequencies for HLA-DR1–DR14 specificities accept the hypothesis of Hardy-Weinberg equilibrium (table 3). The χ^2 test was not carried out for HLA-DR alleles because some specificities were not subtyped in several individuals (table 2).

Heterozygosity levels are .793 and .884, respectively, for HLA-DR specificities and DRB1 alleles. The allospecificity DR6 (DR13 + DR14) reaches a very high frequency (38.3%) compared with those in other sub-Saharan African populations, where it is commonly closer to 20%, as determined by serology, but it is lower than that in the Gambian sample studied by Hill et al. (1992) (pooled DR13 and DR14 frequencies 48.8%). In the oligotyped Norwegian group, it amounts to 16% (Ronningen et al. 1990), and it is as low as 8.4% in a recently oligotyped Chinese sample (Gao et al. 1991). The DRB1*1304 allele identified by DNA sequencing and oligotyping is the major contributor (27.1%) to this high DR6 frequency. This result is similar to the frequency found in Gambia (27.3%; Hill et al. 1992). We also note the prevalence of DR11, DR3, and DR8; low frequencies for DR1, DR2, DR4, and DR7; and an apparent absence of DR15-Dw2 (DRB1*1501–DRB5*0101), which constitutes a common haplotype in caucasoid and Oriental populations (Ronningen et al. 1990; Lee et al.

Table 3

HLA-DR Gene Frequencies

	HLA-DR Specificity Frequency[a]		HLA-DRB1 Allele Frequency[b]		
	Unweighted ($n=191$)	Weighted ($n=118.1$)		Unweighted ($n=191$)	Weighted ($n=118.1$)
1.....	.011	.012	0101	.004	.006
			0102	.007	.006
2.....	.029	.016	1601	.029	.016
3.....	.125	.119	0301	.095	.090
			0302	.029	.028
4.....	.008	.010	0403	.003	.004
			0405	.005	.006
7.....	.068	.075	07*	.068	.075
8.....	.115	.117	0801	.024	.028
			0802	.074	.072
			0803	.012	.013
9.....	.010	.014	0901	.010	.014
10.....	.038	.044	1001	.038	.043
11.....	.172	.169	1101	.094	.097
			1102	.066	.060
			1103	.003	.002
			1104	.008	.007
12.....	.003	.004	1201	.003	.004
13.....	.385	.375	1301	.047	.043
			1302	.049	.045
			1303	.019	.015
			1304	.264	.271
			1305	.008	.005
14.....	.008	.008	1401	.008	.008
x.....	.029	.037	x	.034	.040
χ^{2c}...	94.24[d]	69.67[d]			
df....	65	65			
P.....	<.05	NS			
χ^{2c}...	7.79[e]				
df....	5				
P.....	NS				

	HLA-DRB3 Allele Frequency	
	Unweighted ($n=191$)	Weighted ($n=118.1$)
52a (0101).....	.081	.081
52b (0201/02).....	.568	.555
52c (0301).....	.055	.052
No DRB3	.296	.312

NOTE.—NS = not significant at the 5% level.

[a] Computed from frequencies of genotypes defined by a set of 14 oligonucleotide probes recognizing DR1–DR14 allospecificities.

[b] Computed from frequencies of genotypes defined by subtyping the major DR1–DR14 allospecificities with additional sequence-specific probes (see text). Frequencies of alleles not subtyped for DRB1 (i.e., 01**, 08**, 11**, and 13**) have been shared among subtyped alleles from the same allospecificity, in relative proportions.

[c] Computed from differences between expected and observed genotype frequencies listed in table 2.

[d] Computed on all phenotypic classes.

[e] Computed after clustering classes with expected numbers <5.

1990*b*; Gao et al. 1991). Among the DR11 specificities, the two prevalent subtypes are 1101 and 1102. DR3 haplotypes are approximately equally divided into 0301 (DR17) and 0302 (DR18). It is worth mentioning the presence of the DRB1*0405 allele, very rare in caucasoids or Africans but found mainly in Orientals, whereas the 0401 and 0404 alleles, more frequent in caucasoids, are not detected in Mandenkalu. DRB3*0201/02 (DR52b) presents a very high frequency (80.7%) of all DR52 haplotypes (table 3), because of its association with DR3 (0301), DR11 (all subtypes), and DR13 (some 1301 and all 1304), which are frequent in this population. In all DR2-positive individuals, the DRB1*1601 allele was associated with DRB5*02. In addition, DQB1 oligotyping analysis (Morel et al. 1990) revealed the presence of several DRB1-DQB1 associations, such as DRB1*1302-DQB1*0501, DRB1*0802-DQB1*0301, DRB1*1101-DQB1*0602, or DRB1*0901-DQB1*0201, which have not been found yet in caucasoids (data not shown). Some of these haplotypes have also been observed in other populations of African ancestry (Lee et al. 1990*a*; Hill et al. 1991, 1992). Altogether, at least 40 different DRB1-DRB3-DQB1 haplotypes are detected in our sample. MHC polymorphism of the Mandenka sample is thus characterized not only by its extremely high degree of HLA-DR polymorphism but also by additional diversity of DR-DQ combinations. Similar results were obtained for the Gambian sample analyzed by Hill et al. (1992). However, DRB1 allele frequencies were found to be significantly heterogeneous between the two samples ($P < .01$ for the χ^2 homogeneity test), mostly because of uneven DRB1*0301, DRB1*0802, and DRB1*1302 frequencies (although some alleles were not subtyped in the Gambian sample). Although the two populations are geographically close, the Gambian individuals were sampled among many different ethnic groups (Mandenkalu, Wolof, Jola, Fula, and several others) surrounding Banjul, which may account for these

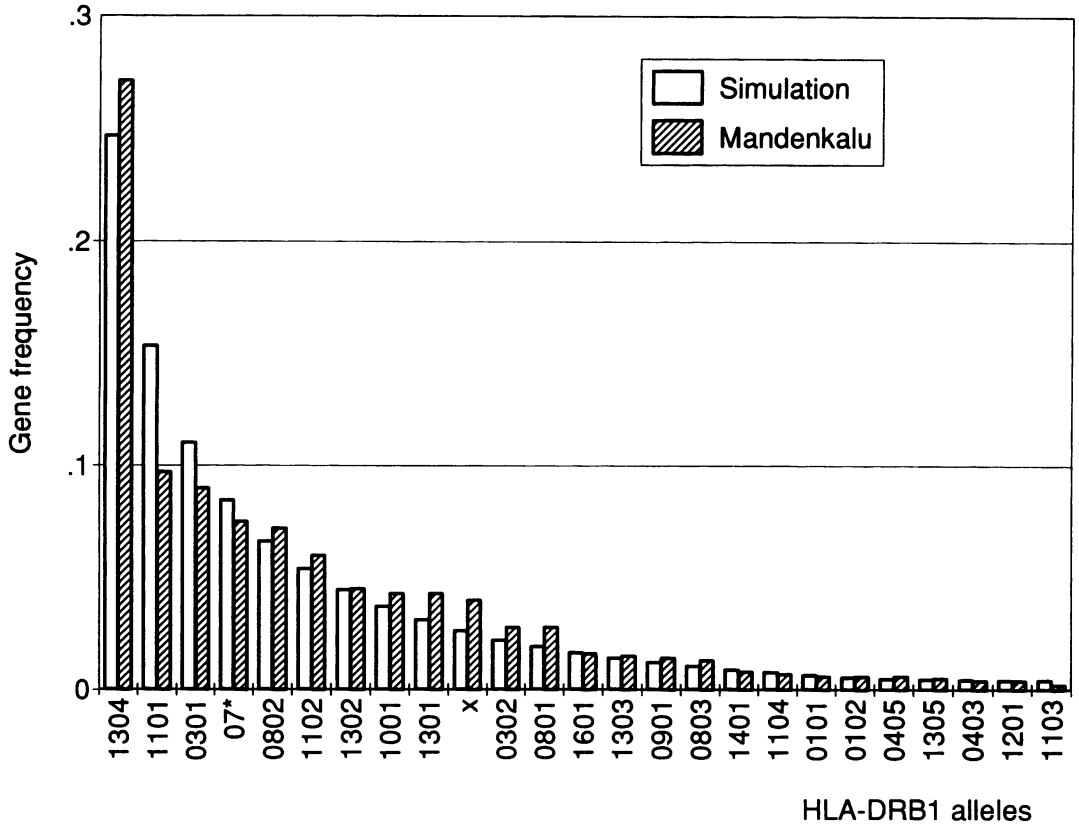

Figure 2 Mandenkalu HLA-DRB1 allele frequencies. Expected values are obtained by simulation of random neutral samples of 25 alleles in 118 diploid individuals (see text).

differences. In particular, the Fula are supposed to be of eastern African ancestry and, for other genetic systems, diverge from western Africans (Excoffier et al. 1987).

HLA-DR Neutrality Tests

Simulation of neutral random samples of populations at equilibrium (Fuerst et al. 1977) provided estimates both for homozygosity levels and for gene frequencies. Simulated mean HLA-DRB1 frequencies are very close to observed values (fig. 2) for an effective Mandenkalu sample size of 118 independent individuals. A goodness-of-fit *G*-test (Sokal and Rohlf 1981, p. 698) between observed and expected frequencies produced nonsignificant results ($G = 10.09$, df = 14, $P = .73$). To study the selective neutrality of the HLA-DRB1 locus, the homozygosity test has been performed for 24 populations worldwide (table 4), including the Mandenkalu. Observed homozygosities are found to be significantly different from neutral expectations for 20 populations, at a confidence level of 5% (fig. 3). In the Mandenkalu, observed homozygosity level computed for HLA-DRB1 alleles ($F_{obs} = .116$) falls within the neutrality range, confirming the results of the *G*-test. A homozygosity test based on DR allospecificities DR1–DR14 also accepts the neutrality hypothesis for the Mandenkalu ($F_{exp} = .56 > F_{obs} = .207 > F_{5\%} = .147$), showing that the apparent neutrality obtained for this sample at the DR locus (see Discussion below) was not affected by subtyping. For each population where the selective neutrality (null hypothesis) of the HLA-DR could not be assumed, equations (5) and (6) have been solved for equilibrium mutation (θ) and selection (α) values (table 5). The mean α value over these 20 populations is 8.84 ± 1.52 (table 5). The expected homozygosity

Table 4

Population Samples Used in Multivariate or Simulation Analyses

Population/Abbreviation (sample size)	No. of DR Alleles[a]	Reference
Afro-Americans/AFA (323)	11	Baur and Danilovs 1980
Asaro (Papuan)/ASA (41)	5	Crane et al. 1985
English/ENG (200)	10	Papiha et al. 1985
Eskimo (Inuit)/ESK 1 (73)	7	Hansen et al. 1986
Eskimo (Inuit)/ESK 2 (46)	8	Hansen et al. 1986
French/FRE (2,718)	11	Cambon-Thomsen and Ohayon 1986
Greeks/GRE (189)	11	Pachoula-Papasteriadis et al. 1989
Han1/HAN1 (221)	13	Pei et al. 1985
Han2/HAN2 (127)	11	Lee et al. 1988
Han3/HAN3 (48)	9	Jaraquemada et al. 1984
Hindu1/HIN1 (134)	10	Mittal et al. 1982
Hindu2/HIN2 (138)	12	Mehra et al. 1986
Japanese/JAP (884)	11	Baur and Danilovs 1980
Koreans/KOR (155)	11	Lee et al. 1988
Malians/MAL (49)	11	Kalidi et al. 1988
Mandenkalu/MAN (118)	25	Present study
Martinicans/MAR (128)	10	Monplaisir et al. 1985
Nigerians/NIG (130)	10	Okoye et al. 1989
Norwegians/NOR (181)	17	Ronningen et al. 1990
San/SAN (212)	10	Du Toit et al. 1990*b*
Saudi Arabs/SAU (100)	9	Ollier et al. 1985
Tunisians/TUN (109)	11	Ayed et al. 1987
American caucasoids/AME (2,569)	14	Lee et al. 1990[b]
Watut/WAT (77)	4	Crane et al. 1985
Xhosa/XHO (322)	11	Du Toit et al. 1987
Zimbabweans/ZIM (104)	11	Du Toit et al. 1990*a*

[a] This value stands for the number of actually detected alleles, plus the blank if the sum of detected frequencies is <1. Except the Mandenka and the Norwegian samples, which have been oligotyped, all samples have been studied by serology.

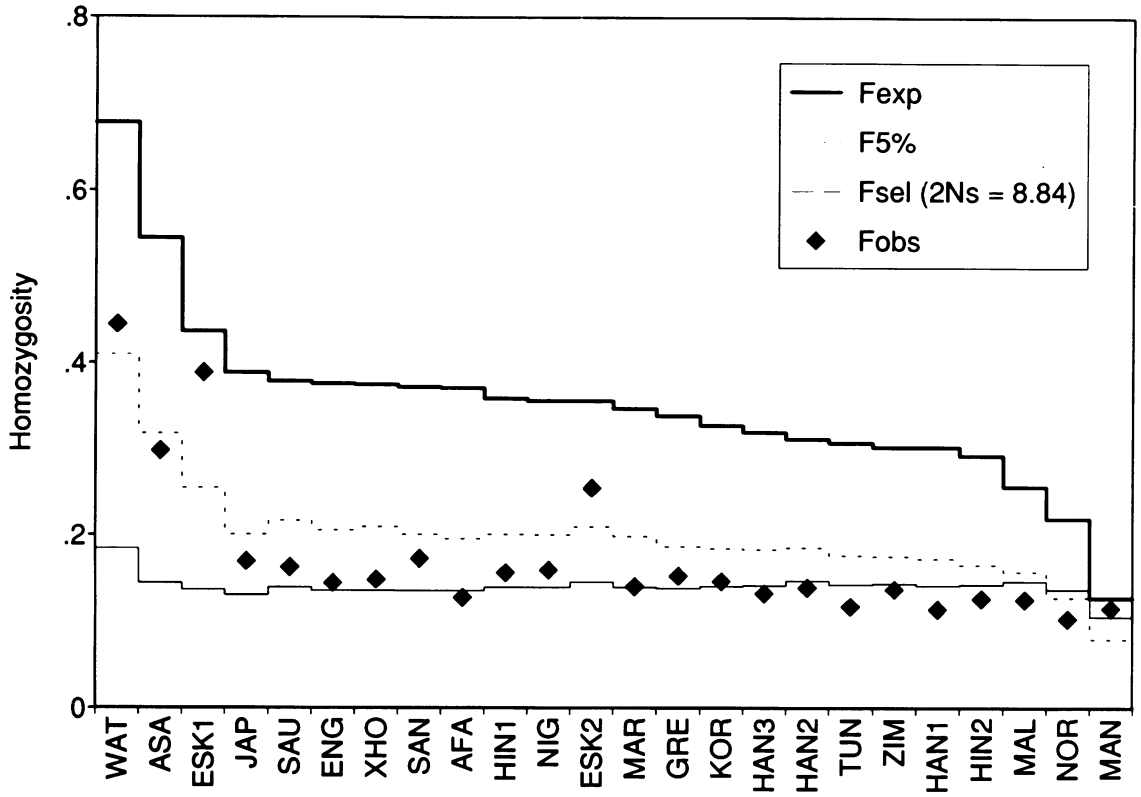

Figure 3 Observed (F_{obs}) and expected (F_{exp}) homozygosity values for 24 populations (referenced in table 4), as well as the limit of a 5%-level confidence interval ($F_{5\%}$). The populations are sorted according to decreasing expected F values. The maximum number of alleles detected by serology or oligotyping was used for each sample. The expected homozygosity under overdominance (F_{sel}) for an α value of 8.84 is also reported.

level corresponding to an α value of 8.84 has then been computed for all populations and is reported in figure 3. An approximate value for the selective advantage s of heterozygotes may be extracted from $\alpha = 2N_e s$, if an estimation for the effective population size is available. When an estimated value of 10^4 is used for human effective population size (Nei and Graur 1984), s can be as small as $(4.42 \pm 0.76) \times 10^{-4}$. Note, however, that this estimate heavily depends on real effective population size, which may vary among populations. Therefore the selective pressure against heterozygotes could also vary from population to population.

HLA-DR Variation Worldwide

The frequency vectors of HLA-DR allospecificities DR1–DR10 are shown, in figure 4, for 26 populations (table 4). With the exception of our sample and the Norwegian group, all remaining 24 populations were analyzed by serology. Human populations show marked variations in their DR frequencies. The most relevant differences are found in Inuit (Eskimo) and Papuan (Asaro and Watut) populations, where DR4 for the former and both DR2 and DR6 for the latter show very high frequencies compared with those in other populations in the world. Note also that a number of DR alleles (DR1, DR3, DR5, DR7, DR9, and DR10) either are not observed or show very low frequencies in these two population groups, contrasting with highly polymorphic distributions observed in other parts of the world. Marked frequency differences are found between Oriental and Occidental groups, since higher DR9 frequencies and lower DR1 frequencies are observed in eastern Asiatics than are observed in either sub-Saharan Africans or caucasoids. These latter two groups diverge mainly by low DR1, DR4, and DR7 in sub-Saharan Africans, associated with a higher DR6 frequency. Among Africans, the Mandenkalu present some singularities, e.g., very low DR1, DR2, and DR4 and very high DR6 frequencies, and the San popula-

Table 5

θ and α Parameters Explaining the Significant Excess of Heterozygotes under an Overdominant Model in 20 Populations for Which Neutrality of DRB1 Locus was Rejected

Population	θ	α	α/θ[a]
Afro-Americans	1.77	9.15	5.17
Asaro (Papuan)	1.00	5.02	5.02
English	1.74	8.51	4.89
Greeks	1.99	8.22	4.13
Han 1	2.39	10.37	4.34
Han 2	2.20	9.32	4.24
Han 3	2.24	9.29	4.15
Hindu 1	1.91	8.09	4.24
Hindu 2	2.42	9.77	4.04
Japanese	1.47	7.57	5.15
Koreans	2.09	8.54	4.09
Malians	2.97	10.63	3.58
Martinicans	1.93	8.80	4.56
Nigerians	1.93	7.96	4.12
Norwegians	3.56	12.55	3.52
San	1.71	7.43	4.35
Saudi Arabs	1.79	7.89	4.41
Tunisians	2.29	10.23	4.47
Xhosa	1.77	8.32	4.70
Zimbabweans	2.32	9.16	3.95

[a] Independent of the effective population size (N_e) and equal to $s/2u$.

tion presents some frequency patterns close to those of other sub-Saharan Africans and to those of caucasoids.

The two first principal planes of a principal component analysis for 26 populations, representing 67% of all genetic information, are shown in figure 5. Populations are grouped into clusters according to their geographical location. Sub-Saharan Africans and caucasoid populations overlap in the projection of the two first principal axes (fig. 5, *top*), whereas eastern Orientals (together with Eskimos) and Papuans are well differentiated from each other and from the former groups. The third axis (fig. 5, *bottom*) contrasts sub-Saharan Africans (except the San) with caucasoids, while southeastern Asiatics are closer to Papuans, and Eskimos are closer to northeastern Asiatics. The geographic groups are thus well differentiated on the two principal planes. There is indeed a significant correlation ($r = .377$; $P < .001$) between geographic and genetic distance matrices.

Discussion

A High Level of Polymorphism in the Mandenkalu

We have presented in this paper an extensive HLA-DRB1, -DRB3, and -DRB5 oligotyping analysis of a

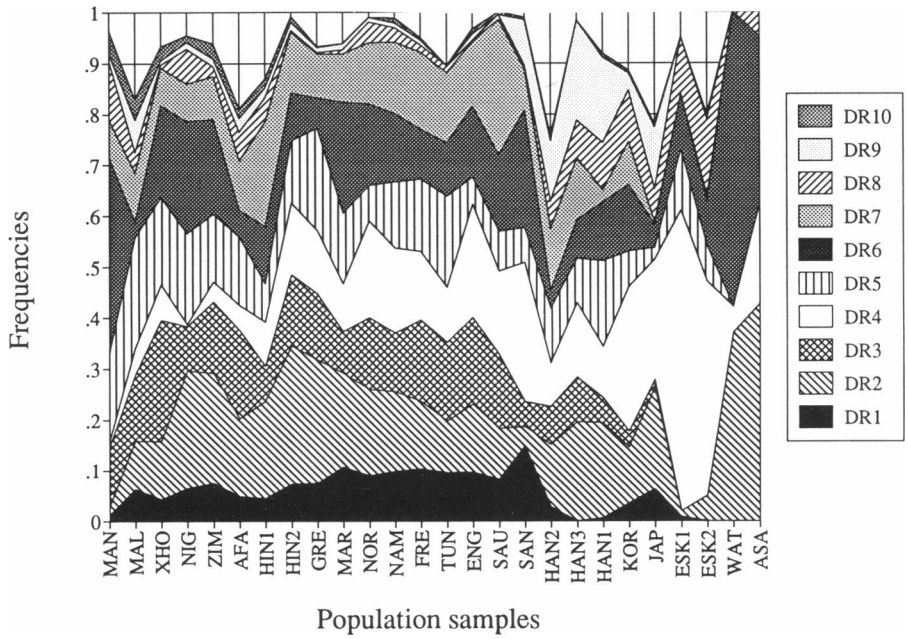

Figure 4 DR1–DR10 allospecificities frequency vectors (*vertical axis*) for 26 worldwide population samples (*horizontal axis*) referenced in table 4. Populations are ordered according to the results of the principal component analysis, and they cluster according to the following ethnic subdivisions: MAN, MAL, XHO, NIG, ZIM, and AFA (all sub-Saharan Africans); HIN1, HIN2, GRE, NOR, NAM, FRE, TUN, ENG, and SAU (all caucasoids); HAN2, HAN3, HAN1, KOR, and JAP (all east Orientals); ESK1 and ESK2 (both aboriginal Americans [Eskimo Inuit from Alaska]); and WAT and ASA (both Oceanians [Papuans from Melanesia]).

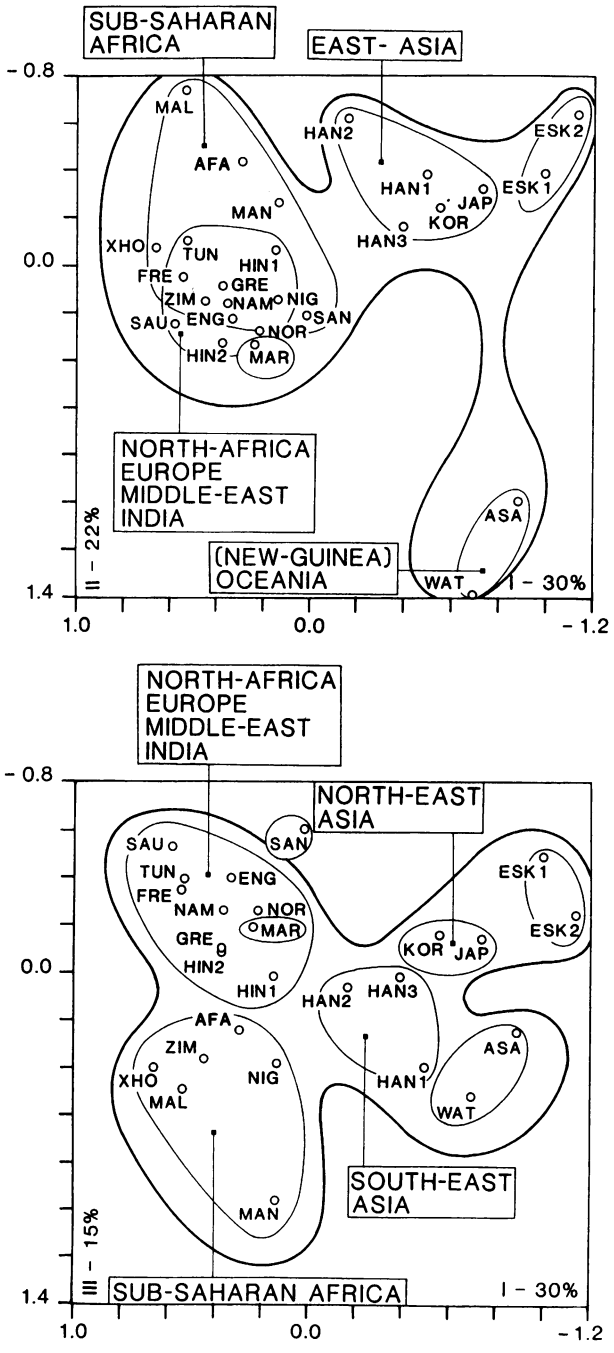

Figure 5 Principal component analysis for 26 population samples (referenced in table 4) tested for HLA-DR broad specificities (DR1, DR2, DR3, DR4, DR5, DR6, DR7, DR8, DR9, and DR10). *Top,* Principal axes I and II. *Bottom,* Principal axes I and III. Populations are clustered according to their geographical locations. The largest correlation coefficients between principal axes and DR allospecificities are as follows: axis I—DR3 (r = .90), DR7 (r = .71), DR1 (r = .67), DR8 (r = −.66), DR4 (r = −.60), and DR5 (r = .58); axis II—DR6 (r = .74), DR2 (r = .71), and DRx (r = −.72); and axis III—DR4 (r = −.64), DR1 (r = −.54), DR10 (r = .65), and DR6 (r = .41).

Senegalese Mandenka sample. The oligotyping performed with 14 probes allowed us to detect the presence of all DR1–DR14 allospecificities at very contrasted frequencies (table 3). Subtyping with additional oligoprobes (table 1), chosen on the basis of present DR specificities, detected a total of 24 DRB1 alleles, 3 DRB3 alleles, and 1 DRB5 allele (table 3 and fig. 2). When compared with those for other samples studied by oligotyping, this number of DRB1 alleles was close to that found in a Chinese sample (26 alleles) (Gao et al. 1991) but was larger than that found in a Norwegian sample (17 alleles) (Ronningen et al. 1990). There was no evidence for putative blank alleles (i.e., PCR-amplified DNA which failed to hybridize to the probes). We cannot, however, exclude the presence of additional micropolymorphism which would have escaped detection by our probes, although the oligonucleotides used here correspond to the most polymorphic amino acid positions within the three hypervariable regions of DRB1 (fig. 1). The estimated blank frequency (4%) is lower than published serological data (5%–24%; fig. 4) for populations of African ancestry. In view of the high diversity, among African populations, in conventional or DNA genetic markers (Wainscoat et al. 1986; Excoffier et al. 1987, 1991; Vigilant et al. 1991), oligotyping or direct gene sequencing of the HLA-DR locus in other African samples is likely to reveal additional polymorphisms still undetected in western African or European populations. It follows that interethnic differences in the distribution of MHC allelic subtypes would certainly increase the difficulty of proper donor-recipient matching for organ transplantation in pluriethnic communities.

Computation of individual weights allows one to get rid of redundant genetic information in samples of related individuals. It provides consistent gene frequency estimates generally not very different from those computed from unweighted gene counts (Chakraborty 1978; Boehnke 1991), although exceptions may occur if rare alleles are passed to the sibship (e.g., DR2; table 2). An interesting outcome of the weighting procedure is the determination of an effective sample size which can be used in the simulation of neutral random samples (fig. 2) and which can lead to more appropriate estimates of gene frequency variances. The limitations of this methodology lie in the accuracy of pedigree information: genealogies may be incomplete and may cover only a few generations.

The DRB1 alleles encoding the serologic specificities DR3, DR8, DR11, and DR13, all belonging to the supertypic group DR52, were the most frequent

(78%) (table 3). In the available oligotyped samples, the total frequency of these specificities is 67.8% in Gambia (Hill et al. 1992), 36% in Norway (Ronningen et al. 1990), and 16.7% in China (Gao et al. 1991). DRB1*0302 (DR18), which is virtually absent in caucasoids, is common in sub-Saharan Africans. DNA sequencing allowed us to identify the recently recognized DRB1*1304 allele, which is not only the most frequent DR13 subtype (71.5% of DR13) but also the most frequent DRB1 allele in this population (27.1%). Oligotyping of more than 600 Europeans from our laboratory did not reveal the presence of this allele, suggesting either its absence or its very low frequency in caucasoids. That DRB1*1304 was first sequenced from an American of African ancestry (Lee et al. 1990*a*) and then was found as a common allele in the Gambia (Hill et al. 1991, 1992) suggests that it may be a common DR13 subtype in many sub-Saharan populations. Difficulties in typing DR13 and its subtypes, essentially because of the lack of monospecific alloantisera, could also account for its nondetection and apparently lower frequency in serologically typed African samples. An alternative explanation for the high frequency of DRB1*1304 in Mandenkalu is that its frequency could have been raised by a pathogen-driven selection mechanism associated with disease protection, as proposed by Hill et al. (1991, 1992). Simulation of neutral random samples shows, however, that its observed frequency (27.1%) is very close to the expected value for the most common allele (24.7%) (fig. 2). The occurrence of a double-selection mechanism acting in the Mandenka population cannot be completely ruled out. A balancing-selection mechanism such as an advantage for heterozygotes (see below), combined with some form of directional selection such as pathogen-driven selection (for DRB1*1304), could mimic the effect of neutral evolution, as also suggested by Hill et al. (1992) for the Gambian sample. A parasitological study is needed to further investigate this double-selection hypothesis in the Mandenkalu, for whom no precise data are available. The most parsimonious hypothesis remains, however, that random genetic drift is the strongest evolutionary force in the Mandenka population.

Evolution of the HLA-DR Genetic System: the Influence of Overdominant Selection

The selective neutrality of the DR locus has been studied not only for the Mandenkalu, the data for whom would have been misleading, but for 23 other populations worldwide. The homozygosities and therefore the allelic frequency distributions at the DRB1 locus are generally found to be significantly different from those predicted under neutrality, contrasting with the apparent Mandenkalu neutrality. The presence of additional blank alleles, most likely in serologically defined samples, could artificially reduce the expected homozygosity level and account for the rejection of the neutrality test. To investigate this effect, the Ewens-Watterson test was done with one additional blank allele introduced into all population samples. The presence of this extra blank allele lowers both the expected homozygosity and the limit of the 5% confidence interval, without changing the result of the test shown in figure 3, except for two populations (San and Watut) whose observed homozygosities fall within the lowered confidence interval (data not shown). The presence of additional blank alleles in the samples would reduce still further the expected homozygosity. However, in most populations, neutrality rejection is associated with the occurrence of a few alleles at high frequencies, which represents a frequency distribution pattern radically different from a neutral distribution (as shown in fig. 2) with a blank fraction made up of several alleles. The rejection of the neutral hypothesis confirms earlier results based on a smaller set of populations (Klitz et al. 1986). Similar observations had been reported for other HLA loci (Hedrick and Thomson 1983; Klitz et al. 1986), which were all compatible with some form of balancing selection. Other results, such as a higher rate of nonsynonymous nucleotide substitution in the antigen-recognition site of HLA class I and class II molecules (Hughes and Nei 1988, 1989; Hedrick et al. 1991*b*) and transspecific polymorphisms (Takahata and Nei 1990; Klein 1991), are well explained by a symmetric overdominant selection model. The magnitude of the heterozygote advantage(s) inferred from α values depends on human effective population size, here taken as 10^4 (Nei and Graur 1984), which could be larger (10^5) or smaller (10^3) in some particular populations. Note that fluctuations in effective population sizes may explain the neutral behavior of some populations. Populations with a small effective population size, submitted to the same selection pressures as larger populations, would present frequency distributions compatible with a neutral hypothesis as $\alpha = 2Ns$ would get closer to 1, the limit of effective selection for a particular allele (Kimura 1968). This is a likely explanation accounting for the apparent neutrality of the two Eskimo populations, the Papuan Watut population, and the Mandenka population. Note also that the neutrality test

is known to be conservative (Ewens 1979): selected samples may thus be found within the neutrality range (type II error), which may also account for the apparent Mandenka neutrality. The α-value fluctuations among populations (table 5) may be due either to variation in N_e or to variation in s within each particular population, both parameters being likely to vary among populations. It is thus difficult to compute a confidence interval around s unless more-accurate effective population sizes are known. Note, however, that the fraction α/θ (equal to $s/2u$ and therefore independent from N_e) shown in table 5 is quite similar in most populations, which would argue in favor of comparable selective pressures in human populations. However, both the exact value for the heterozygote advantage and its variability among populations remain open to further investigation.

Alternative models of selection, such as viability selection, frequency-dependent selection, maternal/fetal interactions, and nonrandom mating, have also been proposed to account for the maintenance of a large number of alleles in the MHC system (reviewed in Hedrick et al. 1991*a;* Nei and Hughes 1991), and more than one selection mode could act on the MHC loci. In its effects, the overdominant model studied here bears many resemblances to some alternative models of selection, making it difficult to determine which is the selection mode at work in human populations. Frequency-dependent selection in the form of minority advantage has mathematical similarities with an overdominant model (Takahata and Nei 1990) and has similar effects on the population allele frequency distribution. In the overdominant model, low-frequency alleles are also at an advantage, as they only appear in heterozygote individuals, compared with common alleles, which appear in homozygotes as well. Viability selection under the form of pathogen-driven selection has been proposed as a major mechanism for the maintenance of HLA diversity in humans, with the demonstration of an association between reduced susceptibility to severe malaria and two combinations of HLA antigens (Hill et al. 1991). The HLA class II allele shown to be involved with this protection in a Gambian sample is DRB1*1302 (Hill et al. 1991). It is found at a low frequency of 4.5% in the total Mandenka population (compared with 16.7% in a Gambian control and with 8.7% in the severely anaemic patients suffering from malaria [Hill et al. 1991]), which leads one to question its protective role against endemic malaria in the Niokholo region. To maintain allelic diversity, viability selection should be effective for many different pathogens. In this case, selected heterozygotes would be able to efficiently present a larger number of pathogens than would homozygotes and could also be at an advantage in the population. As different pathogens are present in different regions of the world, a portion of the correlation between geography and genetics could be due to local adaptations to certain pathogens. More data are needed to establish a firm correlation between pathogen distribution and HLA diversity. If such coevolution needs a long time to have an impact on HLA differentiation (Klein 1991), then its effect would have been relatively weak during the recent evolution of modern humans.

Divergence of Human Populations, for the HLA-DR Locus: the Influence of Genetic Drift

Several lines of evidence suggest that, for the HLA-DR locus, most of the differences among human populations have evolved through both genetic drift and isolation by distance (Malécot 1948; Morton et al. 1972). The continental groups are indeed well separated on the principal plane of the multivariate analysis shown in figure 5. The populations are arranged approximately according to their geographic locations, not according to their neutral or selective behavior inferred from figure 3. The neutral Mandenka population falls within the African cluster. The neutral Inuit (Eskimo) populations, although living in a drastically different environment, are genetically close to northeastern Asiatic populations (Japanese and Korean), reflecting their recent common origin. We also observe a significant positive correlation between geographic and genetic distances and relatively smooth frequency changes between continental groups (fig. 4). Only Papuan (Asaro and Watut) and Inuit (Eskimo) populations exhibit sharp frequency changes for a few specificities, as well as a smaller number of detected allospecificities. Such features are commonly seen in populations having passed through a bottleneck reducing the effective population size, which would explain their apparent neutral homozygosity level (fig. 3), as discussed above. The observed correlation between geography and genetics ($r = .377$) is consistent with estimates obtained for human populations at a continental scale (Barbujani 1991). However, this correlation is likely to be an underestimate, as arc-chord distances among localities are only crude estimates of real migration routes. Its high significance indicates that, for the DRB1 locus, part of the differentiation process of human populations is due to isolation by distance.

A selection mechanism enhancing diversity at the HLA-DR locus is compatible with the fact that human population differences have mainly arisen through genetic drift and historical events. An observed homozygosity level close to .15 in most populations (fig. 3), as well as a fairly constant $s/2u$ factor (table 5), support the view that selective pressures are fairly uniform. We would therefore conclude that the small selective advantage for heterozygotes has had little effect on the pattern of interpopulation variation of DRB1 gene frequencies. The recent settlement of Europe by early farmers is a documented example of an HLA differentiation process associated with historical migrations (Sokal et al. 1991). Under an isolation-by-distance model, local selective adaptations of the MHC may occur, but these adaptative responses are built on an already diversified gene pool, so that existing differences are enhanced and not created de novo.

Because of the known limitations of DR serology, particularly within the DR52 group, it is likely that the DR-typing results of the populations included in this study (table 4) may need to be reevaluated when HLA oligotyping data are available. However, the fact that population data are based on different techniques (serology and oligotyping) does not seem to introduce biases, as the two oligotyped samples (Mandenkalu and Norwegians) are genetically close to surrounding populations (figs. 4 and 5). It suggests that the typing of HLA-DR main specificities (DR1–DR10), on which the multivariate analysis (fig. 5) is based, is basically correct for most populations. A problem would emerge if oligotyping produced results radically different from those of conventional serology, instead of providing mere subspecificities of serologically defined broad specificities. The former case would introduce serious biases into multivariate analyses and neutrality tests, whereas the latter would not have major effects. A clearer picture of HLA-DR variability will emerge with the generalization of oligotyping techniques at the population level. The use of the DNA sequence information which is available for all HLA alleles known so far (Marsh and Bodmer 1991; Zemmour and Parham 1991) should give greater insight into the evolutionary and selection mechanisms at work in human populations.

Acknowledgments

We thank the Mandenka population for its cooperation; A. and S. Epelbouin for blood sampling; Dr. L. Perrin for virology tests; G. Pison for demographic data; B. Battistolo, P. Beffy, P. Dard, L. Graven, E. Martinez, A. Morrison, E. Poloni, and P. Roux-Chabbey for technical assistance; S. Aeschlimann for drawings; and D. Roessli for computer help. We also acknowledge the Senegalese authorites and Air Afrique for permission and assistance. We also thank anonymous reviewers for their comments and suggestions. This study was supported by INSERM (France) grant 900.814 and by FNRS (Switzerland) grants 31-27731.89 and 32–28784.90.

References

Angelini G, de Préval C, Gorski J, Mach B (1986) High resolution analysis of the human HLA-DR polymorphism by hybridization with sequence-specific oligonucleotide probes. Proc Natl Acad Sci USA 83:4489–4493

Ayed KH, Bardi R, Gebuhrer L, Gorgi Y, Bétuel H (1987) HLA-A,B,C and DR antigens in a sample of the Tunisian population. Tissue Antigens 29:225–231

Barbujani G (1991) What do languages tell us about human microevolution? Trends Ecol Evol 6:151–156

Barbujani G, Sokal RR (1991) Genetic population structure of Italy. I. Geographic patterns of gene frequencies. Hum Biol 63:253–272

Baur MP, Danilovs JA (1980) Population analysis of HLA-A,B,C,DR, and other genetic markers. In: Terasaki PI (ed) Histocompatibility testing 1980. UCLA Tissue Typing Laboratory, Los Angeles, pp 955–993

Black FL, Salzano FM (1981) Evidence for heterosis in the HLA system. Am J Hum Genet 33:894–899

Blanc M, Sanchez-Mazas A, Hubert van Blyenburgh N, Seven A, Pison G, Langaney A (1990) Interethnic genetic differentiation: GM polymorphism in eastern Senegal. Am J Hum Genet 46:383–392

Bodmer JG, Marsh SGE, Albert E, Bodmer WF, Dupont B, Erlich HA, Mach B, et al (1991) Nomenclature for factors of the HLA system, 1990. Tissue Antigens 37:97–104

Boehnke M (1991) Allele frequency estimation from data on relatives. Am J Hum Genet 48:22–25

Cambon-Thomsen A, Ohayon E (1986) Analyse des données génétiques sur l'échantillon global des provinces françaises. In: Ohayon E, Cambon-Thomsen A (eds) Colloque INSERM, vol 142: Génétique des populations humaines. INSERM, Paris, pp 297–318

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci USA 85:6002–6006

Chakraborty R (1978) Number of independent genes examined in family surveys and its effect on gene frequency estimation. Am J Hum Genet 30:550–552

——— (1991) Inclusion of data on relatives for estimation of allele frequencies. Am J Hum Genet 49:242–243

Crane G, Bhatia K, Honeyman M, Doran T, Messel N,

Hakos G, Tarlinton D, et al (1985) HLA studies of Highland and Coastal New Guineans. Hum Immunol 12:247–260

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38

Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256:50–52

Du Toit ED, Emmanuel JC, West G, Taljaard DG, Oudshoorn M (1990*a*) HLA frequencies in Black Zimbabweans. Tissue Antigens 36:122–124

Du Toit ED, MacGregor KJ, Taljaard DG, Oudshoorn M (1987) HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African Negroes, Cape Coloureds and South African Caucasoids. Tissue Antigens 31:109–125

Du Toit ED, Taljaard DG, Marshall J, Ritchie C, Oudshoorn M (1990*b*) The HLA genetic constitution of the Bushmen (San). Hum Immunol 28:406–415

Ewens WJ (1979) Testing the generalized neutrality hypothesis. Theor Popul Biol 15:205–216

Excoffier L (1988) Polymorphisme de l'ADN mitochondrial et histoire du peuplement humain. PhD thesis, University of Geneva, Geneva

Excoffier L, Harding RM, Sokal RR, Pellegrini B, Sanchez-Mazas A (1991) Spatial differentiation of RH and GM haplotype frequencies in sub-Saharan Africa and its relation to linguistic affinities. Hum Biol 63:273–307

Excoffier L, Pellegrini B, Sanchez-Mazas A, Simon C, Langaney A (1987) Genetics and history of sub-Saharan Africa. Yearbook Phys Anthropol 30:151–194

Fernandez-Vina MA, Gao X, Moraes ME, Moraes JR, Salatiel I, Miller S, Tsai J, et al (1991) Alleles at four HLA class II loci determined by oligonucleotide hybridization and their associations in five ethnic groups. Immunogenetics 43:299–312

Fuerst PA, Chakraborty R, Nei M (1977) Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. Genetics 86:455–483

Gao X, Sun Y, An J, Fernandez-Vina M, Qou J, Lin L, Stastny P (1991) DNA typing for HLA-DR, and -DP alleles in a Chinese population using the polymerase chain reaction (PCR) and oligonucleotide probes. Tissue Antigens 38:24–30

Gorski J, Rollini P, Mach B (1987) Structural comparison of the genes of two HLA-DR supertypic groups: the loci encoding DRw52 and DRw53 are not trully allelic. Immunogenetics 25:397–402

Hansen JA, Lanier AP, Nisperos B, Mickelson E, Dahlberg S (1986) The HLA system in Inupiat and Central Yupik Alaskan Eskimos. Hum Immunol 16:315–328

Hedrick PW, Klitz W, Robinson WP, Kuhner MK, Thomson G (1991*a*) Population genetics of HLA. In: Selander RK, Clark AG, Whittam TS (eds) Evolution at the molecular level. Sinauer, Sunderland, MA, pp 248–271

Hedrick PW, Thomson G (1983) Evidence for balancing selection at HLA. Genetics 104:449–456

Hedrick PW, Wittham TS, Parham P (1991*b*) Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. Proc Natl Acad Sci USA 88:5897–5901

Helmuth R, Fildes N, Blake E, Luce MC, Chimera J, Madej R, Gorodezky C, et al (1990) HLA-DQα allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. Am J Hum Genet 47:515–523

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennet S, et al (1991) Common West African HLA antigens are associated with protection from severe malaria. Nature 352:595–600

Hill AVS, Allsopp CEM, Kwiatkowski D, Taylor TE, Yates SNR, Anstey NM, Wirima JJ, et al (1992) Extensive genetic diversity in the HLA class II region of Africans, with a focally predominant allele, DRB1*1304. Proc Natl Acad Sci USA 89:2277–2281

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

——— (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Jaraquemada D, Alonso A, Awad J, d'Amaro J, Ollier W, Doyle P, Okoye RC, et al (1984) HLA polymorphisms in a Shanghai Chinese population. Tissue Antigens 23:23–32

Kalidi I, Fofana Y, Rahly AA, Bochu V, Dehay C, Gony J, Hors J (1988) Study of HLA antigens in a population of Mali (West Africa). Tissue Antigens 31:98–102

Kappes D, Strominger JL (1988) Human class II major histocompatibility complex genes and proteins. Annu Rev Biochem 57:991–1028

Kawai J, Ando A, Sato T, Nakatsuji T, Tsuji K, Inoko H (1989) Analysis of gene structure and antigen determinants of DR2 antigens using DR gene transfer into mouse L cells. J Immunol 142:312–317

Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet Res 11:247–269

Klein J (1991) Of HLA, Tryps, and selection: an essay on coevolution of MHC and parasites. Hum Immunol 30: 247–258

Klitz W, Thomson G, Baur MP (1986) Contrasting evolutionary histories among tightly linked HLA loci. Am J Hum Genet 39:340–349

Lalouel J-M, Langaney A (1976) Bedik and Niokholonko of Senegal: intervillage relationship inferred from migration data. Am J Phys Anthropol 45:453–466

Langaney A (1979) Diversité et histoire humaines. Population 34:985–1006

Langaney A, Gomila J (1973) Bedik and Niokholonko: intra- and inter-ethnic migration. Hum Biol 45:137–150

Lee KW, Johnson AH, Hurley CK (1990*a*) Two divergent

routes of evolution gave rise to the DRw13 haplotypes. J Immunol 145:3119–3125

Lee TD, Lee G, Zhao TM (1990*b*) HLA-DR, DQ antigens in North American Caucasians. Tissue Antigens 35:64–70

Lee TD, Zhao TM, Mickey R, Sun YP, Lee G, Song CX, Cheng DZ, et al (1988) The polymorphism of HLA in the Chinese. Tissue Antigens 32:188–208

Mach B, Berte C, Tiercy J-M, Reith W (1990) Diversity and regulation of MHC class II genes. In: Silver J (ed) Molecular biology of HLA class II antigens. CRC, Boca Raton, FL, pp 201–223

Mach B, Tiercy J-M (1991) Genotypic typing of HLA class II: from the bench to the bedside. Hum Immunol 30:278–284

Malécot G (1948) Les mathématiques de l'hérédité. Masson, Paris

Mantel G (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Marsh SGE, Bodmer JG (1991) HLA class II nucleotide sequences, 1991. Tissue Antigens 37:181–189

Mehra NK, Taneja V, Kailash S, Raizada N, Vaidya MC (1986) Distribution of HLA antigens in a sample of the North Indian Hindu population. Tissue Antigens 27:64–74

Mittal KK, Naik S, Sansonetti N, Cowherd R, Kumar R, Wong DM (1982) The HLA antigens in Indian Hindus. Tissue Antigens 20:223–226

Monplaisir N, Valette I, Lepage V, Dijon V, Lavocat E, Ribal C, Raffoux C (1985) Study of HLA antigens of the Martinican population. Tissue Antigens 26:1–11

Morel C, Zwahlen F, Jeannet M, Mach B, Tiercy J-M (1990) Complete analysis of HLA-DQB1 polymorphism and DR-DQ linkage disequilibrium by oligonucleotide typing. Hum Immunol 29:64–77

Morton NE, Yee S, Harris DE, Lew R (1972) Bioassay of kinship. Theor Popul Biol 2:507–524

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. Methods Enzymol 155:335–350

Nei M, Graur D (1984) Extent of protein polymorphism and the neutral mutation theory. Evol Biol 17:73–118

Nei M, Hughes AL (1991) Polymorphism and evolution of the major histocompatibility complex in mammals. In: Selander RK, Clark AG, Whittam TS (eds) Evolution at the molecular level. Sinauer, Sunderland, MA, pp 222–247

Nei M, Roychoudhury AK (1982) Genetic relationship and evolution of human races. Evol Biol 14:1–59

Obata F, Ito K, Kaneko T, Yang Y-G, Onda K, Ito I, Yabe N, et al (1991) HLA-DR gene frequencies in the Japanese population obtained by oligonucleotide genotyping. Tissue Antigens 38:124–132

Okoye RC, Ollier W, Jaraquemada D, Awad J, Navarrete C, Cutbush S, Carthy D, et al (1989) HLA-D region heterogeneity in a Nigerian population. Tissue Antigens 33: 445–456

Ollier W, Doyle P, Alonso A, Awad J, Williams E, Gill D, Welch S, et al (1985) HLA polymorphisms in Saudi Arabs. Tissue Antigens 25:87–95

Pachoula-Papasteriadis C, Ollier W, Cutbush S, Economodou J, Festenstein H (1989) HLA antigen and haplotype frequencies in Greeks. Tissue Antigens 33:488–490

Papiha SS, Duggan-Keen MF, Rodger RSE (1985) DR antigens and Bf allotypes in North-East England. Hum Hered 35:246–249

Pei J, Tokunaga K, Araki C, Yang CL, Sun SX, Mi XY, Saji H, et al (1985) HLA polymorphisms in a Han Chinese population of Sichuan. Tissue Antigens 26:323–331

Piazza A, Menozzi P, Cavalli-Sforza LL (1980) The HLA-A, B gene frequencies in the world: migration or selection? Hum Immunol 4:297–304

Powell JR, Levene H, Dobzhanski T (1972) Chromosomal polymorphism in *Drosophila pseudoobscura* used for diagnosis of geographic origin. Evolution 26:553–559

Rohlf FJ (1990) Numerical taxonomy and multivariate analysis system (NTSYS), ver 1.60. Exeter, Setauket

Ronningen KS, Spurkland A, Markussen G, Iwe T, Vartdal F, Thorsby E (1990) Distribution of HLA class II alleles among Norwegian Caucasians. Hum Immunol 29:275–281

Sanchez-Mazas A (1990) Polymorphisme des systèmes immunologiques Rhésus, GM et HLA et histoire du peuplement humain. PhD thesis, University of Geneva, Geneva

Sanchez-Mazas A, Langaney A (1988) Common genetic pools between human populations. Hum Genet 78:161–166

Sokal RR (1988) Genetic, geographic, and linguistic distances in Europe. Proc Natl Acad Sci USA 85:1722–1726

Sokal RR, Oden NL, Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. Science 351:143–145

Sokal RR, Rohlf FJ (1981) Biometry. Freeman, New York

Takahata N, Nei M (1990) Allelic geneaology under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124:967–978

Tiercy J-M, Gorski J, Bétuel H, Freidel AC, Gebuhrer L, Jeannet M, Mach B (1989) DNA typing of DRw6 subtypes: correlation with DRB1 and DRB3 allelic sequences by hybridization with oligonucleotide probes. Hum Immunol 24:1–14

Tiercy J-M,Gorski J, Jeannet M, Mach B (1988) Identification and distribution of three serologically undetected HLA-DR alleles by oligonucleotide. DNA typing analysis. Proc Natl Acad Sci USA 85:198–202

Tiercy J-M, Goumaz C, Mach B, Jeannet M (1991*a*) Application of HLA-DR oligotyping to 110 kidney transplant patients with doubtful serological typing. Transplantation 51:1110–1114

Tiercy J-M, Jeannet M, Mach B (1990*a*) A new approach for the analysis of HLA class II polymorphism: "HLA oligotyping." Blood Rev 4:9–15

——— (1990*b*) A new HLA-DRB1 allele within the DRw52

supertypic specificity (DRw13-DwHAG): sequencing and direct identification by oligonucleotide typing. Eur J Immunol 20:237–241

——— (1991*b*) Oligonucleotide typing analysis for the linkage disequilibrium between the polymorphic DRB1 and DRB5 loci in DR2 haplotypes. Tissue Antigens 37:161–164

Trowsdale J, Ragoussis J, Campbell RD (1991) Map of the humam MHC. Immunol Today 12:443–446

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, et al (1986) Evolutionary relationship of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491–493

Watterson GA (1977) Heterosis or neutrality? Genetics 85: 789–814

——— (1978) The homozygosity test of neutrality. Genetics 88:405–417

Zemmour J, Parham P (1991) HLA class I nucleotide sequences, 1991. Tissue Antigens 37:174–180