K. Cao
A.M. Moormann
K.E. Lyke
C. Masaberg
O.P. Sumba
O.K. Doumbo
D. Koech
A. Lancaster
M. Nelson
D. Meyer
R. Single
R.J. Hartzman
C.V. Plowe
J. Kazura
D.L. Mann
M.B. Sztein
G. Thomson
M.A. Fernández-Viña

Key words:
genetic distance; haplotype frequency; HLA class I genes; linkage disequilibrium; sequence-specific oligonucleotide probe

Acknowledgments:
Part of this work was performed at the National Histocompatibility Laboratory, American Red Cross at Baltimore, MD. AL, MN, PM, RS, and GT were supported by the National Institutes of Health (NIH) grant GM3J326 and AI49213. The research was partially funded by the NIH AI43906 'Immunologic Memory to pre-erythrocytic *Plasmodium falciparum*,' and we thank the Director of KEMRI for permission to publish this work; partial funding for this work was provided by the contract number N01-AI-85346 from the NIH; part of this work was supported by funding from the Office of Naval Research N00014-00-1-0898 to the C.W. Bill Young Marrow Donor Recruitment and Research Program. We thank Dr Steve Mack, Dr Carolyn Katovich Hurley, and Dr Jennifer Ng for their critical comments and input.

Received 19 July 2003, revised 9 October 2003, accepted for publication 23 October 2003

Tissue Antigens 2004: **63:** 293–325
Printed in Denmark. All rights reserved

Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci

Abstract: The allelic and haplotypic diversity of the HLA-A, HLA-B, and HLA-C loci was investigated in 852 subjects from five sub-Saharan populations from Kenya (Nandi and Luo), Mali (Dogon), Uganda, and Zambia. Distributions of genotypes at all loci and in all populations fit Hardy–Weinberg equilibrium expectations. There was not a single allele predominant at any of the loci in these populations, with the exception of A*3002 [allele frequency (AF) = 0.233] in Zambians and Cw*1601 (AF = 0.283) in Malians. This distribution was consistent with balancing selection for all class I loci in all populations, which was evidenced by the homozygosity *F* statistic that was less than that expected under neutrality. Only in the A locus in Zambians and the C locus in Malians, the AF distribution was very close to neutrality expectations. There were six instances in which there were significant deviations of allele distributions from neutrality in the direction of balancing selection. All allelic lineages from each of the class I loci were found in all the African populations. Several alleles of these loci have intermediate frequencies (AF = 0.020–0.150) and seem to appear only in the African populations. Most of these alleles are widely distributed in the African continent and their origin may predate the separation of linguistic groups. In contrast to native American and other populations, the African populations do not seem to show extensive allelic diversification within lineages, with the exception of the groups of alleles A*02, A*30, B*57, and B*58. The alleles of human leukocyte antigen (HLA)-B are in strong linkage disequilibrium (LD) with alleles of the C locus, and the sets of B/C haplotypes are found in several populations. The associations between A alleles with C-blocks are weaker, and only a few A/B/C haplotypes (A*0201-B*4501-Cw*1601; A*2301-B*1503-Cw*0202; A*7401-B* 1503-Cw*0202; A*2902-B*4201-Cw*1701; A*3001-B*4201-Cw*1701; and A*3601-B*5301-Cw*0401) are found in multiple populations with intermediate frequencies [haplotype frequency (HF) = 0.010–0.100]. The strength of the LD associations between alleles of HLA-A and HLA-B loci and those of HLA-B and HLA-C loci was on average of the same or higher magnitude as those observed in other non-African populations for the same pairs of loci. Comparison of the genetic distances measured by the distribution of alleles at the HLA class I loci in the sub-Saharan populations included in this and other studies indicate that the Luo population from western Kenya has the closest distance with virtually all sub-Saharan population so far studied for HLA-A, a finding consistent with the putative origin of modern humans in East Africa. In all African populations, the genetic distances between each other are greater than those observed between European populations. The remarkable current allelic and haplotypic diversity in the HLA system as well as their variable distribution in different sub-Saharan populations is probably the result of evolutionary forces and environments that have acted on each individual population or in their ancestors. In this regard, the genetic diversity of the HLA system in African populations poses practical challenges for the design of T-cell vaccines and for the transplantation medical community to find HLA-matched unrelated donors for patients in need of an allogeneic transplant.

Authors' affiliation:

K. Cao[1,2],
A.M. Moormann[3],
K.E. Lyke[4],
C. Masaberg[1,2],
O.P. Sumba[5],
O.K. Doumbo[6],
D. Koech[7],
A. Lancaster[8],
M. Nelson[8],
D. Meyer[8],
R. Single[8,9],
R.J. Hartzman[2],
C.V. Plowe[4],
J. Kazura[3],
D.L. Mann[10],
M.B. Sztein[4],
G. Thomson[8],
M.A. Fernández-Viña[1,2]

[1]Department of Oncology, Georgetown University, Washington DC, WA, USA

[2]C.W. Bill Young DoD Marrow Donor Program, Naval Medical Research Institute/ Georgetown University, Kensington, MD, USA

[3]The Center for Global Health and Diseases at Case Western Reserve University, Cleveland, OH, USA

[4]The Center for Vaccine Development, University of Maryland, Baltimore, MD, USA

[5]The Center for Vector Biology and Control Research, KERMI, Kisumu, Kenya

[6]Malaria Research and Training Center, Pharmacy and Dentistry, University of Bamako, Mali, West Africa

[7]Kenya Medical Research Institute, Nairobi, Kenya

[8]Department of Integrative Biology, University of California, Berkeley, CA, USA

[9]Medical Biostatistics, University of Vermont, Burlington, VT, USA

[10]Immunogenetics Division, Department of Pathology, University of Maryland Medical System, Baltimore, MD, USA

Correspondence:
Marcelo A. Fernández-Viña, PhD
C.W. Bill Young/DoD Marrow Donor Program
Naval Medical Research Institute/Georgetown University
5516 Nicholson Lane,
Building A, 4th Floor
Kensington, MD 20895, USA
Tel: +1-301-998-8900
Fax: +1-301-998-8946
e-mail: mfervina@erols.com

Africa is thought to be the homeland of all modern humans. Data from both archeological and genetic studies support the model of a recent African origin for human evolution (1–3). This model proposes that all non-African human populations descended from a common *Homo sapiens* ancestor that evolved in Africa and radiated to the rest of the world. Under this model, it is predicted that all genetic lineages derive from recent common African ancestors, and all the genetic variations of non-African populations should be found as a subset of the modern African populations. Additional race-specific allelic and haplotypic diversification may have arisen after the out-of-Africa migrations. The genetic makeup and diversity of modern African populations probably reflects the effects of both diversifications, before the separation of racial and ethnic groups and evolutionary events that occurred after the migrations out of the African continent. Studies using mitochondrial and nuclear DNA markers are highly illustrative of this model and consistently indicate that Africa is the most genetically diverse region in the world (4–10).

The major histocompatibility complex (MHC) is one of the most polymorphic genetic systems of many species, including human leukocyte antigen (HLA) in humans. The class I and class II MHC genes encode cell-surface heterodimers that play an important role in antigen presentation, tolerance, and self/non-self recognition (11–13). The HLA molecules bind intracellularly processed antigenic peptides, forming complexes that are the ligands of the antigen receptors of T lymphocytes (13–16). The so-called peptide-binding specificity pockets of the class I molecules accommodate the peptide's side chains (13–17). The extensive population polymorphism of the *MHC* genes may have resulted from selective pressures and functional adaptations (18–20). It has been shown that the highest degree of variability of MHC proteins is found in residues pointing to the peptide-binding region (21, 22). This variation results in differences in immune responses among individuals. In addition to their natural biological function, i.e. to bind and present peptides, the class I and class II histocompatibility antigens play an important role in allogeneic transplantation. Matching for the alleles at the class I and class II MHC loci impacts the outcome of both solid-organ (23, 24) and hematopoietic stem cell (24–26) allogeneic transplants.

Several studies of the population distribution of HLA alleles among Africans have shown significant HLA polymorphism (27–52). Most of the studies that utilized molecular typing methods in African populations have focused on the analysis of loci of the class II region (34–47). Most analyses of frequencies of HLA class I alleles in populations from this continent were performed by serological procedures or by utilizing intermediate-/low-resolution tests (27–33). The International Histocompatibility Workshops held in 1991 and 1996 (34–36) utilized molecular methods to investigate the distribution of HLA class II alleles in various ethnic groups, and highly informative data were collected from those studies and many other independent studies (37–46). In contrast, only a few studies performing high-resolution molecular testing have analyzed the distribution of alleles of the classical class I loci in sub-Saharan populations (47–52). Unfortunately, complete haplotypic information could not be gathered from these studies, because all class I loci were not tested in the same participants.

There is extensive allelic variation in the class I loci with more than 250 HLA-A, 500 HLA-B, and 120 HLA-C alleles (for more information, see website at http://www.anthonynolan.com). The performance of high-resolution molecular typing warrants the precise examination of the population distribution of alleles at the class I loci and allows a detailed analysis of the haplotypic relationships between alleles of different loci. High-resolution testing of alleles at the class I loci of non-African populations has shown that stronger evidence for balancing selection is usually seen at the HLA-B locus when compared to the HLA-A locus (51–64). Linkage disequilibrium (LD) is the condition by which the frequency of a haplotype is significantly different from that expected from the product of the allelic frequencies at each locus. Differences between observed and expected haplotype frequencies (HFs) [delta (Δ)] determine the strength of the associations between alleles at different loci. It can be postulated that the ancestral African populations should have fewer sites in LD and shorter haplotype blocks, when compared with non-African populations, because they have maintained larger population sizes and there has been more time for recombination and mutation to decrease the levels of LD. In a previous study, we observed that the relative Δ-values (D'_{ij}) for the haplotypes defined by alleles at the HLA-A and HLA-B loci of African Americans were an average, lower than those of other populations (63). However, it can be anticipated that examination of patterns of LD in single, distinct African population may be different, and the strength of the associations may be of similar magnitude as the one observed in non-African populations.

The aims of the present study were to investigate the extent of the HLA polymorphism and the distribution of alleles in each of the HLA class I loci and to define the levels of haplotypic diversity in five sub-Saharan populations from different geographic areas. These results will help define and understand the genetic relationships determined by the precise analysis of the distribution of HLA alleles between the populations included in the present and previous studies.

Materials and methods

Subjects studied

We studied 852 unrelated individuals from five African populations. These samples included 241 study participants from Nandi, 265 Luo

from Kenya, 138 from Mali, 163 from Uganda, and 45 from Zambia. The alleles at the HLA-A, HLA-B, and HLA-C loci were typed in all participants.

In Kenya, there are many different populations and more than 60 languages are spoken. Nandi and Luo populations included in this study are believed to be highly homogeneous. The participants from Kanyawegi (−0.11 degrees latitude; 34.63 degrees longitude), which is located in the Kisumu district of the Nyanza Province in western Kenya, were predominantly (97%) of Luo ethnicity and their primary language was Luo. The group from Kipsamoite (0.33 degrees latitude; 35.01 degrees longitude), which is located in the Nandi district in the Rift Valley Province of western Kenya, included all subjects of Nandi ethnicity, a subgroup of Kalenjin, and their primary language was Nandi. Both Luo and Nandi languages belong to the major linguistic Nilo-Saharan and Eastern Sudanic-Nilotic groups.

The population studied from Mali is relatively homogeneous. The samples included 80.7% of subjects self-reported as Dogon, 5.3% as Peuhl (also known as Fulani), 3.5% as Bambara, and 10.5% from other ethnic groups. The Dogon speak various dialects of the Dogon language. Other languages like Peuhl, Bambara, and French are also spoken in the study area. The Dogon language belongs to the broad group Niger/Atlantic/Volta Congo.

Samples from Uganda and Zambia were collected from random individuals living in the capital cities, Kampala and Lusaka, respectively. Both countries are located in Central East Africa and represent a variety of populations and languages. Even though the collected samples may have limited application for anthropological studies, the results reported here illustrate the genetic diversity of HLA in Africa, as reported in studies that examined other loci.

Typing of alleles at the HLA-A, HLA-B, and HLA-C loci

HLA-A, HLA-B, and HLA-C alleles were typed using the polymerase chain reaction and sequence-specific oligonucleotide probe (PCR-SSOP) hybridization methods, as described previously (63, 65). Briefly, extracted DNA was amplified using intronic locus-specific primers that span exons 1, 2, and 3. The polymerase chain reaction (PCR) products were immobilized onto positively charged nylon membranes and were hybridized with sets of probes matching polymorphic sequences of HLA-A, HLA-B, and HLA-C alleles. The hybridization signal was detected using a chemiluminescent substrate on radiography film. Alleles were assigned based on the hybridization patterns with sets of oligonucleotide probes previously described (65) and with additional probes (Cao et al., unpublished data) to unambiguously identify all genotypes for HLA class I alleles defined up to June 2001. All genotypic ambiguities were tested and resolved using group-specific amplification and SSOP hybridization methods, with the exception of alleles presenting differences outside exons 2 and 3 (63, 65). The typing systems allowed the identification of alleles differing by silent substitutions; however, because some genotypes were ambiguous in the fifth and sixth digit, all groups of alleles differing only by synonymous nucleotide substitutions were collapsed into a four-digit assignment. In the present study, we did not test for polymorphisms outside exons 2 and 3; therefore, some groups of alleles with sequence differences only in the other exons were not distinguished. Indistinguishable alleles were assigned as the allele with the lowest number followed by a letter G, indicating that this is a group of alleles (e.g. Cw*0701_G designates the ambiguity Cw*0701 or Cw*0706). Eight groups of HLA-A alleles had this type of ambiguity: A*0101_G (A*01011 and A*0104N); A*0201_G (A*02011, A*0209, and A*0243N); A*0207_G (A*0207 and A*0215N); A*2301_G (A*2301 and A*2307N); A*2402_G (A*2402101, A*2402102L, A*2409N, and A*2411N); A*2403_G (A*2403 and A*2433); A*6801_G (A*68012 and A*6811N); and A*7401_G (A*7401, A*7402). For HLA-B, there were seven groups of alleles differing only outside exons 2 and 3: B*0705_G (B*0705 and B*0706); B*1512_G (B*1512 and B*1519); B*1801_G (B*1801 and B*1817N); B*2705_G (B*2705 and B*2713); B*3501_G (B*3501, B*3540N, and B*3542); B*4402_G (B*4402, B*4419N, and B*4427); and B*5101_G (B*51011, B*5111N, B*5130, and B*5132). There were six groups of HLA-C indistinguishable alleles that included: Cw*0401_G (Cw*0401101, Cw*0401102, and Cw*0409N); Cw*0501_G (Cw*0501 and Cw*0503); Cw*0701_G (Cw*07011, Cw*07012, and Cw*0706); Cw*0704_G (Cw*0704 and Cw*0711); Cw*1701_G (Cw*1701, Cw*1702, and Cw*1703); and Cw*1801_G (Cw*1801 and Cw*1802).

The distinction between alleles differing only by silent substitutions was not made consistently, because the typing reagents did not allow for their assignment in all genotypes; therefore, here we describe frequencies and associations of alleles defined at the fourth-digit level, reflecting protein-sequence differences only.

Putative novel alleles with hybridization patterns that do not match previously described alleles were found in four instances. Heterozygous sequence-based typing confirmed the existence of the novel patterns (data not shown) and was locally designated as a variant (V) of the allele with highest sequence homology. In a few samples, genotype ambiguities were not resolved at the allele level, and the alleles were designated as undefined.

Allelic lineages

Following previous classifications, alleles were grouped according to their nucleotide-sequence similarity (18, 66–68).

Statistical analysis

Most analyses were performed using the computer package PYPOP (69, 70). Allele frequencies (AFs) (denoted as p_i or q_j) were obtained by direct counting, assuming no blank frequencies. The frequency for the ij^{th} haplotype is denoted as HF_{ij} (below).

Test of fit to Hardy–Weinberg genetic equilibrium

Overall deviations from Hardy–Weinberg proportions (HWPs) were tested using the exact test of Guo and Thompson (71), implemented through ARLEQUIN (72) via PYPOP and by χ^2 testing when expected values were ≥ 5. χ^2 tests were investigated for overall common genotypes, all heterozygotes, overall heterozygotes for a specific allele (e.g. A*0101, X, where X represents all non-A*0101 alleles), and individual genotypes.

Allele frequency heterogeneity

Heterogeneity of AF differences for each locus was investigated using contingency table methods. Classes with expected frequencies less than 5 were combined before χ^2 testing.

Homozygosity statistic

Watterson's (73, 74) homozygosity F statistic ($F = \Sigma p_i^2$), calculated as the expected proportion of homozygotes under HWPs, was used as a measure of the AF distribution and compared with the distribution expected under the neutral model. The test compares the observed F with the distribution of values expected in a neutral sample with the same number of alleles and sample size. The homozygosity test was applied using the exact test described by Slatkin (75).

Haplotype frequencies and linkage disequilibrium estimation

Two- and three-locus HFs were estimated via the expectation–maximization (EM) algorithm (72) within a module of PYPOP (69, 70). Deviations from random association of alleles at two loci were calculated for each haplotype ($D_{ij} = HF_{ij} - p_i q_j$). To account for differing AFs at the loci, Lewontin's (1964) normalized disequilibrium value, also referred to as the relative normalized Δ-value, was used ($D'_{ij} = D_{ij}/D_{max}$).

Overall LD was estimated using three different approaches. The overall D' statistic described by Hedrick (1987) was computed by summing the absolute value of the D'_{ij} over all haplotypes in a multi-allelic two-locus system, using the products of AFs at the loci as weights (69). The significance of overall LD between two loci was tested using the permutation distribution of the likelihood ratio test (Excoffier and Slatkin 1995).

A second measurement, W_n, also known as Cramer's V statistic (Cramer 1946), was also used to measure global LD (69).

We developed an additional statistic, the average positive relative Δ-value (APRDV), to measure the strength of the associations between alleles at two different loci. This measure gives the average of the positive normalized disequilibrium values (relative delta values). The APRDV is the sum of the frequency-weighted D'_{ij} values over the set of haplotypes that have positive disequilibrium coefficients (i.e. $D'_{ij} > 0$). For this statistic, the HFs are used as weights.

$$\mathrm{APRDV} = \sum_{D'_{ij}>0} HF_{ij} D'_{ij}.$$

The APRDV statistic ranges from 0 to 1, reaching its maximum if all haplotypes for a pair of loci had absolute associations ($D'_{ij} = 1$). APRDV will be zero if there were no LDs between all pairs of alleles ($D'_{ij} = 0$).

The statistical significance of the LD between each pair of alleles of different loci was evaluated using Mickey's formula (76):

$$X^2 = \left[\frac{4n\left[\mathrm{HF}(ij) - \mathrm{GF}(i)\mathrm{GF}(j)^2\right]}{\mathrm{GF}(i)\mathrm{GF}(j)[2 - \mathrm{GF}(i)[2 - \mathrm{GF}(j)]]}\right] \left[1 + \frac{[\mathrm{HF}(ij) - \mathrm{GF}(i)\mathrm{GF}(j)]}{2[1 - \mathrm{GF}(i)][1 - \mathrm{GF}(j)]}\right]^2.$$

Genetic distances between populations

Genetic distances (d) were estimated using the definition given by Cavalli-Sforza and Bodmer (77)

$$d_{j,k} = \sqrt{1 - \sum i \sqrt{P_{i,j} P_{i,k}}}.$$

where $d_{j,k}$ is the distance between population j and population k and i,j is the AF of allele i in population j. Genetic distance estimates the lack of overlap between alleles at each locus from two different populations.

Results

Among the 852 participants from five African sub-Saharan populations, we identified 45 HLA-A alleles, 73 HLA-B alleles, and 30 HLA-C alleles. The number of alleles found in each population ranged from 20 to 35 for HLA-A, from 30 to 51 for HLA-B, and from 13 to 25 for

Heterozygosity at the HLA-A, HLA-B, and HLA-C loci in five African groups

Ethnic group	Number of subjects[a]	Number of alleles[b]	Observed heterozygosity[c]	Expected HW heterozygosity[d]
HLA-A				
Kenyan Nandi	241	28	0.90	0.93
Kenyan Luo	265	30	0.95	0.94
Malians	138	21	0.91	0.90
Ugandans	163	35[e]	0.91	0.93
Zambians	43	20	0.88	0.90
HLA-B				
Kenyan Nandi	240	39	0.95	0.95
Kenyan Luo	265	47	0.94	0.94
Malians	138	31	0.95	0.92
Ugandans	161	51[e]	0.95	0.97
Zambians	44	30	0.95	0.95
HLA-C				
Kenyan Nandi	240	21	0.88	0.89
Kenyan Luo	265	22	0.89	0.91
Malians	129	19	0.88	0.84
Ugandans	163	25[e]	0.91	0.92
Zambians	45	13[e]	0.89	0.90

HLA, human leukocyte antigen; HW, Hardy–Weinberg.
[a]The number of subjects studied (*n*).
[b]The number of alleles identified (*k*).
[c]The observed heterozygosity is obtained by direct counting.
[d]The expected value is calculated under the assumption of HW equilibrium.
[e]At the four-digit level, there was one allele less (HW and the Ewen–Watterson neutrality testing were carried out on this smaller number).

Table 1

HLA-C (Table 1). The smallest and largest number of alleles at each of these class I loci was consistently observed in the Zambian and Ugandan populations, respectively. These observations may be related to the small sample size collected from the former population, while the age of the population or/and genetic admixture may explain the allelic diversity observed in the latter (see below).

Hardy–Weinberg equilibrium and high frequency of heterozygous genotypes at all the HLA class I loci

The overall distribution of genotypes at each of the HLA class I loci did not show significant deviations from those expected under HWP (Table 1). All loci displayed relatively higher levels of heterozygosity that ranged from 0.875 (HLA-C locus in Malians) to above 0.95 (HLA-B locus in Nandi from Kenya, Ugandans, and Zambians). As observed in Caucasian and native American populations (53–57, 63) the HLA-B locus has the largest number of alleles and the highest frequencies of heterozygous genotypes. Testing by χ^2 of overall deviation from HW, all heterozygotes, heterozygotes by a specific allele, and common genotypes showed no evidence of significant deviations.

For all population and loci, the homozygosity *F* statistic was less than that expected under neutrality. Each of the three loci had at least two populations where the AF distributions significantly differed from neutrality in the direction of balancing selection (Table 2). Interestingly, because of the high frequency of A*3002 in Zambians and Cw*1601 in Malians (see below, Tables 3 and 5), the AF distributions for HLA-A and HLA-C in Zambians and Malians, respectively, are virtually identical to those expected under neutrality.

The AFs of HLA-A, HLA-B and, HLA-C alleles are summarized in Tables 3, 4, and 5, respectively. To compare the population distribution of alleles of the same locus in different populations, we sorted the alleles in decreasing order of frequencies and plotted their cumulative frequencies and the corresponding number of alleles. Graphical representations of allele distributions for HLA-A, HLA-B, and HLA-C loci are shown in Fig. 1(A), (B), and (C), respectively. Figure 1(A) shows the distribution of HLA-A alleles in the five populations. It can be seen that the curves corresponding to the distributions of alleles of HLA-A in Zambians and Malians virtually overlap in all points

Homozygosity test of neutrality at HLA-A, HLA-B, and HLA-C loci in five African groups

			Slatkin's implementation of EW homozygote test of neutrality		
Ethnic Group	Number of Chromosome (2*n*)	Number of alleles	Observed[a]	Expected[b]	*P*-value of *F*[c]
HLA-A					
Kenyan Nandi	482	28	0.073	0.134	0.009
Kenyan Luo	530	30	0.060	0.127	0.001
Malians	276	21	0.107	0.161	NS
Ugandans	326	35[d]	0.068	0.097	NS
Zambians	86	20	**0.112**	**0.118**	NS
HLA-B					
Kenyan Nandi	480	39	0.054	0.092	0.012
Kenyan Luo	530	47	0.057	0.076	NS
Malians	276	31	0.088	0.103	NS
Ugandans	322	51[d]	0.037	0.059	0.003
Zambians	88	30	0.059	0.069	NS
HLA-C					
Kenyan Nandi	480	21	0.113	0.183	NS
Kenyan Luo	530	22	0.098	0.177	0.025
Malians	258	19	**0.160**	**0.176**	NS
Ugandans	326	25[d]	0.087	0.145	NS
Zambians	90	13[d]	0.106	0.219	0.001

HLA, human leukocyte antigen; EW, Ewen–Watterson; NS, not significant.
[a]The observed homozygosity H statistic is the EW test of neutrality.
[b]The expected homozygosity under neutrality is that obtained from the Ewens neutrality sampling formula for the same 2*n* and *k* as observed in the sample.
[c]The *P*-value is one-sided; significant observed *F*-value less than expected is indicative of balancing selection.
[d]At the four-digit level there was one allele less.

Table 2

below the cumulative frequency of 0.80. The curves corresponding to both Kenyan population and Ugandans are shifted to the right, and these indicate that more alleles are required to cover the same combined cumulative frequency. This observation also indicates that the A locus is more diverse in Kenyans and Ugandans, when compared with Zambians and Malians. Figure 1(B) shows that fewer HLA-B alleles in Malians cover the same proportion of alleles (cumulative frequency), when compared with those in other populations, indicating less diversity for this locus in this population. Both Kenyan populations at all loci and the Zambians at HLA-B and HLA-C present intermediate ranges of diversity. The HLA-B locus appears to be extremely diverse in Ugandans. Figure 1(C) shows a less diverse distribution of HLA-C alleles in Malians than in the other groups.

Population distribution of HLA-A alleles

Up to 12 HLA-A allelic lineages have been defined so far in humans (18, 66–68). Alleles of each of them were observed in the five populations analyzed in the present study. The lineages A*01/36, A*02, A*03, A*25/26/34/43/66, A*29, A*30, A*32/74, A*31/33, and A*68/69 were present with multiple alleles. In some populations, there were at least two alleles of the same family that we found with frequencies of similar magnitude (Table 3). Examples of balanced frequencies of multiple alleles were those of A*02 subtypes in Nandi and the subtypes of A*30 (3001 and 3002) in most of the populations.

Table 3 also shows that 25 of the 45 HLA-A alleles identified in this study were present in three or more populations. The alleles found in multiple populations, when combined, account for a large proportion of the HLA-A alleles of each population (sum of AFs of shared allele was 0.969 in Nandi, 0.942 in Luo, 0.989 in Malians, 0.914 in Ugandans, and 0.977 in Zambians). Despite the sharing of alleles, there are significant differences in the distributions of these alleles in each group (see below, analysis of genetic distances). Both Kenyan populations show similarities in the allele distributions, because there is not a single predominant HLA-A allele, and the nine most frequent alleles have frequencies ranging from 0.118 to 0.050. Among these alleles, $A*0201_G$, $A*0101_G$, A*6802, $A*2301_G$, A*6601, and A*2902 are present

Allele frequencies for HLA-A in five African populations

HLA-A	Kenyan Nandi	Kenyan Luo	Malians	Ugandans	Zambians	Zulu ([51])	Cameroon ([47])	African specific
A*0101[e]	0.1183[b]	0.0736[c]	0.0073	0.0828[c]	0.0349	0.0400	0.0110	
A*0102	0.0270	0.0038	0.0109	0.0031				Yes
A*0103	0.0166							
A*0201[e]	0.1183[b]	0.1151[b]	0.0833[c]	0.1841[a]	0.1279[b]	0.0250	0.0710[c]	
A*0202	0.0664[c]	0.0302	0.0761[c]	0.0337	0.0233	0.0350	0.0820[c]	Yes
A*0204							0.0060	
A*0205	0.0871[c]	0.0264	0.0217	0.0215		0.0550[c]	0.0220	
A*0206				0.0031	0.0116			
A*0211							0.0060	
A*0214	0.0083	0.0151		0.0061	0.0116		0.0060	Yes
A*0217			0.0036					
A*0225		0.0038						Yes
A*0301	0.0311	0.0359	0.0435[d]	0.0552[c]	0.0581[c]	0.0600[c]	0.0830[c]	
A*0302				0.0031				
A*1101				0.0429[d]				
A*2301[e]	0.0705[c]	0.0887[c]	0.2283[a]	0.0675[c]	0.0814[c]	0.1000[c]	0.1870[a]	
A*2402[e]	0.0104	0.0057		0.0429[d]		0.0100	0.0050	
A*2403[e]				0.0061				
A*2406			0.0036					
A*2413			0.0036					
A*2501				0.0123				
A*2601	0.0083	0.0057		0.0153	0.0116	0.0100	0.0170	
A*2608				0.0061				
A*2612		0.0019					0.0060	Yes
A*2901	0.0125	0.0113		0.0061		0.0150		
A*2902	0.0477[d]	0.0528[c]	0.0326	0.0307	0.0581[c]	0.1100[b]	0.1040[b]	
A*3001	0.0477[d]	0.0642[c]	0.1413[b]	0.0307	0.1395[b]	0.0950[c]	0.0550[c]	
A*3002	0.0373	0.0585[c]	0.0399	0.0399	0.2326[a]	0.0850[c]	0.0600[c]	
A*3003							0.0110	
A*3004	0.0042	0.0132				0.0150	0.0110	Yes
A*3006							0.0060	
A*3101	0.0042	0.0132		0.0245	0.0349	0.0050	0.0110	
A*3103	0.0062	0.0076						
A*3104				0.0031				Yes
A*3201	0.0187	0.0264	0.0073	0.0215		0.0250	0.0110	
A*3301	0.0021	0.0132	0.0181	0.0153	0.0116			
A*3303	0.0021	0.0094	0.0942[c]	0.0307		0.0050	0.0280	
A*3402	0.0498[d]	0.0264	0.0362	0.0276	0.0116	0.0550[c]	0.0220	Yes
A*3601	0.0125	0.0302		0.0123	0.0349		0.0220	Yes
A*4301					0.0116	0.0300		Yes
A*6601	0.0498[d]	0.0679[c]		0.0276	0.0116	0.0200	0.0600[c]	
A*6602	0.0042	0.0094				0.0150	0.0110	Yes
A*6801[e]	0.0062	0.0151	0.0362	0.0276		0.0300	0.0060	

Table 3

Continued

HLA-A	Kenyan Nandi	Kenyan Luo	Malians	Ugandans	Zambians	Zulu ([51])	Cameroon ([47])	African specific
A*6802	0.1162[b]	0.0755[c]	0.0616[c]	0.0522[c]	0.0465[d]	0.0950[c]	0.0280	
A*6901				0.0031			0.0060	
A*7401[e]	0.0166	0.0717[c]	0.0362	0.0522[c]	0.0233	0.0650[c]	0.0500[c]	Yes
A*7403		0.0283		0.0031				Yes
A*8001			0.0145	0.0031	0.0233			Yes
A-2601V[f]				0.0031				
2n	482	530	276	326	86	200	182	

HLA, human leukocyte antigen.
[a]AF $\geq$ 0.15.
[b]0.10 $<$ AF $<$ 0.15.
[c]0.05 $\leq$ AF $\leq$ 0.10.
[d]0.04 $<$ AF $<$ 0.05.
[e]A group of alleles which were not distinguished in this study. These alleles of the group only differ by nucleotide substitution, outside exons 2 and 3 (*Materials and methods*).
[f]A new allele, related in sequence to A*2601; however, it has not been officially named.

Table 3

in both Kenyan groups with relative high frequencies. The alleles A*0202, A*0205, A*3402, A*7401$_G$, A*3001, and A*3002 are present with relatively high frequency in one population but with low frequency in the other Kenyan population. Because of the relatively even frequencies of several HLA-A alleles, both Kenyan populations have low frequencies of homozygous genotypes, and their distributions significantly depart from the expectations under the theory of neutral distribution of alleles (Table 2). A similar observation was made in another study of the Zulu population (51, 52) that showed several alleles with intermediate frequencies (A*2902, AF = 0.11; A*2301, AF = 0.1; A*3001, AF = 0.095; A*6802, AF = 0.095; and A*3002, AF = 0.085).

In contrast to both Kenyan groups, the HLA-A locus presents a few predominant alleles in Ugandans, Zambians, and Malians. The frequencies of a few predominant alleles contrast with those of the remaining alleles at the locus. The most frequent allele in each of these populations was almost two times more frequent than the second most frequent one. The alleles with the highest frequencies observed in these populations are A*2301$_G$ and A*3001 in Malians, A*0201 in Ugandans, and A*3002 followed by A*3001 and A*0201$_G$ in Zambians (Table 3). In a population study of Cameroonians (47), the most frequent allele observed was A*2301 (AF = 0.187), which also showed a contrasting frequency with the next most frequent allele (A*2902 and AF = 0.104) (47).

In this study, we identified several alleles that appear to be specific to African populations. The alleles A*0202, A*3601, A*3402, and A*7401$_G$ have intermediate frequencies in all African populations and are not found in other populations studied so far (47, 48, 51–63). Other alleles found to be specific only to some African populations and observed at lower frequencies were A*0102, A*0214, A*0225, A*4301, A*6602, A*7403, and A*8001. A new allele, provisionally named A-2601 V, was identified in one Ugandan subject.

Population distribution of HLA-B alleles

As observed with HLA-A, there were many HLA-B alleles identified in all the populations described in this study. These included 17 alleles present in all populations, nine alleles found in four populations, and six additional alleles found in three populations. These 32 alleles, when combined, account for 77–82% of the AFs of Ugandans and Malians and as much as 92% of the AFs of both Kenyan populations and Zambians. Examples of all 20 allelic lineages defined for the HLA-B locus in humans (18, 66–68) were found in each of the African populations described here.

As has been observed in many populations from around the world, there are no predominant HLA-B alleles in the populations included in the present study. Several intermediate-frequency alleles were found in each of the African populations. Ugandans present many intermediate- and low-frequency alleles, with only three alleles (B*1503, B*0801, and B*0702) having frequencies higher than 0.05. The other four populations present five to eight alleles with intermediate frequencies ranging from 0.05 to 0.16. The alleles, B*1503, B*4201, and B*5301, that are not common in Ugandans are found among the top five highest frequency alleles in the other four populations. B*4501 is found among the most frequent alleles in three populations.

As observed for the HLA-A, both Kenyan populations presented similarities in the distribution of HLA-B alleles. In addition to the four HLA-B alleles mentioned above, both Kenyan populations presented two subtypes of B58, B*5801 and B*5802, among the most frequent alleles. Neither of the subtypes of B58 ranks among the

Allele frequencies for HLA-B in five African populations

HLA-B	Kenyan Nandi	Kenyan Luo	Malians	Ugandans	Zambians	Zulu ([52])	Cameroon ([47])	African specific
B*0702	0.0104	0.0245[d]	0.0580[c]	0.0590[c]	0.0455[d]	0.0450[d]	0.0600[c]	
B*0705[e]	0.0021	0.0038			0.0114	0.0200		
B*0801	0.0458[d]	0.0302[d]	0.0073	0.0621[c]	0.0341[d]	0.0750[c]	0.0540[c]	
B*1302	0.0229[d]	0.0113	0.0036	0.0280[d]	0.0114	0.0150	0.0270[d]	
B*1303	0.0083							Yes
B*1401	0.0021	0.0113	0.0109	0.0155	0.0455[d]	0.0400[d]	0.0110	
B*1402	0.0063	0.0434[d]	0.0109	0.0373[d]	0.0341[d]	0.0250[d]	0.0050	
B*1403		0.0019					0.0160	Yes
B*1501				0.0248[d]	0.0227[d]	0.0050	0.0110	
B*1503	0.0792[d]	0.0887[c]	0.0688[c]	0.0621[c]	0.0568[c]	0.0950[c]	0.0490[d]	Yes
B*1509	0.0021		0.0073					
B*1510	0.0167	0.0359[d]	0.0217[d]	0.0248[d]	0.0455[d]	0.0850[c]	0.0110	Yes
B*1516	0.0104	0.0038	0.0109	0.0062		0.0050	0.0110	Yes
B*1517	0.0104	0.0132	0.0036	0.0031			0.0110	
B*1518		0.0019	0.0036		0.0114			
B*1531		0.0038		0.0062				Yes
B*1537		0.0019						Yes
B*1555				0.0031				Yes
B*1801[e]	0.0479[d]	0.0434[d]	0.0073	0.0497[d]	0.0341[d]	0.0200	0.0270[d]	
B*1803		0.0057		0.0031	0.0114			
B*2703	0.0250[d]	0.0076	0.0036	0.0062			0.0050	Yes
B*2705[e]		0.0019		0.0124				
B*3501[e]	0.0333[d]	0.0340[d]	0.1268[b]	0.0373[d]	0.0227[d]	0.0400[d]	0.0710[c]	
B*3502	0.0063			0.0155				
B*3503				0.0062				
B*3508				0.0062				
B*3701	0.0083			0.0031			0.0050	
B*3801				0.0062				
B*3901		0.0057		0.0155				
B*3903		0.0057						
B*3910	0.0104	0.0113	0.0073			0.0150	0.0050	Yes
B*4001				0.0155	0.0227[d]			
B*4002				0.0155				
B*4012	0.0125	0.0113						Yes
B*4016		0.0038					0.0110	Yes
B*4101	0.0313[d]	0.0113		0.0093	0.0114	0.0050	0.0050	
B*4102	0.0063	0.0057	0.0073	0.0031	0.0227[d]	0.0050		
B*4201	0.0688[c]	0.0774[c]	0.1377[b]	0.0124	0.1477[b]	0.1200[b]	0.0490[d]	Yes
B*4202		0.0038		0.0093				Yes
B*4402[e]				0.0342[d]				
B*4403	0.0083	0.0170	0.0145	0.0466[d]	0.0227[d]	0.1050[b]	0.0650[c]	
B*4405					0.0114			
B*4407							0.0220[d]	

Table 4

Continued

HLA-B	Kenyan Nandi	Kenyan Luo	Malians	Ugandans	Zambians	Zulu ([52])	Cameroon ([47])	African specific
B*4415	0.0063	0.0019						
B*4501	0.0566[c]	0.0660[c]	0.0616[c]	0.0280[d]	0.0455[d]	0.0350[d]	0.0330[d]	
B*4701	0.0125				0.0114		0.0110	
B*4702		0.0019		0.0031				
B*4703	0.0021	0.0132		0.0031			0.0050	Yes
B*4901	0.0458[d]	0.0038	0.0254[d]	0.0373[d]	0.0114	0.0100	0.0540[c]	
B*5001		0.0019	0.0109				0.0050	
B*5101[e]	0.0417[d]	0.0226[d]	0.0254[d]	0.0497[d]	0.0568[c]		0.0160	
B*5108				0.0031				
B*5201			0.0833[c]	0.0155				
B*5301	0.0875[c]	0.0679[c]	0.1594[a]	0.0497[d]	0.1023[b]	0.0150	0.1090[b]	
B*5501				0.0124				
B*5601	0.0021		0.0073	0.0031				
B*5701	0.0083	0.0076		0.0311[d]	0.0114			
B*5702	0.0042	0.0057	0.0109		0.0227[d]	0.0100		Yes
B*5703	0.0292[d]	0.0094	0.0073	0.0124	0.0568[c]	0.0400[d]	0.0270[d]	Yes
B*5801	0.1000[c]	0.0698[c]	0.0217[d]	0.0404[d]		0.0400[d]	0.0540[c]	
B*5802	0.0854[c]	0.1245[b]		0.0435[d]	0.0227[d]	0.0800[c]	0.1090[b]	Yes
B*6701						0.0050		
B*7301	0.0021	0.0019						
B*7801			0.0688[c]					Yes
B*8101	0.0396[d]	0.0623[c]		0.0093	0.0227[d]	0.0350[d]	0.0440[d]	Yes
B*8201	0.0021	0.0189		0.0031		0.0050		Yes
B*13und[f]					0.0114			
B*15und[f]				0.0062				
B*35und[f]			0.0036					
B*39und[f]				0.0031				
B*40und[f]				0.0031				
B-0702 V[g]				0.0031				
B-5104 V[g]			0.0036					
2n	480	530	276	322	88	200	184	

HLA, human leukocyte antigen.
[a]AF $\geq$ 0.15.
[b]0.10 $<$ AF $<$ 0.15.
[c]0.05 AF $\leq$ 0.10.
[d]0.02 $<$ AF $<$ 0.05.
[e]A group of alleles which were not distinguished in this study. These alleles of the group only differ by nucleotide substitution, outside exons 2 and 3 (*Materials and methods*).
[f]Undefined alleles from ambiguous genotypes including only alleles of the B*13, B*15, B*39, and B*40 groups.
[g]New alleles related in sequence to B*0702 and B*5104. These alleles have not received official designation.

Table 4

higher frequency alleles in Malians or Zambians. The latter populations show alleles of B5-B35 groups (B^*3501_G, B*7801, and B*5201 in Malians and B^*5101_G in Zambians), B*0702 (in Malians), and B*5703 (in Zambians) among the most frequent ones.

In this study, we were able to categorize alleles that seem to be found strictly in African populations. These included (i) five alleles (B*1503, B*4201, B*5301, B*5802, and B*5703) that were observed with high or intermediate frequencies in most populations, (ii) five

Allele frequencies for HLA-C in five African populations

HLA-C	Kenyan Nandi	Kenyan Luo	Malians	Ugandans	Zambians	Mandenka ([49])	African specific
Cw*0102	0.0021		0.0078	0.0245[d]		0.0312[d]	
Cw*0202	0.0875[c]	0.1000[c]	0.0543[c]	0.0920[c]	0.1000[b]	0.0711[c]	
Cw*0302	0.0458[d]	0.0321[d]		0.0184		0.0671[c]	
Cw*0303		0.0019	0.0116	0.0368[d]	0.0222[d]		
Cw*0304	0.0417[d]	0.0472[d]	0.0155	0.0491[d]	0.0667[c]	0.0743[c]	
Cw*0305				0.0031			
Cw*0401[e]	0.1146[b]	0.1321[b]	0.2132[a]	0.1411[b]	0.1444[b]	0.1946[a]	
Cw*0404	0.0125	0.0076	0.0039				
Cw*0501[e]	0.0167	0.0113	0.0078	0.0399[d]		0.0333[d]	
Cw*0602	0.2167[a]	0.1868[a]	0.0543[c]	0.1166[b]	0.0333[d]	0.0394[d]	
Cw*0701[e]	0.1521[a]	0.1170[b]	0.0620[c]	0.1595[a]	0.1222[b]	0.0869[c]	
Cw*0702	0.0083	0.0377[d]	0.0465[d]	0.0614[c]	0.0444[d]	0.0212[d]	
Cw*0704[e]	0.0646[c]	0.0509[c]		0.0184	0.0222[d]		
Cw*0705		0.0076					
Cw*0801				0.0061			
Cw*0802	0.0063	0.0509[c]	0.0233[d]	0.0614[c]	0.0667[c]		
Cw*0804	0.0146	0.0076	0.0078				Yes
Cw*0802/04						0.0281[d]	
Cw*1202			0.0078	0.0092		0.0182	
Cw*1203	0.0104	0.0113	0.0194	0.0368[d]		0.0030	
Cw*1402	0.0104	0.0038	0.0194	0.0153			
Cw*1403						0.0061	
Cw*1502	0.0063	0.0019		0.0123		0.0343[d]	
Cw*1503				0.0031			
Cw*1505	0.0083	0.0057		0.0031			
Cw*1601	0.0438[d]	0.0453[d]	0.2830[a]	0.0491[d]	0.1000[b]	0.1681[a]	
Cw*1602	0.0125	0.0057		0.0031			
Cw*1604			0.0039				
Cw*1701[e]	0.1021[b]	0.0868[c]	0.1434[b]	0.0276[d]	0.1556[a]	0.0485[d]	
Cw*1801[e]	0.0229[d]	0.0491[d]	0.0155	0.0092	0.1000[b]	0.0250[d]	Yes
Cw*03und						0.0122	
Cw*null						0.0374[d]	
Cw-0304 V[f]				0.0124	0.0222[d]		
2n	480	530	258	326	90	330	

HLA, human leukocyte antigen.
[a]AF $\geq$ 0.15.
[b]0.10 $<$ AF $<$ 0.15.
[c]0.05 $\leq$ AF $\leq$ 0.10.
[d]0.02 $<$ AF $<$ 0.05.
[e]A group of alleles, which were not distinguished in this study. These alleles of the group only differ by nucleotide substitution, outside exons 2 and 3 (*Materials and methods*).
[f]A new allele variant of Cw*030401, which has not been officially designated.

Table 5

alleles (B*1510, B*1516, B*5702, B*8101, and B*7801) found in various populations with intermediate frequencies, (iii) four alleles (B*1803, B*2703, B*3910, and B*8201) found in several populations with relatively low frequency, and (iv) 12 additional alleles that had low frequencies and were found in one or two populations. In addition to the previously described alleles, we identified two novel alleles, provisionally designated B-5104V and B-0702V. These were identified in participants from Mali and Uganda, respectively.

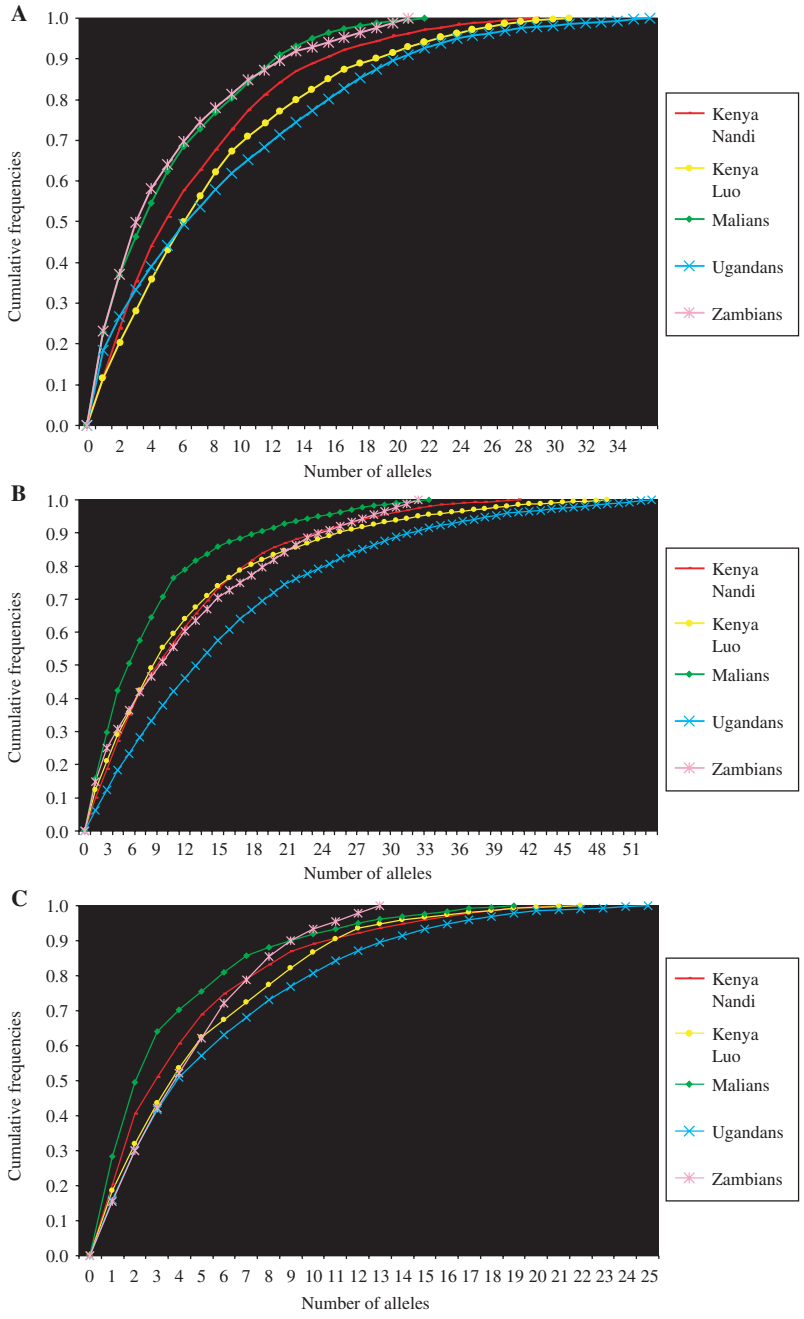

Fig. 1. **Population coverage by human leukocyte antigen (HLA) alleles in five Africans populations.** The alleles of HLA-A (A), HLA-B (B), and HLA-C (C) were sorted according to their allele frequencies in descending order; cumulative frequencies were plotted according the number of alleles. (A) Population coverage at HLA-A; (B) Population coverage at HLA-B; (C) Population coverage at HLA-C.

Population distribution of HLA-C alleles

Ten HLA-C alleles were present in all African populations. These were Cw*0202, Cw*0304, Cw*0401$_G$, Cw*0602, Cw*0701$_G$, Cw*0702, Cw*0802, Cw*1601, Cw*1701$_G$, and Cw*1801$_G$. The combined frequencies of these alleles account for approximately 76% of the *HLA-C* genes in Ugandans, 79% in the Nandi, 85% in the Luo, 91% in Malians, and 93% in Zambians. Five additional alleles were found in four of the five populations, and these widely distributed alleles, when combined, account for more than 90% of the allele pool of each population (Table 5).

Despite the large number of shared alleles among various populations and the observation that these cover a large proportion of the allele pool of each population, the African populations have significant differences in the individual AFs. For example, Cw*0602 has a relatively low frequency in Zambians (0.03), intermediate values in Ugandans and Malians, and high frequencies in both Kenyan populations. Cw*1701$_G$ has low frequency in Ugandans, intermediate frequencies in both Kenyan populations, and high frequencies in Zambians and Malians. Perhaps, one of the most significant fluctuations is the one observed for Cw*1601 that is found with relatively

low or intermediate frequency in most populations (AF ranging from 0.04 to 0.10) and reaches a striking high frequency in Malians (AF = 0.283), which results in nearly half of the population having at least one dose of this allele.

As discussed above, there are differences in the distribution modality from population to population. In Malians, the three most frequent and highly predominant alleles are Cw*1601, Cw*0401_G, and Cw*1701_G, and the combined frequencies account for more than 64% of the HLA-C allele pool. The frequency of these alleles strikingly contrasts with that of the remaining alleles. The fourth most frequent allele, Cw*0701_G, is present with a frequency less than one-half of the preceding allele (Cw*1701_G). The unequal distribution of alleles results in a higher frequency of subjects homozygous for the three predominant HLA-C alleles with high frequency. This distribution of genotypes is close to the one expected under the neutrality theory (Table 2). In contrast, the Ugandans, Zambians, and Luo have several HLA-C alleles with intermediate frequencies, and none of them showed a single predominant allele at this locus.

The allele Cw*1801_G was found in all African populations, with frequencies ranging from as low as <0.01 in Ugandans to more than a 10 times higher frequency in Zambians (AF = 0.100). This allele appears to be one of the few alleles of HLA-C that is exclusively found in Africans. This allele is found forming haplotypes with the African-specific alleles B*8101, 5702, and 5703. Interestingly, Cw*0804 is another African-specific allele that also was found in association with B*8101 (see below). Two other African-specific alleles with silent mutations (Cw*020204 and Cw*030402) also appear associated with the African-specific HLA-B alleles B*1503 and B*1510, respectively.

The frequencies of alleles defined by silent mutations are not individually listed in Table 5 and are collapsed and listed as alleles encoding the same protein sequence. Even though we could not assign alleles differing by silent substitutions consistently, in those samples in which the assignment of silent substitution was possible, we observed invariably tight and absolute associations of HLA-C alleles with alleles of HLA-B. This observation indicates that the silent substitutions may be used as markers of evolutionary events.

Genetic distances estimated on the basis of distributions of alleles of HLA class I loci

We compared the frequencies of alleles at the class I loci and calculated genetic distances (Table 6) between the African populations described in this study and in other studies including the Zulu tribe from South Africa (A and B loci) (51–52), the population of Cameroon (A and B loci) (47), and the Mandenka tribe from Senegal (C locus) (49). We also compared the genetic distances with an admixed population of Kenyans from another study (A and B loci) (48). This admixed population from Kenya (48) showed systematically larger genetic distances with all populations than those observed for the comparisons made with either the Luo or the Nandi Kenyan populations described in the present study. It is possible that the Kenyan populations described here are more homogeneous and, therefore, provide a more accurate estimation of the genetic relationships with other sub-Saharan populations.

Genetic distances with African Americans (63) and Caucasians from US (63) were calculated as well and served as references. Table 6 summarizes the genetic distances calculated from the distribution of alleles at the HLA-A, HLA-B, and HLA-C loci, respectively. It was observed that the closest genetic distances between populations were those between the Luo and Nandi groups. The genetic distances calculated on the basis of the distribution of HLA-A and HLA-B alleles show that the Luo population from Kenya had either the lowest (HLA-A and some HLA-B) or second lowest distance (some HLA-B) with each sub-Saharan African population, with the exception of the Malians. For each African population, both the Malians and the Zambians consistently showed the largest or the second largest genetic distances. The Malians appeared to have the closest distances with Cameroonians (A and B loci), Zambians (B and C loci), Zulus (A locus), and Mandenkans (C locus). Zambians had the closest distances with the Luo population (A and C loci) and the Zulus (B locus). The Ugandans appeared to cluster closer to the two Kenyan populations.

The magnitude of the genetic distances was greater with calculations made using the distribution of HLA-B alleles. Genetic distance values obtained with HLA-A AF data were higher than those obtained with HLA-C frequencies. If neighboring populations are not considered (Ugandans are close to Kenyans, and Zambians appear close to both Kenyans and Zulus), it appears that the genetic distances between populations that are geographically distant have similar magnitudes.

Interestingly, each African population with the exception of the Zulus appears to have close genetic distances with the African Americans. The distances with African Americans ranked either first or second, when HLA-A and -C (Table 6) allele distributions were used for the calculations of genetic distances, and were first to third, when the distances were calculated on the basis of the distribution of HLA-B alleles (Table 6). The Ugandan population shows the closest distance with the African American population; in addition, this population from Central Africa appears to be closer to Caucasians from US, while the other populations from continental Africa present little overlap with Caucasians in the distribution of alleles of HLA-A and B loci (with genetic distances ranging from $d = 0.56$ to $d = 0.74$).

Genetic distances between eight African populations estimated by the distribution of HLA-A, HLA-B, and HLA-C alleles

Ethnic group	Kenyan Nandi[a]	Kenyan Luo[a]	Malian[a]	Ugandan[a]	Zambian[a]	Zulu[b] ([51])	Cameroonian[c] ([47])	Kenyans[d] ([48])	Mandenka[f] ([49])	AFAM[e]	Caucasians[e]
HLA-A											
Kenyan Nandi	0	0.27[g]	0.47	0.36[h]	0.48	0.35	0.39	0.39[h]		0.33	0.56
Kenyan Luo	0.27[g]	0	0.45	0.32[g]	0.41[g]	0.33[g]	0.33[g]	0.33[g]		0.30	0.56
Malian	0.47	0.45	0	0.46	0.49	0.42	0.41	0.48		0.37	0.68
Ugandan	0.36	0.32[h]	0.46	0	0.46	0.43	0.41	0.40		0.23	0.37
Zambian	0.48	0.41	0.49	0.46	0	0.45	0.43	0.44		0.42	0.63
Zulu	0.35[h]	0.33	0.42[h]	0.43	0.45	0	0.35[h]	0.41		0.36	0.62
Cameroonian	0.39	0.33	0.41[g]	0.41	0.43[h]	0.35[h]	0	0.41		0.33	0.62
Kenyans	0.39	0.33	0.48	0.40	0.44	0.41	0.41	0		0.39	0.64
AFAM	0.33	0.30	0.37	0.23	0.42	0.36	0.33	0.39		0	0.46
CAU	0.56	0.56	0.68	0.37	0.63	0.62	0.62	0.64		0.46	0
HLA-B											
Kenyan Nandi	0	0.30[g]	0.54	0.46	0.47	0.44	0.35[g]	0.33[g]		0.43	0.72
Kenyan Luo	0.30[g]	0	0.55	0.45[g]	0.45[h]	0.37[g]	0.37[h]	0.36[h]		0.41	0.72
Malian	0.54	0.55	0	0.58	0.52	0.56	0.53	0.53		0.44	0.72
Ugandan	0.46	0.45	0.58	0	0.49	0.50	0.45	0.48		0.36	0.46
Zambian	0.47	0.45	0.52[g]	0.49	0	0.42	0.48	0.48		0.44	0.66
Zulu	0.44	0.37[h]	0.56	0.50	0.42[g]	0	0.39	0.43		0.44	0.72
Cameroonian	0.34[h]	0.37	0.53[h]	0.45[h]	0.48	0.39[h]	0	0.39		0.40	0.68
Kenyans	0.33	0.36	0.53	0.48	0.48	0.43	0.39	0		0.46	0.74
AFAM	0.43	0.41	0.44	0.36	0.44	0.44	0.40	0.46		0	0.54
CAU	0.72	0.72	0.72	0.46	0.66	0.72	0.68	0.74		0.54	0
HLA-C											
Kenyan Nandi	0	0.16[g]	0.44	0.33[h]	0.41				0.42	0.29	0.44
Kenyan Luo	0.16[g]	0	0.41	0.31[g]	0.33[g]				0.41	0.21	0.42
Malian	0.44	0.41	0	0.42	0.35[h]				0.38[h]	0.33	0.48
Ugandan	0.33[h]	0.31[h]	0.42	0	0.40				0.38[g]	0.26	0.28
Zambian	0.41	0.33	0.35[g]	0.40	0				0.44	0.33	0.52
Mandenka	0.42	0.41	0.38[h]	0.38	0.44				0	0.36	0.47
AFAM	0.29	0.21	0.33	0.26	0.33				0.36	0	0.34
CAU	0.44	0.42	0.48	0.28	0.52				0.47	0.34	0

AFAM, African Americans. HLA, human leukocyte antigen.
[a]The data from this study.
[b,c,d,f]Genetic distances calculated from published data;
[e]Data from our previous study (63).
[g]The population from each column with the lowest distance.
[h]The second lowest distance between two populations.

Table 6

The diversification of allele families does not contribute significantly to the genetic differentiation between African populations

In contrast to HLA-A, in which we observed multiple HLA-A alleles of the same lineage with similar frequencies in the same population, HLA-B showed less examples of balanced diversity within the same group. Strikingly, only a few families of B-locus sequence-related alleles showed the occurrence of multiple alleles in the same population with balanced frequencies. Only the groups of alleles associated with the serotypes B70 (B*1503/B*1510), B63 (B*1516/B*1517), B57, and B58 presented more than one allele with intermediate frequencies simultaneously in the same

population (Table 4). Similarly, HLA-C showed only two allelic groups, Cw*03 and Cw*07, that presented multiple alleles with balanced frequencies in the same population (Table 5).

In order to evaluate whether genetic diversification by the generation of novel alleles made major contributions to the differentiation of the African populations, we collapsed the AFs from all the subtypes with first two identical digits into common groups (e.g., the AFs of the B*07 subtypes, B*0702, B*0705_G, etc. were added up and provided the estimation of AF of the B*07 group). Then we calculated the genetic distances between populations, utilizing both the allelic level data and broad grouping data. We observed that the genetic distances calculated with allele level data (Table 6) were not larger than 1.59 times of the genetic distances calculated with the broad grouping data. The lowest ratio was observed for the genetic distance between Zulus and Malians with B-locus data (B-locus allele level, $d = 0.56$; B-locus broad grouping level, $d = 0.51$; ratio 1.09), and the highest ratio was observed for the genetic distance between Ugandans and Nandi for HLA-C (C-locus allele level, $d = 0.33$; C-locus broad grouping data $d = 0.21$; ratio 1.59). The absolute contributions of alleles to genetic distances showed that the highest increase was observed for the B-locus between Cameroonians and Ugandans, in which the genetic distance calculated with broad grouping data ($d = 0.26$) raised by 0.19, when estimated by the distribution of allele-level types ($d = 0.45$; ratio 1.73). The increases between allelic level data and broad grouping data in genetic distances between African populations are not as remarkable as the ones observed between other populations (50–63). The genetic distances between American Indian populations increase remarkably from broad grouping level (serologic-equivalent) to allele-level resolution. For example, the genetic distance between the Wichi Indians from Argentina and the Mixteco Indians from Mexico was raised dramatically from 0.45 to 0.88 (ratio 1.96), when analyzing the distribution of HLA-B alleles and broad groups. The differences in the contributions of alleles and groups to genetic distances are likely to be related to the evolutionary histories of the HLA loci in different continents.

Haplotypes formed by alleles of the HLA-A, HLA-B, and HLA-C loci

A large number of HLA-A/B, HLA-B/C, HLA-A/C, and HLA-A/B/C haplotypes were observed in each of the African populations described in this study. There were examples of both, haplotypes that are identified in all populations and population-specific haplotypes. Most of the HLA-B/C haplotypes identified were observed in more than one population.

HLA-B/C haplotypes

Most of the associations between alleles of the B and C loci were strong; many pairs of HLA-B and HLA-C alleles had absolute (i.e., $D'_{ij} = 1$) or almost complete associations (Table 7). Significant LD ($P < 0.001$, $P < 0.01$, or $P < 0.05$) was observed for almost all HLA-B/C haplotypes with frequencies higher than 0.01 (Table 7).

The overall analysis of the associations between alleles of HLA-B and HLA-C suggests that in each African population, almost invariably, each allele of the B-locus has predominant or absolute associations with only one allele of the C-locus (Table 7). Some exceptions were observed, however, for B*5301 (associated with Cw*0401_G and Cw*0602 in Kenyans and Malians), B*4501 (associated with Cw*1601 and Cw*0602 in both Kenyan groups and Ugandans), B*1801_G (associated with Cw*0501_G, Cw*0704_G, Cw*0701_G, and Cw*1203 in Ugandans), B*5101_G (with variable associations with Cw*1601, Cw*0202, Cw*1602, and Cw*1402 in various populations), B*5801 (associated with Cw*0701_G and Cw*0302 in Kenyans and Ugandans), B*8101 (associated with Cw*0401_G, Cw*0804, and Cw*1801_G in Kenyans), and B*3501_G (associated with Cw*0401_G and Cw*1601 in Malians). Some B-locus alleles showed associations with different HLA-C alleles in different populations; for example, B*5703 was associated with Cw*1801_G in Zambians and with Cw*0701_G in the Luo population.

In this study, 72 B/C haplotypes were identified (Table 7). Only five B/C haplotypes were present in all the populations studied; the haplotypes common to all populations include B*1503-Cw*0202, B*1510-Cw*0304, B*4501-Cw*1601, B*5101_G -Cw*1601, and B*5301-Cw*0401_G (Table 7). Five additional B/C haplotypes were observed in four populations: B*0702-Cw*0702, B*1402-Cw*0802, B*3501_G-Cw*0401_G, B*4201-Cw*1701_G, and B*5801-Cw*0701_G (Table 7). These 10 widely distributed haplotypes include almost all the high-frequency haplotypes of all the African populations. The groups of the most frequent B/C haplotypes of all populations are completed with the inclusion of only a few additional population-specific haplotypes; these include B*5802-Cw*0602 (in Kenyans and Ugandans), B*0801-Cw*0701_G (in Ugandans), and several haplotypes including Cw*1601 (B*5201-Cw*1601 and B*7801-Cw*1601 in Malians). The 58 remaining sets of B/C haplotypes had intermediate or low frequencies (HF < 0.05). The most frequent B/C haplotypes identified in each population contained the most frequent HLA-B alleles of each population group.

Interestingly, we observed several high-frequency B/C haplotypes that contained the allele Cw*1601 in Malians; the increased frequency of several haplotypes containing this highly predominant HLA-C allele reinforces the hypothesis of the occurrence of directional selection for this allele in this population from West Africa.

Common HLA-B/C haplotype frequencies in five African populations

Haplotype		Kenyan Nandi ($n=240$)			Kenyan Luo ($n=265$)			Malians ($n=129$)			Ugandans ($n=161$)			Zambians ($n=43$)		
HLA-B	HLA-C	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P
0702	0702				0.0189	0.76	<0.001	0.0426[d]	0.91	<0.001	0.0559[c]	0.91	<0.001	0.0349[d]	0.77	<0.001
0801	0602	0.0104	0.01													
0801	0701	0.0146	0.20								0.0528[c]	0.82	<0.001	0.0233[d]	0.64	<0.05
0801	0702				0.0151	0.48	<0.001									
0801	0704	0.0167	0.32	<0.001	0.0151	0.47	<0.001									
1302	0602	0.0146	0.54	<0.05	0.0076	0.59	<0.05				0.0218[d]	0.75	<0.001			
1401	0202													0.0233[d]	0.46	
1401	0802							0.0116	1.00	<0.001	0.0155	1.00	<0.001			
1402	0802				0.0415[d]	0.95	<0.001	0.0116	1.00	<0.001	0.0373[d]	1.00	<0.001	0.0349[d]	1.00	<0.001
1501	0303										0.0124	0.48	<0.001			
1503	0202	0.0604[c]	0.74	<0.001	0.0755[c]	0.86	<0.001	0.0504[c]	0.92	<0.001	0.0559[c]	0.89	<0.001	0.0349[d]	0.57	<0.01
1503	0401	0.0188	0.14		0.0113	0.00										
1503	1202							0.0078	1.00	<0.001						
1510	0304	0.0146	0.87	<0.001	0.0359[d]	1.00	<0.001	0.0116	0.74	<0.001	0.0217[d]	0.87	<0.001	0.0349[d]	0.74	<0.001
1516	1402	0.0104	1.00	<0.001				0.0078	0.71	<0.001						
1517	1701				0.0075	0.53	<0.01									
1801	0501	0.0125	0.74	<0.001				0.0078	1.00	<0.001	0.0093	0.19	<0.05			
1801	0701				0.0076	0.06					0.0124	0.11				
1801	0704	0.0313[d]	0.63	<0.001	0.0283[d]	0.63	<0.001				0.0124	0.66	<0.001			
1801	1203										0.0124	0.30	<0.01			
1801	1801													0.0233[d]	0.65	<0.05
2703	0202	0.0250[d]	1.00	<0.001	0.0076	1.00	<0.001									
2705	0102										0.0093	0.74	<0.001			
3501	0401	0.0125	0.27	<0.05	0.0208[d]	0.55	<0.001	0.0773[c]	0.50	<0.001	0.0218[d]	0.52	<0.01			
3501	0704	0.0104	0.25	<0.01												
3501	0705				0.0076	1.00	<0.001									
3501	1601							0.0429[d]	0.08							
3502	0401										0.0155	1.00	<0.001			
3901	1203										0.0124	0.79	<0.001			
3910	1203	0.0104	1.00	<0.001	0.0094	0.83	<0.001	0.0078	1.00	<0.001						
4001	0304										0.0155	1.00	<0.001	0.0233[d]	1.00	<0.01
4012	0404	0.0125	1.00	<0.001	0.0076	1.00	<0.001									
4101	1701	0.0271[d]	0.85	<0.001	0.0094	0.81	<0.001									
4201	1701	0.0688[c]	1.00	<0.001	0.0736[c]	0.92	<0.001	0.1395[b]	1.00	<0.001				0.1512[a]	1.00	<0.001
4402	0501										0.0217[d]	0.62	<0.001			
4403	0303										0.0093	0.22	<0.05			
4403	0401										0.0249[d]	0.46	<0.01			
4403	0701													0.0233[d]	1.00	<0.01
4403	1203							0.0078	0.53	<0.001						
4403	1601										0.0093	0.16				
4501	0401				0.0076	−0.02										
4501	0602	0.0375[d]	0.57	<0.001	0.0302[d]	0.33	<0.01				0.0093	0.25				
4501	1601	0.0146	0.29	<0.001	0.0283[d]	0.60	<0.001	0.0540[c]	0.83	<0.001	0.0186	0.65	<0.001	0.0465[d]	1.00	<0.001
4701	0602	0.0125	1.00	<0.01												
4703	0701				0.0094	0.68	<0.001									

Table 7

Continued

Haplotype		Kenyan Nandi ($n=240$)			Kenyan Luo ($n=265$)			Malians ($n=129$)			Ugandans ($n=161$)			Zambians ($n=43$)		
HLA-B	HLA-C	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P
4901	0701	0.0458[d]	1.00	<0.001				0.0233[d]	0.91	<0.001	0.0280[d]	0.70	<0.001			
5001	0602							0.0078	0.70	<0.001						
5101	1402							0.0078	0.39	<0.001						
5101	0202										0.0124	0.17				
5101	1601	0.0271[d]	0.63	<0.001	0.0170	0.74	<0.001	0.0194	0.67	<0.001	0.0093	0.15		0.0349[d]	0.57	<0.01
5101	1602	0.0125	1.00	<0.001												
5201	1202											0.0093	1.00	<0.001		
5201	1601							0.0853[c]	1.00	<0.001						
5301	0304	0.0167	0.34	<0.001												
5301	0401	0.0458[d]	0.48	<0.001	0.0491[d]	0.68	<0.001	0.1047[b]	0.56	<0.001	0.0404[d]	0.78	<0.001	0.1047[b]	1.00	<0.001
5301	0602	0.0187	0.00		0.0170	0.08		0.0388[d]	0.66	<0.001						
5601	0102							0.0078	1.00	<0.001						
5701	0701				0.0076	1.00	<0.001									
5701	0602											0.0217[d]	0.66	<0.001		
5702	1801							0.0116	1.00	<0.001				0.0233[d]	1.00	<0.01
5703	0701	0.0250[d]	0.83	<0.001												
5703	1801													0.0581[c]	1.00	<0.001
5801	0302	0.0417[d]	0.90	<0.001	0.0132	0.37	<0.001					0.0093	0.49	<0.001		
5801	0602	0.0167	−0.06													
5801	0701	0.0417[d]	0.31	<0.001	0.0434[d]	0.59	<0.001	0.0194	0.89		0.0218[d]	0.45	<0.05			
5802	0602	0.0854[c]	1.00	<0.001	0.1170[b]	0.93	<0.001					0.0435[d]	1.00	<0.001		
7801	1601							0.0698[c]	1.00	<0.001						
8101	0401				0.0170	0.16										
8101	0704													0.0233[d]	1.00	<0.001
8101	0804	0.0146	1.00	<0.001	0.0076	1.00	<0.001									
8101	1801	0.0146	0.62	<0.001	0.0321[d]	0.63	<0.001									
8201	0302					0.0170	0.90	<0.001								

HLA, human leukocyte antigen. HLA-B/C haplotypes with frequencies >0.007 (Kenyan Nandi, Kenyan Luo, Malians, and Ugandans) and >0.02 (Zambians) are shown.
[a]HF > 0.15.
[b]0.10 < HF < 0.15.
[c]0.05 < HF < 0.10.
[d]0.02 < HF < 0.05.
[e]HFs were calculated by maximum likelihood methods (*Materials and methods*).

Table 7

There were several examples of B-alleles shared by both African and non-African populations. Some of these 'universally distributed' alleles form one or a few haplotypes that are also widely distributed in other populations from different continents including B*0801-Cw*0701$_G$, B*1801$_G$-Cw*0501$_G$, B*1801$_G$-Cw*0701$_G$, B*1801$_G$- Cw*1203, B*5301-Cw*0401$_G$, B*3501$_G$ -Cw*0401$_G$ and B*5801-Cw*0302. In addition to these 'universal' or 'out of Africa' haplotypes, the same B-locus alleles form haplotypes that are only found in African populations (B*0801$_G$-Cw*0704$_G$ in the Kenyans; B*3501$_G$-Cw*1601 in Malians); and the haplotypes B*1801$_G$-Cw*0704$_G$, B*5301-Cw*0602, and B*5801-Cw*0701$_G$ are found in several African populations reported in this study.

HLA-A/B haplotypes

The estimated HFs and the corresponding D'_{ij} values of the most frequent pairs of alleles at the HLA-A and HLA-B loci in each population are shown in Table 8. The strength of the associations between pairs of alleles of the HLA-A and HLA-B loci (Table 8) measured by D'_{ij} was weaker than those observed for pairs of alleles of the HLA-B and

Common HLA-A/B haplotype frequencies in five African populations

Haplotype		Kenyan Nandi ($n=240$)			Kenyan Luo ($n=265$)			Malians ($n=138$)			Ugandans ($n=161$)			Zambians ($n=43$)		
HLA-B	HLA-C	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P
0101	0801	0.0138	0.21		0.0170	0.53	<0.001				0.0373[d]	0.56	<0.001			
0101	1516	0.0104	1.00	<0.001												
0101	4402										0.0093	0.21				
0101	4501	0.0104	0.08													
0101	5301	0.0168	0.09													
0101	5701										0.0093	0.24				
0101	5801	0.0111	−0.01													
0101	8101	0.0167	0.34	<0.01	0.0170	0.21	<0.01									
0102	5801	0.0167	0.58	<0.001												
0201	0702										0.0114	0.00				
0201	0801	0.0178	0.31	<0.01							0.0124	0.02				
0201	1402										0.0155	0.29				
0201	1503	0.0114	0.03		0.0336[d]	0.31	<0.001				0.0124	0.02				
0201	1510													0.0233[d]	0.44	
0201	1801	0.0150	0.22	<0.05							0.0124	0.08				
0201	2705										0.0093	0.69	<0.05			
0201	3501							0.0217[d]	0.15							
0201	4101	0.0104	0.24													
0201	4201				0.0139	0.07										
0201	4402										0.0217[d]	0.55	<0.01			
0201	4501	0.0120	0.11		0.0168	0.16		0.0318[d]	0.47	<0.001	0.0186	0.59	<0.05	0.0233[d]	0.44	
0201	5101				0.0132	0.53	<0.001				0.0129	0.09		0.0233[d]	0.32	
0201	5301				0.0088	0.02					0.0124	0.08				
0201	5701										0.0151	0.37				
0201	5703	0.0104	0.27													
0201	5801				0.0115	0.06										
0202	1503	0.0250	0.32	<0.001												
0202	5301							0.0432[d]	0.48	<0.001						
0202	5802	0.0271	0.35	<0.001	0.0151	0.43	<0.01				0.0186	0.53	<0.001			
0202	7801							0.0109	0.23							
0205	1801	0.0161	0.27	<0.01	0.0113	0.37	<0.001									
0205	5101	0.0228	0.50	<0.001												
0205	5801	0.0168	0.10		0.0130	0.42	<0.001				0.0186	0.86	<0.001			
0301	4901							0.0181	0.70	<0.001						
0301	5802	0.0104	0.27	<0.05	0.0144	0.32	<0.05									
1101	5101										0.0182	0.39	<0.001			
1101	5501										0.0093	0.74	<0.001			
2301	0702							0.0206[d]	0.17							
2301	1401													0.0233[d]	0.47	<0.05
2301	1503	0.0113	0.09		0.0124	0.06		0.0154	−0.01							
2301	1510							0.0109	0.24							
2301	2703	0.0142	0.53	<0.001												
2301	3501							0.0602[c]	0.32	<0.05						
2301	4201				0.0086	0.02		0.0118	−0.18							
2301	4403										0.0248[d]	0.50	<0.001			

Table 8

Continued

Haplotype		Kenyan Nandi ($n=240$)			Kenyan Luo ($n=265$)			Malians ($n=138$)			Ugandans ($n=161$)			Zambians ($n=43$)		
HLA-B	HLA-C	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P	HF[e]	D'	P
2301	5101							0.0145	0.44							
2301	5201							0.0401[d]	0.33	<0.05						
2301	5301	0.0108	0.07					0.0393	0.02							
2301	5801				0.0113	0.09										
2301	5802				0.0116	0.01										
2902	3501							0.0109	0.57							
2902	4201	0.0245[d]	0.45	<0.001	0.0264[d]	0.46	<0.001	0.0100	0.20					0.0349[d]	0.53	<0.05
2902	4403										0.0093	0.27	<0.01	0.0233[d]	1.00	<0.001
2902	4501	0.0109	0.17	<0.05	0.0113	0.16	<0.05									
3001	1302										0.0140	0.48	<0.001			
3001	4201	0.0292[d]	0.58	<0.001	0.0226[d]	0.30	<0.001	0.1044[b]	0.72	<0.001				0.1163[b]	0.80	<0.001
3001	5301							0.0113	−0.09							
3001	5802				0.0109	0.05										
3002	0702							0.0109	0.12	<0.05						
3002	1801				0.0189	0.40	<0.001				0.0093	0.19	<0.01			
3002	4501				0.0109	0.13										
3002	5301	0.0166	0.39	<0.001										0.0581[c]	0.44	
3002	5703													0.0465[d]	0.76	<0.05
3002	8101													0.0233[d]	1.00	
3101	4001													0.0233[d]	1.02	<0.001
3201	8101	0.0104	0.54	<0.001												
3301	7801							0.0109	0.35	<0.01						
3303	1516							0.0109	0.09	<0.001						
3303	3501							0.0089	−0.04							
3303	5201							0.0095	0.02							
3303	7801							0.0290[d]	0.36	<0.001						
3402	3501	0.0199	0.54	<0.001												
3402	4901	0.0153	0.30	<0.001												
3402	5301							0.0254[d]	0.64	<0.01						
3601	5301	0.0125	1.00	<0.001	0.0188	0.60	<0.001				0.0093	0.75	<0.001	0.0233[d]	0.63	<0.05
6601	4501	0.0108	0.17	<0.05												
6601	5802	0.0288[d]	0.54	<0.001	0.0491[d]	0.68	<0.001				0.0186	0.66	<0.001			
6801	5101										0.0124	0.42	<0.001			
6802	0702				0.0113	0.42	<0.001				0.0124	0.19	<0.05	0.0233[d]	0.49	<0.01
6802	1302	0.0104	0.38	<0.05												
6802	1510	0.0146	0.86	<0.001	0.0128	0.31	<0.01				0.0093	0.34	<0.05			
6802	4901										0.0124	0.30	<0.05			
6802	5301	0.0149	0.07		0.0164	0.18	<0.01	0.0179	0.16							
6802	5801	0.0193	0.09													
6802	7801							0.0109	1.00							
7401	0702										0.0093	0.12				
7401	1402				0.0151	0.29	<0.001									
7401	1503				0.0211[d]	0.20	<0.01	0.0126	0.30	<0.01	0.0280[d]	0.51	<0.001	0.0233[d]	1.00	<0.001
7401	3501				0.0104	0.25	<0.05									

Table 8

Continued

Haplotype		Kenyan Nandi (n = 240)			Kenyan Luo (n = 265)			Malians (n = 138)			Ugandans (n = 161)			Zambians (n = 43)		
HLA-B	HLA-C	HF[e]	*D′*	P	HF[e]	*D′*	P	HF[e]	*D′*	P	HF[e]	*D′*	P	HF[e]	*D′*	P
7401	5201							0.0091	0.18							
7401	5802				0.0103	0.01										
7403	8101				0.0113	0.47	<0.001									
7403	8201				0.0113	0.59	<0.001									
8001	5702													0.0233[d]	1.00	<0.001

HLA, human leukocyte antigen; *D′*, relative normalized Δ-value. HLA-A/B haplotypes with frequencies >0.008 (Kenyan Nandi, Kenyan Luo, Malians, and Ugandans) and >0.02 (Zambians) are shown.
[a]HF > 0.15.
[b]0.10 < HF < 0.15.
[c]0.05 < HF < 0.10.
[d]0.02 < HF < 0.05.
[e]HFs were calculated by maximum likelihood methods (*Materials and methods*).

Table 8

HLA-C loci (Table 7). A more accurate comparison of the strengths of associations between alleles of different loci in various populations was made by measuring APRDV (Table 9). The comparisons between APRDV show that, as expected, the associations between alleles of HLA-A and HLA-B loci are invariably weaker than the B/C associations in each of the populations analyzed here.

Table 8 illustrates that the degree of haplotypic diversity varied from population to population. The most predominant A/B haplotype (A*3001-B*4201) in both Malians and Zambians was observed with frequencies above 0.10; in contrast, the most frequent haplotypes A*3001-B*4201 in Kenyan Nandi, A*6601-B*5802 in Kenyan Luo, and A*0101_G-B*0801 in Ugandans had frequencies below 0.05. Similarly, the distribution of haplotypes analyzed as cumulative frequencies showed distinct population coverage in different populations. The cumulative frequencies of the 15 most frequent haplotypes in both Malians and Zambians accounted for approximately 50% of the haplotype pool, while in the population from Uganda and from both Kenyan groups, the same number of haplotypes accounted for a much smaller proportion of the haplotype pool (approximately 30% of all estimated A/B haplotypes in each population).

When we evaluated the occurrence of haplotypes in multiple populations, we identified a group of 15 HLA-A/B haplotypes that were present in at least three populations of this study. The haplotype A*0201_G-B*4501 was identified in all populations. The haplotypes A*2902-B*4201, A*3001-B*4201, A*3601-B*5301, and A*7401_G-B*1503 were identified in four populations; 10 additional haplotypes were identified in three populations (A*0101_G-B*0801, A*0201_G-B*1503, A*0201_G-B*5101_G, A*0202-B*5802, A*0205-B*5801, A*2301_G-B*1503, A*6601-B*5802, A*6802-B*0702, A*6802-B*1510, and A*6802-B*5301). The combined frequency of these 15 haplotypes in the African continent accounts for as high as 0.30 in the Luo population from Kenya and for less than 0.20 in Ugandans and Malians. The finding of a higher prevalence of these ubiquitous African haplotypes in Kenyan Luo is in line with the observation that

Global linkage disequilibrium[a] between pairs of loci

		A/B			A/C			B/C		
Ethnic group	Subjects	ARPDV	*D′*	W_n	ARPDV	*D′*	W_n	ARPDV	*D′*	W_n
Caucasians[b]	287	0.27	0.58	0.48	0.20	0.49	0.41	0.79	0.93	0.76
African Americans[b]	251	0.24	0.66	0.43	0.17	0.54	0.33	0.58	0.88	0.73
Kenyan Nandi	240	0.33	0.69	0.46	0.28	0.61	0.40	0.67	0.87	0.79
Kenyan Luo	265	0.29	0.69	0.48	0.25	0.61	0.38	0.67	0.89	0.74
Malians	129	0.36	0.66	0.49	0.31	0.58	0.40	0.74	0.88	0.77
Ugandans	161	0.34	0.77	0.56	0.24	0.62	0.39	0.62	0.90	0.72
Zambians	43	0.61	0.74	0.74	0.48	0.77	0.66	0.80	0.97	0.92

APRDV, average of the positive relative Δvalue; *D′*, relative normalized Δ-value; W_n, Cramer's V statistic.
[a]Global linkage disequilibrium was estimated by three different approaches. These include APRDV, *D′*, and W_n (*Materials and methods*).
[b]Data extracted from a previous study of populations living in the US (63).

Table 9

the Kenyan Luo population has the closest genetic distances with each of the African populations, as measured by differences in the distribution of alleles at various class I loci individually.

We observed that the most frequent haplotypes of each population contained the most frequent HLA-A or HLA-B alleles identified in each of them. Interestingly, we observed that the most frequent alleles had multiple associations with alleles of the other locus. The allele A*3002, the most predominant HLA-A allele in Zambians, was found forming several frequent haplotypes in this population with the alleles B*5301, B*5703, and B*1801_G. In Ugandans, we observed the haplotype A*3002-B*1801_G, which seems to be frequent also in Southern European and Mediterranean populations (78–81). Similarly, B*5802, the most frequent HLA-B allele in the Kenyan Luo, and A*2301_G, the most frequent allele in Malians, were observed forming haplotypes with multiple alleles of the other loci. These multiple haplotypes were present with intermediate frequencies, and their occurrence correlates well with the lower D'_{ij} values observed for the A/B haplotypes.

HLA-A/C haplotypes

The associations between HLA-A and HLA-C alleles were weaker than those observed for B/C haplotypes (data not shown). In general, the A/C associations reflected the A/B and the B/C associations described above.

Global linkage disequilibria between pairs of HLA class I loci

All three pairs of overall pairwise LD were highly significant in all populations ($P < 0.001$). The strength of the association between alleles at two loci was estimated using APRDV, D', and W_n for each pair of loci. Even though these measurements have different magnitudes, the pairs of loci and populations rank in similar positions for each of the three measurements. The values measuring the strength of global LD for the A/B, B/C, and A/C pairs are shown in Table 9. This table includes the estimations for the five sub-Saharan populations reported in the present study and those of Caucasians from US and African Americans included in a previous report (63). The latter populations were included as a reference. The comparison of LD measurements made for different pairs of loci clearly indicates that the B/C associations are always stronger than the A/B and A/C associations. W_n and APRDV present wider ranges of variation for the values found for each pair of loci and, therefore, appear to be more discriminative for the evaluation of strength of global LD between pairs of loci in different populations.

The Zambian population, perhaps, because of the limited sample size, showed the highest value for each pair of loci by virtually all the measurements. All the comparisons between the same pair of loci had similar values in different populations. The strength of global LD measured by either W_n or APRDV, for each of the class I pairs of loci, places the African Americans as the population with lower values. However, this is not the case for the D'_{ij} measure. The weaker associations observed in the latter population may result from the contribution of multiple haplotypes from several populations. In contrast, all the populations living in this continent show global LD values (D'_{ij}, APRDV and W_n) that are in the same ranges as those observed in Caucasians.

Haplotypes comprising alleles of three class I loci (A/B/C haplotypes)

The patterns of population distribution of haplotypes defined by alleles at the three class I loci (A, B, and C), in each population, are similar to those observed for the haplotypes defined by alleles at the HLA-A and HLA-B loci. The haplotypes formed by alleles of the HLA-A, HLA-B, and HLA-C loci estimated to be present in at least two subjects of one population are shown in Table 10.

The Malian and Zambian populations presented a single prevalent haplotype. The haplotype A*3001-B*4201-Cw*1701_G was widely distributed in these populations (HF = 0.1044 and 0.1163, respectively) (Table 10).

In the Luo population, the most frequent haplotype, A*6601-B*5802-Cw*0602, was present with relatively high frequency (HF = 0.0509; Table 10).

The most frequent haplotype in Ugandans (A*0101-B*0801-Cw*0701, HF = 0.0342) and the most frequent haplotypes in Nandi (A*3001-B*4201-Cw*1701, HF = 0.0292; and A*6601-B*5802-Cw*0602, HF = 0.0292) had much lower frequencies.

The distribution of haplotypes formed by alleles at three class I loci (HLA-A, HLA-B, and HLA-C) are similar to the ones observed for haplotypes defined by alleles at the HLA-A and HLA-B loci. For example, the cumulative frequencies of the 15 most frequent haplotypes defined by alleles at three loci (HLA-A, HLA-B, and HLA-C) do not differ significantly from the frequencies of haplotypes defined by alleles at the HLA-A and HLA-B loci. This is probably due to high LD between HLA-B and HLA-C loci.

In the Nandi population, the cumulative frequency of the 15 most frequent A/B/C haplotypes (A/B/C-CF_{15}) is 0.298, while the A/B-CF_{15} is 0.314; and in the Luo population, the A/B/C-CF_{15} is 0.303 and the A/B-CF_{15} is 0.318. In Malians, the A/B/C-CF_{15} is 0.455 and the A/B-CF_{15} is 0.494. In Ugandans, the A/B/C-CF_{15} is 0.258 and the A/B-CF_{15} is 0.287, while in Zambians, the A/B/C-CF_{15} is identical to the A/B-CF_{15} (0.512). We plotted the cumulative frequencies of alleles and haplotypes for different populations. Figure 2 shows the distribution of the cumulative frequencies of the class I alleles and haplotypes, which are observed for the Kenyan Luo population; it can be

Common HLA-A/B/C haplotype frequencies in five African populations

Haplotype			Haplotype frequency[e]				
HLA-A	HLA-B	HLA-C	Kenyan Nandi ($n=240$)	Kenyan Luo ($n=265$)	Malians ($n=129$)	Ugandans ($n=161$)	Zambians ($n=43$)
0101	0801	0701				0.034[d]	
0101	0801	0704	0.014	0.015			
0101	1516	1402	0.010				
0101	5301	0401	0.015				
0101	5701	0602				0.012	
0101	5801	0302	0.017				
0101	8101	1801	0.013	0.017			
0102	5801	0302	0.015				
0201	0801	0602	0.010				
0201	0801	0701				0.012	
0201	1402	0802				0.016	
0201	1501	0303				0.009	
0201	1503	0202	0.011	0.034[d]			
0201	1510	0304					0.023[d]
0201	1801	0701				0.012	
0201	1801	0704	0.015				
0201	3501	0401			0.012		
0201	4101	1701	0.010				
0201	4201	1701		0.014			
0201	4402	0501				0.019	
0201	4501	1601	0.010	0.013	0.035[d]	0.019	0.023[d]
0201	5101	0202				0.012	
0201	5101	1601		0.009			0.023[d]
0201	5301	0401		0.009		0.012	
0201	5703	0701	0.010				
0202	1503	0202	0.025[d]				
0202	5301	0401			0.030[d]		
0202	5802	0602	0.027[d]	0.013		0.019	
0205	1801	0704	0.012	0.011			
0205	5101	1601	0.023[d]				
0205	5801	0701	0.015	0.013		0.009	
0301	1503	0202			0.012		
0301	4901	0701			0.019		
0301	5802	0602	0.010	0.013			
1101	5201	1202				0.009	
2301	0702	0702			0.024[d]		
2301	1401	0202					0.023[d]
2301	1503	0202	0.012	0.013	0.011	0.009	
2301	2703	0202	0.015				
2301	3501	0401			0.034[d]		
2301	3501	1601			0.016		
2301	4201	1701		0.009	0.011		

Table 10

Continued

Haplotype			Haplotype frequency[e]				
HLA-A	HLA-B	HLA-C	Kenyan Nandi ($n=240$)	Kenyan Luo ($n=265$)	Malians ($n=129$)	Ugandans ($n=161$)	Zambians ($n=43$)
2301	4403	0303				0.009	
2301	4403	0401				0.016	
2301	5101	1601			0.012		
2301	5201	1601			0.050[c]		
2301	5301	0602			0.031[d]		
2301	5802	0602		0.010			
2902	1402	0802				0.009	
2902	4201	1701	0.025[d]	0.026[d]	0.012		0.035[d]
2902	4403	0701					0.023[d]
2902	4501	0602	0.010	0.015			
3001	1302	0602				0.016	
3001	4201	1701	0.029[d]	0.0207[d]	0.104[b]		0.116[b]
3001	5802	0602		0.010			
3002	0702	0702				0.009	
3002	1801	0704		0.017			
3002	4501	1601		0.009			
3002	5301	0304	0.017				
3002	5301	0401					0.058[c]
3002	5703	1801					0.047[d]
3002	7801	1601			0.012		
3002	8101	0704					0.023[d]
3101	4001	0304					0.023[d]
3201	8101	0804	0.010				
3301	7801	1601			0.012		
3303	3501	0401			0.008		
3303	5301	0401			0.012		
3303	7801	1601			0.027[d]		
3402	3501	0401	0.010				
3402	3501	0704	0.010				
3402	4901	0701	0.015				
3402	5301	0401			0.019		
3601	5301	0401	0.013	0.019		0.009	0.023[d]
6601	4501	0602	0.010				
6601	5802	0602	0.029[d]	0.051[c]		0.019	
6802	0702	0702		0.013		0.015	0.023[d]
6802	1302	0602	0.010				
6802	1510	0304	0.015	0.013		0.012	
6802	5301	0401	0.010		0.027[d]		
6802	5301	0602		0.011			
6802	5801	0701	0.018				
6802	7801	1601			0.012		

Table 10

Continued

Haplotype			Haplotype frequency[e]				
HLA-A	HLA-B	HLA-C	Kenyan Nandi ($n=240$)	Kenyan Luo ($n=265$)	Malians ($n=129$)	Ugandans ($n=161$)	Zambians ($n=43$)
7401	1402	0802		0.015			
7401	1503	0202		0.019	0.015	0.0249[d]	0.023[d]
7401	3501	0401		0.011			
7401	4901	0701				0.009	
7401	5802	0602		0.010			
8001	5702	1801					0.023[d]

HLA, human leukocyte antigen.
[a]HF > 0.15;
[b]0.10 < HF < 0.15;
[c]0.05 < HF < 0.10;
[d]0.02 < HF < 0.05.
[e]HFs were calculated by maximum likelihood methods (*Materials and methods*).
HLA-A/B/C haplotypes with frequencies >0.008 (Kenyan Nandi, Kenyan Luo, Malians, and Ugandans) and >0.02 (Zambians) were shown.

Table 10

observed that the curves for haplotypes defined by A and B alleles virtually overlap with the curve of the cumulative frequencies of haplotypes defined by alleles at the three class I loci. This clearly indicates that the addition of HLA-C typing does not significantly increase the haplotypic diversity in this population. We observed similar levels of coverage by A/B/C and A/B haplotypes in the Kenyan Nandi, Zambians, and Malians (data not shown). In contrast, HLA-C in Ugandans seems to provide an additional and more significant level of haplotypic diversity with a population coverage, for the 15 most frequent A/B/C haplotypes, that is approximately one tenth (1/10) lower than the frequency obtained for haplotypes defined by A and B alleles. Among the populations described here, the Ugandans presented the largest haplotypic diversity with lowest population coverage, when comparing the cumulative frequencies of haplotypes defined by alleles at two and three class I loci.

Table 10 presents the haplotypes defined by alleles at the HLA-A, HLA-B, and HLA-C loci found in each of the populations. These haplotypes include associations that contained the most common two-locus haplotypes defined by alleles at the A and B loci shown in Table 8. Only one haplotype (A*0201-B*4501-Cw*1601) is present in all populations. The haplotypes containing alleles A*7401$_G$, A*3601, B*1503, and B*4201 are only found in populations of the African continent or those with recent African ancestry (e.g., African Americans [63]).

Interestingly, we observed four three-locus haplotypes, whose distributions correlate with the geographic location of the populations. These were A*0202-B*5802-Cw*0602, A*0205-B*5801-Cw*0701$_G$, A*6601-B*5802-Cw*0602, and A*6802-B*1510-Cw*0304, which were shared by Ugandans and Kenyans and were absent in the other African populations. A similar relationship between geographic proximity and sharing haplotypes was indicated by the presence of

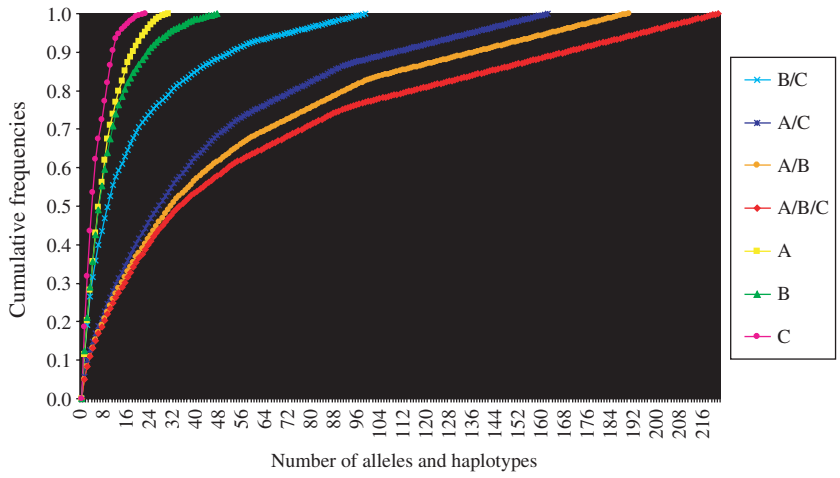

Fig. 2. rowsep="1" **Population coverage by human leukocyte antigen (HLA)-A, HLA-B, HLA-C alleles, and haplotypes in the Luo population from western Kenya.** Cumulative frequencies of alleles and haplotypes were sorted from the most frequent to the least frequent and were plotted against the number of alleles and haplotypes, which account for each cumulative frequency.

at least seven additional haplotypes, in both Kenyan populations. There are few HLA-A/B/C haplotypes widely distributed in the African populations; in fact, with the exception of only five haplotypes containing B*0702, B*4201, and B*5301, no other HLA-A/B/C haplotypes were observed in multiple populations. This observation suggests that there is a large haplotypic variation in all African populations. It could be postulated that the haplotypes found in multiple populations are the only ones that indicate the ancestral relationships between populations of the African continent.

The haplotypes A*0101$_G$-B*0801-Cw*0701$_G$, A*0201$_G$-B*4402$_G$-Cw*0501$_G$, A*3001-B*1302-Cw*0602, A*0201$_G$-B*0801-Cw*0701$_G$, A*0201$_G$-B*1801$_G$-Cw*0701$_G$, and A*0201$_G$-B*1501-Cw*0303 were only found in Ugandans and had frequencies above 0.01. These haplotypes are frequently found in Caucasians (53, 63). The low values for the estimated genetic distances between Ugandans and Caucasians, and the sharing of extended haplotypes between these populations indicate a closer relationship between Ugandans and Caucasians than between the latter and the other sub- Saharan populations described here and in other studies (34–36).

Different selective events may have contributed to the composition of HLA class I haplotypes found in Africa, and perhaps, two different types of associations may reflect these events. It is worth mentioning that some alleles of the same lineage or family (A*3001/A*3002, B*5701/B*5702/B*5703, and B*5801/B*5802) display distinct associations with alleles at other class I loci. In contrast, some alleles of the same family (the B5 CREG family in Malians, B*5101$_G$/B*5201/B*7801/B*3501$_G$) were found in association with the same alleles of the other class I loci (Cw*1601 and A*2301$_G$).

Discussion

In the present study, we examined the distribution of alleles of the class I loci defined by the application of molecular high-resolution testing methods. The use of these procedures allowed us to characterize and make distinctions between alleles having slight but highly informative structural and functional variation. These differences may go undetected, if either serologic- or intermediate-resolution molecular methods are applied.

We observed many alleles of HLA-A, HLA-B, and HLA-C loci, which are only found in the African continent; these alleles, therefore, may have arisen in Africa either after the 'migrations out of Africa' or were not transmitted to the non-African populations. Several of these African-specific alleles have intermediate or high AFs in various populations, and interestingly, these are the ones present in multiple African populations (A*0202, A*3402, A*3601, A*7401; B*1503, B*4201, B*5703, and B*5802). Therefore, these 'African widespread' alleles may be relatively old and their origins may predate the separation of linguistic groups within Africa. This hypothesis is reinforced by the observation that some of these alleles are found forming sets of haplotypes with the same alleles at other loci in various populations. Examples of widespread African-specific haplotypes are the three-locus haplotypes, such as A*7401-B*1503-Cw*020204, A*6601-B*5802-Cw*0602, A*0202-B*5802-Cw*0602, A*3001-B*4201-Cw*1701$_G$, and A*2902-B*4201-Cw*1701$_G$. The latter observation appears to reject the hypothesis that the location of the African-specific alleles in different areas of the African continent was the result of convergent evolution.

The present and other studies (47, 51, 52) indicate that a significant number of alleles of the HLA-A (20 alleles) and HLA-B (34 alleles) loci were found restricted to only one or two African populations. We observed, however, that virtually all these 'private' alleles had low frequencies. It can be postulated that several of these alleles may have been generated recently and may have originated by mutations in the most common alleles. This observation contrasts with the finding of 'novel' HLA class I alleles with significantly high frequency in indigenous populations of the American continent (54–57). Contrasting with the populations from Africa, the native populations from America present only a few allelic lineages of both HLA class I (54–57, 82–85) and class II loci (82, 83, 86–90). In spite of the reduced number of lineages, the indigenous tribes from South and Meso-America display a rich variation in alleles of these few lineages, and some of the novel alleles are predominant. It has been proposed that rapid evolution of MHC class I loci has taken place in extended geographic areas of the American continent (54, 84, 85, 91) with a trend for 'allele turnover,' in which the new alleles not only supplement the founder alleles but in some populations almost completely supplant the older alleles. This dynamic process of allele replacement seems to have minor impact in the modern populations from Africa, because virtually none of the putatively new HLA class I alleles contributes significantly to the gene pool of each individual population. It has been speculated that because the native American populations went through population bottlenecks, and only a few alleles entered or were maintained in the American continent, selective forces may have favored the increase in frequency and the maintenance of the newly generated alleles, because these alleles enlarged the peptide-binding repertoire of the HLA molecules of the individuals of these populations. It appears that in the African populations, a large allelic pool has been permanently available with all lineages present. Therefore, selection for the newly generated alleles may not have operated, because there were no diversity constraints in the African continent. Even though evolutionary bottlenecks may have occurred in a few regions of this continent, a

genetic pool of significant population size was continuously preserved after the initial population expansion took place in the region.

The analysis of alleles present in multiple African populations suggests that the most important diversification within allelic lineages occurred before the separation of linguistic groups. Supporting this hypothesis is the observation that several evolutionary groups of alleles of both HLA-A and HLA-B present multiple alleles that are widespread in the majority of the African populations. Many pairs of alleles included in the same lineage differ from each other by one or two stretches of DNA. Mechanisms involving one or two gene conversion or mutational events may be invoked for the generation of new alleles (18, 66–68). In HLA-A, the alleles A*0101$_G$, A*0102, and A*3601 are included in the evolutionary group A1/36 and have been observed with significant frequencies in several African populations. A*3601 differs from A*010101 by only four nucleotide substitutions in a DNA stretch, spanning codons 163–167 of exon-3. The nucleotide sequence found in A*3601, in this DNA segment, is also found in A*3001 (this nucleotide sequence segment is also present in several other alleles of HLA-A), and a gene conversion event may be invoked for the generation of A*3601 (16, 66–68). At the protein-structure level, A*3601 and A*0101 differ by just three amino acid replacements at residues 163, 166, and 167. Analysis of the three-dimensional structure shows that residue 9 of the HLA class I molecule points toward the substructure of the peptide-binding pockets B and C, while residues 163, 166, and 167 are located in pocket A (17). The distinguishing amino acid substitutions between A*0101 and 3601 are likely to determine some distinctive peptide-binding characteristics for each of the molecules encoded by these alleles (92).

A gene conversion event may also be invoked for the generation of A*0102 which may have arisen as well from A*0101. Similar to the A1/36 family, the sequence-related alleles, A*0201$_G$, A*0202, A*0205, A*0214, and A*0225, integrate the A2 group and are also present in multiple groups. Other examples of multiple alleles widely distributed are found in the A10 family with two or more of the alleles, A*2501, A*2601, A*3402, and A*6601, observed in various sub-Saharan populations. Other examples can be observed for the A30 group (A*3001, A*3002, and A*3004), the A31/33 group (A*3101, A*3103, A*3104, A*3301, and A*3303), and the A32/74 group (A*3201 and A*7401$_G$). It should be noted that many of these alleles differ from each other by a small number of amino acid substitutions; invariably, these substitutions are located in residues that are thought to affect the antigen recognition site and, therefore, likely to determine distinct peptide-binding repertoires for each of these alleles.

HLA-B presents several alleles that are also found exclusively in Africans, and these are present in several sub-Saharan populations. These alleles include the highly widespread and high-frequency alleles, such as B*1503, B*4201, B*5802, and B*5703; the intermediate frequency alleles, such as B*1510, B*1516, B*5702, B*8101, and B*7801; and the alleles found in several populations with relatively low frequency like B*3910 and B*8201. Most of the B-locus alleles can be grouped into evolutionary families on the basis of sequence similarity (18, 66–68). These include the subfamilies of the B15 group that includes the B70-related alleles, B*1503 and B*1510; the B63-related alleles, B*1516 and B*1517; the B17 family that includes B*5701, B*5702, B*5703, B*5801, and B*5802; and the B5 family that includes B*5101$_G$, B*5201, and B*7801. Interestingly, other alleles such as B*4201, B*8101, and B*8201 cannot be related to a defined allelic family but appear to present rather complex sequences, and perhaps, these alleles arose through more complex mechanisms (18, 66–68). The nucleotide sequence of B*4201 is identical to B*0801, with the exception of the sequences spanning exon 2 in which B*4201 is identical to B*070201. It has been postulated that this hybrid allele arose as a fusion between the alleles B*070201 and B*0801, or as an exon shuffling in which the second exon of B*0702 was inserted in a backbone structure provided by B*0801 (18).

The allele B*4201 seems to be predominant in most of the populations described here, with the exception of the population of Uganda. It appears that B*4201 may have replaced partially or completely its putative ancestor alleles, B*070201 and B*0801, suggesting that this allele was positively selected in Africa. It could be speculated that selection may have been driven by its unique peptide-binding characteristics. Interestingly, B*4201 does not seem to relate to either B*070201 or B*0801 in its haplotypic association, because this B-locus allele is associated with A*3001 and Cw*1701$_G$; neither B*0801 nor B*0702 are associated with these alleles. In fact, B*4201 appears to have unique associations with alleles of the other loci, such as MICA, complotype FC, and tumor necrosis factor-α in the HLA region (93, 94).

The pair of sequence-related alleles B*5801 and B*5802 provides an example in which both have been maintained in most populations. These alleles differ by substitutions at residues 94, 95, and 97, and it has been speculated that B*5802 arose from a gene conversion between B*1402 (or B*1401, donor) and B*5801 (recipient) (95). The resulting recombinant gives B*5802 a dual peptide-binding specificity, and one subset of the peptides bound by B*5802 possesses the typical B17-binding motif, while a second subset displays an additional anchor (arginine) at position 6 (95). The peptides, eluted from B*5802, possessing the arginine 6 anchor are shorter (7 or 8 amino acids long) and lack the typical carboxyl terminal anchor of phenylalanine or tryptophan found predominantly in the nonamers eluted from B*5802 and the other alleles members of the B17 family (95). In light of this observation, it could be argued that this specialized

binding capacity may provide (or has provided) the subjects carrying B*5802 with an ability to bind a particular peptide and mount an effective response to a yet-to-be-identified parasite. The subtle amino acid sequence differences between alleles may, however, provide a significant selective advantage to fight infections effectively. Parasite driven selection does not necessarily lead the advantageous allele to rise to very high frequency, because in an immunologically diverse environment, multiple alleles are necessary to provide effective coverage to respond to multiple infectious challenges. In this particular example, B*5801 could have been maintained in the population, because of a particular advantage provided by its unique sequence.

Interestingly, we have observed that both related alleles, B*5801 and B*5802, have distinct associations with alleles of the other class I loci; B*5801 is found in association with Cw*0302 in some African and in non-African populations (53, 58–60, 63) or Cw*0701_G in several African populations described in this study; in contrast, B*5802 is only found in African populations and in strong association with Cw*0602. The extended haplotypes, A*0205-B*5801-Cw*0701_G, A*0202-B*5802-Cw*0602, and A*6601-B*5801-Cw*0602, were observed in both Kenyan populations, in the Ugandans described in the present study and in the study of African Americans (63), suggesting that both the allele and the haplotypic differentiation may predate the separation of African populations. This observation seems to correspond to a more generalized observation, in which sequence-related alleles that have intermediate or high frequencies in the same population appear to have distinctive associations with alleles of other class I loci (63).

We hypothesize that the different LD patterns of sequence-related alleles are related to the peptide-binding differences that are found between them. It could be argued that the sequence-related alleles, because of their similarity, can bind a discrete and overlapping set of peptides, and because of their differences, could have the capacity to bind some peptides but not others. We propose that the similar intermediate/high frequency of sequence-related alleles in the same population is the result of each one of them having provided selective advantages, because of their abilities to present peptides from different pathogens. Also, it could be argued that the amino acid substitutions that provide an allele with the ability to present some peptides may also result in the loss of the capability of binding other peptides. The latter deficits could be compensated by the presence of other alleles of the same locus in heterozygous individuals or by the alleles of other loci, because the molecules encoded by different class I loci have an overlapping function to selectively bind peptides. Therefore, the maintenance of several relatively conserved haplotypes in the same population may have been favored from the complementary/compensatory peptide-binding abilities of alleles from different loci carried on the same chromosome. Furthermore, extended haplotypes may in some aspects be considered as functional units that include alleles that combine to provide a peptide-binding repertoire that ensures effective responses to multiple common pathogens found in the same environment.

Because of the sequence similarities and the readily recognizable putative mechanisms for the generation of some alleles by mutations in the backbones of other progenitor alleles, it can be postulated that some alleles arose from others (18, 66–68). The common occurrence of sequence-related alleles in multiple African populations strongly suggests that their generation occurred before the expansion and differentiation into the modern African populations. Because the ancient African populations of modern humans may have had limited sizes, the HLA system may have gone through an evolutionary bottleneck. Constraints imposed by the limited number of individuals living in immunologically challenging tropical environments may have, therefore, favored heterozygosity. The consequent balancing selection may have contributed to the permanent establishment of many alleles. Many alleles that are unique to African populations appear to be old, because they are widely distributed in the continent. Our hypothesis is consistent with studies of other genetic markers conducted in other African populations. Analyses of pair-wise mitochondrial DNA-sequence divergence indicate that African populations expanded in size earlier than either Asian or European populations: 99,000 years ago *vs* 52,000 years ago and 23,000 years ago in Asia and Europe, respectively (96, 97). An earlier African expansion is also supported by studies of highly variable microsatellites (98–100).

The distribution of all genotypes including alleles at each of the class I loci of all the African populations reported here are consistent with the proportions expected under HWP. Even though the test of fit may be a weak test for assessing distortions in the distribution of genotypes, it appears that no major temporary selective events were taking place at the time of sample collection. It should be considered, however, that the new infectious agents like the human immunodeficiency virus (HIV) epidemic in Africa are likely to have an important impact in the distribution of HLA alleles and genotypes in African populations. It appears that both heterozygous HLA class I genotypes and the presence of specific alleles have an important impact in delaying the progression to acquired immune deficiency syndrome (AIDS), if HIV replication inhibitory drugs are not utilized (101–107). Our report provides important information of allele and HFs in various African populations, and this information may be useful in the design of modern strategies to fight and prevent infections endemic to the African continents. Knowledge of the make up and distribution of HLA alleles of the local populations is essential for both the selection of peptide vaccines that are designed to cover a large proportion of the population to be

vaccinated and the selection of methods to assess the effectiveness of novel vaccines and/or delivery systems.

The understanding of the genetic diversity and the haplotypic complexity of the HLA system of African populations are essential for the design of registries of volunteer bone marrow donors and cord blood banks. These registries are designed to identify donors and to provide HLA-matched allogeneic hematopoietic stem cells. In the current and other studies (47–52), many alleles and haplotypes were identified and appear to be uniquely found in subjects with African ancestry.

The present study clearly demonstrates that a large proportion of subjects with African descent will only find donors matching their own HLA alleles in subjects with the same ancestry. The difficulties that African recipients have to find suitable donors of allogeneic bone marrow and stem cells are not only related to the occurrence of unique African alleles, but also related to the diversity in HLA alleles and haplotypes. In addition to the high allelic diversity in HLA-B locus, most of the African populations described here present high allelic diversity in HLA-A and HLA-C loci. The diversity of a locus is not necessarily reflected by the number of alleles identified in a particular population, but is better defined by the patterns of allelic distributions. The frequencies of heterozygous genotypes at the HLA-A locus are often lower than those of other HLA loci in the non-African populations (51–63). Most of the African populations described here and the population of African Americans (63) show the lack of one or a few highly predominant alleles at the HLA-A locus. With a few exceptions, the distribution of alleles at all class I loci in the African populations result in low frequencies of homozygous genotypes. In other ethnic groups, the HLA-B locus seems to display the highest diversity and plasticity, with genotype distributions resulting in heterozygosity close to the maximum expected under genetic equilibrium. These allele distributions are consistent with those expected for the overdominant selection model (18–22, 108). Because the African populations have been and are exposed to a wider range of pathogens, a larger allelic repertoire with more alleles may be required to provide the necessary peptide-binding structures to present their antigenic peptides to T-lymphocytes. Because the molecules encoded by alleles of different class I loci have the same redundant function (which is to bind and present peptides to T-lymphocytes), the needs for more allelic diversity could have been met by the relative equalization of the frequencies of multiple alleles of the HLA-A locus.

In the majority of non-African populations, HLA-A presents one or a few alleles with relatively high frequency (5–63); however, because none of these frequently observed alleles predominates overwhelmingly, a significant proportion of the subjects in the non-African populations are heterozygous at HLA-A. This type of distribution may result from a selective advantage provided by one allele of the HLA-A locus, while a meaningful degree of diversity is still maintained for the locus. It is conceivable that in addition to basic common framework functions, the different HLA class I loci perform some specialized functions, which in different environments or circumstances may be either reduced or expanded. We could speculate that the Kenyan and Ugandan populations, and to a lesser extent the Malian population, present frequencies of homozygosity in HLA-A comparable to those observed for the genotypes of HLA-B. In contrast, the Zambian population presents A*3002 as the single most predominant allele (AF = 0.233), which was present in more than 40% of the subjects of this population. It can be argued that directional selection for this allele may have taken place; the observation of multiple HLA-A/B haplotypes bearing A*3002, including A*3002-B*5703-Cw*1801_G, appears to support the hypothesis of directional selection for A*3002 in Zambia. It is worth noting that the frequency of A*3002 in Zambia appears to be one of the highest in the world; the increased frequencies of both B*5703 and Cw*1801_G in this population, compared with other African populations, may be the result of hitch-hiking with an allele that provided a selective advantage.

We also observed low expected homozygosity of HLA-C genotypes that were below neutrality expectations. These were observed in all the African populations, with the exception of the population from Mali. In Malians, the homozygosity statistic, *F*, frequency of homozygous genotypes was virtually identical to that expected under neutrality, and this was almost entirely due to the high frequency of HLA-Cw*1601. Similar to A*3002 in Zambians, the allele Cw*1601 has a distribution that suggests that it may have been positively selected; and again, the observation of several haplotypes bearing this allele supports this hypothesis. We observed higher frequencies of the alleles A*2301_G, B*5201 (with the allele with the silent mutation, B*520102, determined in all informative cases), and B*7801 in the population from Mali, compared to their frequencies in other African populations. The increased frequencies of these alleles that form haplotypes with Cw*1601 in the Malian population may be the result of hitch-hiking. We noticed that many of the haplotypes with Cw*1601 contained alleles related in sequence to the B5 group. It could be argued that some of these alleles may have been generated *de novo* and selected in order to increase or maintain the diversity of the HLA-B locus in this population.

The analysis of the associations between alleles of the three classical HLA class I loci shows that all the African populations have high levels of HLA-A/B/C haplotypic diversity. The haplotype A*3001-B*4201-Cw*1701_G was the only combination of alleles found with relatively high frequency in the present study; it was observed with frequencies between 0.10 and 0.12 in Malians and Zambians, with intermediate frequencies in both Kenyan populations, and was virtually absent from Ugandans. Three additional haplotypes

were observed with frequencies exceeding 0.05; these were the haplotypes A*3002-B*5301-Cw*0401$_G$ in Zambia, A*2301$_G$-B*5201-Cw*1601 in Malians, and A*6601-B*5802-Cw*0602 in the Nandi population from Kenya. We observed a myriad of haplotypes with low or intermediate frequencies, which delineate a vast diversity in HLA haplotypes; this seems to be a common characteristic of all the African populations. This level of diversity does not result from weak LD between alleles of different loci. In fact, our calculations of global LD, utilizing APRDV, D', and W_n (Table 9), indicate that the strength of the associations between alleles of HLA-A and HLA-B loci and those between alleles of HLA-B and HLA-C loci are of similar magnitude as the ones observed for the same pairs of loci in other populations from other regions of the world (data not shown) (63). Therefore, the remarkable haplotypic diversity observed in sub-Saharan populations seems to originate exclusively from the lack of predominant alleles at all loci.

The distribution of haplotypes defined by alleles at the HLA-A and HLA-B loci is similar to that of haplotypes defined by alleles at the three class I loci (Fig. 2) (data not shown). This observation seems to indicate that most of the haplotypic variation in the class I region is related to the strength of the associations between the alleles of the HLA-A and HLA-B loci and the consequent A/B haplotypic diversity. As shown previously, for other segments of extended haplotypes (107), it appears that when the alleles at the loci on the extremes of the haplotypic block are defined, there is limited or no allelic variation at the intervening loci. It is important to pinpoint that the HLA-C alleles can be predicted in African haplotypes by the knowledge of what alleles are found at the HLA-A and HLA-B loci. The majority of the subjects in each population do not carry extended haplotypes. This may be due to the low to intermediate ranges, as measured by APRDV (Table 9), between the HLA-A and HLA-B alleles. The predictive value for HLA-C may decrease in the unassociated haplotypes. The observation of larger haplotypic diversity in the Ugandan population may, therefore, be related to the lack of predominant alleles at HLA-A.

The Ugandan population presents the alleles B*4402$_G$ and Cw*0501$_G$ with frequencies significantly higher than the other African populations in this study. These alleles form the extended haplotype A*0201$_G$-B*4402$_G$-Cw*0501$_G$, which is frequently found in Caucasians; their finding in Ugandans, combined with the frequent observation in this population of haplotypes commonly found in Caucasoids and the closer genetic distances between Caucasians and Ugandans (Table 6), could indicate a recent gene flow of Caucasian or North African genes into the Ugandan population. This observation could explain the greater allelic and haplotypic HLA diversity of the Ugandan population.

Opposing the view of recent admixture, it could be argued that, as has been shown for other loci in the geographically close populations from Ethiopia and Somalia (110–113), the Ugandans present a subset of the variability that is present in other sub-Saharan African populations. Some populations in north-east Africa might have diverged from the rest of sub-Saharan Africa, early in the history of modern African populations; a subset of this north-east African population may have then migrated out of Africa and populated other regions of the world (110–114). Studies of mitochondrial DNA (4, 6, 7, 115) and Y-chromosome diversity (116–119) in other populations, and the distribution of HLA alleles in Ugandans reported here are consistent with the hypothesis of the migration of modern humans out of an East African population. The absence of the widely distributed African haplotype A*3001-B*4201-Cw*1701$_G$ in Ugandans may reflect an early differentiation between Ugandans and other African populations.

Our analysis of genetic distances between the sub-Saharan African populations shows that all populations analyzed here are significantly distant from each other. For example, the distances between each African population exceed those obtained when comparing the distances between various Caucasian populations. In our previous report (63), we observed that the highest genetic distances between Caucasians from the US (63), Irish (51, 52, 120, 121), and those collected in the Centre d'Etude du Polymorphisme Humain (CEPH) study (53) were observed, when comparing the distribution of HLA-B alleles; in our analysis, all the genetic distances between Caucasian populations were below 0.25 (63). In the present study, the Luo and Nandi populations from Kenya have larger distances from each other than those observed between the Caucasian populations. The analysis of the distances obtained for each population shows that the Luo population from Kenya was the closest, or among the two closest populations, for each sub-Saharan population, when analyzing the distributions of alleles at HLA-A and HLA-B. This observation is consistent with the putative origin of modern humans in East Africa (1), followed by the most ancient radiation to the south and west sub-Sahara (1). The Luo population shows the highest frequency of heterozygous genotypes in HLA-A and a large haplotypic diversity. From this observation, it can be concluded that of all populations compared in the present study, the Kenyan Luo population appears to be the oldest or the ancestral population. This observation does not appear to result from admixture, because a large proportion of subjects identified themselves as Luo. All the African populations included in this study showed their highest genetic distance with the predominantly Dogon population from Mali. Timbuktu in Central Mali was an important trade center between West African, Saharan, and North African populations; it is, therefore, conceivable that the influx of alleles from these differentiated populations may contribute to these increased genetic distances between Malians and other sub-Saharan populations. Even though the Cameroonians

have large distances with Zulus and Zambians, the Cameroonians ranked closer to each one of these populations, compared to other African populations; this observation is consistent with other findings that indicate a most recent Bantu expansion from Cameroon to the southern areas of Africa (1, 117, 118, 122, 123).

The findings mentioned above indicate that strong selection has operated at various levels on the HLA system. The most accepted hypothesis indicates that modern humans originated in Africa (1–10). Therefore, the examination of the diversity of the HLA and other genetic systems in contemporary African populations is informative, because significant differentiation between populations living in relatively close geographic areas could only be achieved in longer time elapsed from their origins.

The current level of diversity and the variable allelic distributions observed in different populations are probably the result of evolutionary forces and environments, which have acted on each individual population or in their ancestors. In this regard, the genetic diversity of the HLA system in the African populations poses practical challenges for the design of T-cell vaccines that should effectively cover a large proportion of the individuals. The extensive diversity in this complex genetic system presents difficulties to the transplantation medical community to find HLA-matched transplant pairs that arise from the remarkable allelic and haplotypic variation, which is likely to result from the biological function of the MHC genes and the evolutionary forces that acted on them.

References

1. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet* 2002: **3**: 611–21.
2. Lahr MM, Foley RA. Towards a theory of modern human origins: geography, demography and diversity on recent human evolution. *Yb Phys Anthropol* 1998: **41**: 137–76.
3. Relethford JH. Genetics and the search for modern human origins. New York: John Wiley and Sons, 2001.
4. Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 1995: **57**: 133–49.
5. Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC. mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am J Hum Genet* 2000: **66**: 1362–83.
6. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 2000: **66**: 979–88.
7. Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Paabo S. mtDNA sequence diversity in Africa. *Am J Hum Genet* 1997: **61**: 980–3.
8. Watson E, Forster P, Richards M, Bandelt HJ. Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 1997: **61**: 691–704.
9. Ingman M, Kaessmann H, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature* 2000: **408**: 708–13.
10. Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA et al. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* 1995: **57**: 523–38.
11. Davis MM, Bjorkman PJ. A model for T cell receptor and MHC/peptide interaction. *Adv Exp Med Biol* 1989: **254**: 13–6.
12. Marrack P, Kappler J. The antigen specific, major histocompatibility complex restricted receptor on T cells. *Adv Immunol* 1986: **38**: 1–30.
13. Marrack P, Bender J, Jordan M, Rees W, Robertson J, Schaefer BC et al. Major histocompatibility complex proteins and TCRs: do they really go together like a horse and carriage? *J Immunol* 2001: **167**: 617–21.
14. Marrack P, Kappler J. The T cell repertoire for antigen and MHC. *Immunol Today* 1988: **9**: 308–15.
15. Saper MA, Bjorkman PJ, Wiley DC. Refined structure of the human histocompatibility antigen HLA A2 at 2.6A resolution. *J Mol Biol* 1991: **219**: 277–319.
16. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL et al. Three dimensional structure of the human class II histocompatibility antigen HLA DR1. *Nature* 1993: **364**: 33–9.
17. McFarland BJ, Beeson C. Binding interactions between peptides and proteins of the class II major histocompatibility complex. *Med Res Rev* 2002: **22**: 168–203.
18. Lawlor DA, Zemmour J, Ennis PD, Parham P. Evolution of class I MHC genes and proteins: from natural selection to thymic selection. *Annu Rev Immunol* 1990: **8**: 23–63.
19. Hughes AL, Ota T, Nei M. Positive Darwinian selection promotes charge profile diversity in the antigen binding cleft of class I major histocompatibility complex molecules. *Mol Biol Evol* 1990: **7**: 515–24.
20. Gustafsson K, Wiman K, Emmoth E, Larhammar D, Bohme J, Hyldig Nielsen JJ et al. Mutations and selection in the generation of class II histocompatibility antigen polymorphism. *EMBO J* 1984: 1655–61.
21. Hughes AL, Nei M. Maintenance of MHC polymorphism. *Nature* 1992: **355**: 402–3.
22. Hughes AL, Nei M. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 1989: **86**: 958–62.
23. Opelz G, Wujciak T, Dohler B, Scherer S, Mytilineos J. HLA compatibility and organ transplant survival. Collaborative Transplant Study. *Rev Immunogenet* 1999: **1**: 334–42.
24. Erlich HA, Opelz G, Hansen J. HLA DNA typing and transplantation. *Immunity* 2001: **14**: 347–56.
25. Hansen JA, Yamamoto K, Petersdorf E, Sasazuki T. The role of HLA matching in hematopoietic cell transplantation. *Rev Immunogenet* 1999: **1**: 359.
26. Petersdorf EW, Anasetti C, Martin PJ, Hansen JA. Tissue typing in support of unrelated hematopoietic cell transplantation. *Tissue Antigens* 2003: **61**: 1–11.
27. Muller JY, Halle L, Jaeger G. HLA-A and B antigens in AKA Pygmies. *Tissue Antigens* 1981: **17**: 372–5.

28. Modiano D, Luoni G, Petrarca V, Sodiomon Sirima B, De Luca M, Simpore J et al. HLA class I in three West African ethnic groups: genetic distances from sub-Saharan and Caucasoid populations. *Tissue Antigens* 2001: **57**: 128–37.
29. Renquin J, Sanchez-Masas A, Halle L, Rivallano S, Jaeger G, Mbayo K et al. HLA class II polymorphism in Aka Pygmies and Bantu Congolese and a reassessment of HLA-DRB1 African diversity. *Tissue Antigens* 2001: **58**: 211–22.
30. Novitsky V, Flores-Villanueva PO, Chigwedere P, Gaolekwe S, Bussman H, Sebetso G et al. Identification of most frequent HLA class I antigen specificities in Botswana: relevance for HIV vaccine design. *Hum Immunol* 2001: **62**: 146–56.
31. Tang J, Naik E, Costello C, Karita E, Rivers C, Allen S et al. Characteristics of HLA class I and class II polymorphisms in Rwandan women. *Exp Clin Immunogenet* 2000: **17**: 185–98.
32. Krausa P, McAdam S, Bunce M, Whitworth J, Biryahwaho B, French N et al. HLA-A-B-C-DRB1, DRB3, DRB4, DRB5 and DQB1 polymorphism detected by PCR-SSP in a semi-urban HIV-positive Ugandan population. *Exp Clin Immunogenet* 1999: **16**: 17–25.
33. de Pablo R, Garcia-Pacheco JM, Vilches C, Vilches C, Moreno ME, Sanz L et al. HLA class I and class II allele distribution in the Bubi population from the island of Bioko (Equatorial Guinea). *Tissue Antigens* 1997: **50**: 593–601.
34. Hors J, Chenawi FE, Djoulah S, Hafez M, Abbas F, El Borai M-H et al. HLA in North African populations: 12th International Histocompatibility Workshop NAFR report. In: Charron D, ed. *HLA – Genetic Diversity of HLA Functional and Medical Implication. Anthropology: Contribution of HLA*, Vol. 1. EDK: Medical and Scientific International Publisher, 1997: 328–34.
35. Hammond MG, du Toit ED, Sanchez-Mazas A, Andrien M, Coluzzi M, De Pablo MR et al. HLA in sub-Saharan Africa: 12th International Histocompatibility Workshop SSAF report. In: Charron D, ed. *HLA – Genetic Diversity of HLA Functional and Medical Implication. Anthropology: Contribution of HLA*, Vol. 1. EDK: Medical and Scientific International Publisher, 1997: 345–53.
36. Imanishi T, Akaza T, Kimura A, Tokunaga K, Gojobori T. Reference tables. In: K Tsuji, M Aizawa, T Sasazuki, eds. *HLA 1991 – Proceedings of the Eleventh International Histocompatibility Workshop and Conference*, Vol. 1. Oxford: University press, 1991: 67–1072.
37. Arnaiz-Villena A, Gomez-Casado E, Martinez-Laso J. Population genetic relationships between Mediterranean populations determined by HLA allele distribution and a historic perspective. *Tissue Antigens* 2002: **60**: 111–21.
38. Oumhani K, Canossi A, Piancatelli D, Di Rocco M, Del Beato T, Liberatore G et al. Sequence-based analysis of the HLA-DRB1 polymorphism in Metalsa Berber and Chaouya Arabic-speaking groups from Morocco. *Hum Immunol* 2002: **63**: 129–38. (Erratum In: *Hum Immunol* 2002: 63: 626).
39. Pimtanothai N, Hurley CK, Leke R, Klitz W, Johnson AH. HLA-DR and –DQ polymorphism in Cameroon. *Tissue Antigens* 2001: **58**: 1–8.
40. Arnaiz-Villena A, Martinez-Laso J, Alonso-Garcia J. The correlation between languages and genes: the Usko-Mediterranean peoples. *Hum Immunol* 2001: **62**: 1051–61.
41. Gomez-Casado E, del Moral P, Martinez-Laso J, Garcia-Gomez A, Allende L, Silvera-Redondo C et al. HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens* 2000: **55**: 239–49.
42. Sanchez-Mazas A. African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection. *Hum Immunol* 2001: **62**: 937–48.
43. Migot-Nabias F, Fajardy I, Danze PM, Everaere S, Mayombo J, Minh TN et al. HLA class II polymorphism in a Gabonese Banzabi population. *Tissue Antigens* 1999: **53**: 580–5.
44. Fort M, de Stefano GF, Cambon-Thomsen A, Giraldo-Alvarez P, Dugoujon JM, Ohayon E et al. HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. *Tissue Antigens* 1998: **51**: 327–36.
45. Izaabel H, Garchon HJ, Caillat-Zucman S, Beaurain G, Akhayat O, Bach JF et al. HLA class II DNA polymorphism in a Moroccan population from the Souss, Agadir area. *Tissue Antigens* 1998: **51**: 106–10.
46. Schnittger L, May J, Loeliger CC, Gallin MY, Erttmann KD, Bienzle U et al. HLA DRB1-DQA1-DQB1 haplotype diversity in two African populations. *Tissue Antigens* 1997: **50**: 546–51.
47. Ellis JM, Mack SJ, Leke RF, Quakyi I, Johnson AH, Hurley CK. Diversity is demonstrated in class I HLA-A and HLA-B alleles in Cameroon, Africa: description of HLA-A*03012, *2612, *3006 and HLA-B*1403, *4016, *4703. *Tissue Antigens* 2000: **56**: 291–302.
48. Luo M, Embree J, Ramdahin S, Ndinya-Achola J, Njenga S, Bwayo JB et al. HLA-A and HLA-B in Kenya, Africa: allele frequencies and identification of HLA-B*1567 and HLA-B*4426. *Tissue Antigens* 2002: **59**: 370–80.
49. Sanchez-Mazas A, Steiner QG, Grundschober C, Tiercy JM. The molecular determination of HLA-Cw alleles in the Mandenka (West Africa) reveals a close genetic relationship between Africans and Europeans. *Tissue Antigens* 2000: **56**: 303–12.
50. Ellis JM, Hoyer RJ, Costello CN, Mshana RN, Quakyi IA, Mshana MN et al. HLA-B allele frequencies in Cote d'Ivoire defined by direct DNA sequencing: identification of HLA-B*1405, B*4410, and B*5302. *Tissue Antigens* 2001: **57**: 339–43.
51. Middleton D, Williams F, Meenagh A, Daar AS, Gorodezky C, Hammond M et al. Analysis of the distribution of HLA-A alleles in populations from five continents. *Hum Immunol* 2000: **61**: 1048–52.
52. Williams F, Meenagh A, Darke C, Acosta A, Daar AS, Gorodezky C et al. Analysis of the distribution of HLA-B alleles in populations from five continents. *Hum Immunol* 2001: **62**: 645–50.
53. Bugawan TL, Klitz W, Blair A, Erlich HA. High-resolution HLA class I typing in the CEPH families: analysis of linkage disequilibrium among HLA loci. *Tissue Antigens* 2000: **56**: 392–404.
54. Lazaro AM, Moraes ME, Marcos CY, Moraes JR, Fernandez-Vina MA, Stastny P. Evolution of HLA-class I compared to HLA-class II polymorphism in Terena, a South-American Indian tribe. *Hum Immunol* 1999: **60**: 1138–49.
55. Fernandez-Vina MA, Lazaro AM, Marcos CY, Nulf C, Raimondi E, Haas EJ et al. Dissimilar evolution of B-locus versus A-locus and class II loci of the HLA region in South American Indian tribes. *Tissue Antigens* 1997: **50**: 233–50.
56. Little AM, Scott I, Pesoa S, Marsh SG, Arguello R, Cox ST et al. HLA class I diversity in Kolla Amerindians. *Hum Immunol* 2001: **62**: 170–9.
57. Hollenbach JA, Thomson G, Cao K, Ferdandez-Vina M, Erlich HA, Bugawan TL et al. HLA diversity, differentiation, and haplotype evolution in Mesoamerican natives. *Hum Immunol* 2001: **62**: 378–90.
58. Saito S, Ota S, Yamada E, Inoko H, Ota M. Allele frequencies and haplotypic associations defined by allelic DNA typing at HLA class I and class II loci in the Japanese population. *Tissue Antigens* 2000: **56**: 522–9.
59. Tokunaga K, Ishikawa Y, Ogawa A, Wang H, Mitsunaga S, Moriyama S et al. Sequence-based association analysis of HLA class I and II alleles in Japanese supports conservation of common haplotypes. *Immunogenetics* 1997: **46**: 199–205.

60. Bannai M, Ohashi J, Harihara S, Takahashi Y, Juji T, Omoto K et al. Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens* 2000: **55**: 128–39.
61. Bugawan TL, Mack SJ, Stoneking M, Saha M, Beck HP, Erlich HA. HLA class I allele distributions in six Pacific/Asian populations: evidence of selection at the HLA-A locus. *Tissue Antigens* 1999: **53**: 311–9.
62. Main P, Attenborough R, Chelvanayagam G, Bhatia K, Gao X. The peopling of New Guinea: evidence from class I human leukocyte antigen. *Hum Biol* 2001: **73**: 365–83. (Erratum In: *Hum Biol* 2001: 73: 782).
63. Cao K, Hollenbach J, Shi X, Shi W, Chopek M, Fernandez-Vina MA. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum Immunol* 2001: **62**: 109–30.
64. Tsai Y, Thomson G. Selection intensity differences in seven HLA loci in many populations. In: Hansen JA, Dupont B, eds. *HLA 2002 – Immunobiology of the Human MHC*. IHWG Press, 2004. (in press).
65. Cao K, Chopek M, Fernandez-Vina MA. High and intermediate resolution DNA typing systems for class I HLA-A, B, C genes by hybridization with sequence-specific oligonucleotide probes (SSOP). *Rev Immunogenet* 1999: **1**: 177–208.
66. Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 1988: **335**: 268–71.
67. Parham P, Lawlor DA, Lomen CE, Ennis PD. Diversity and diversification of HLA-A, B, alleles. *J Immunol* 1989: **142**: 3937–50.
68. Parham P, Adams EJ, Arnett KL. The origins of HLA-A, B, C polymorphism. *Immunol Rev* 1995: **143**: 141–80.
69. Lancaster A, Nelson MP, Single RM, Meyer D, Thomson G. Software framework for the biostatistics core. In: Hansen JA, Dupont B, eds. *HLA 2002 – Immunobiology of the Human MHC*. IHWG Press, 2004. (in press).
70. Single R, Meyer D, Thomson G. Statistical Methods for analysis of population genetic data. In: Hansen JA, Dupont B, eds. *HLA 2002 – Immunobiology of the Human MHC*. IHWG Press, 2004. (in press).
71. Guo SW, Thompson EA. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 1992: **48**: 361–72.
72. Schneider S, Kueffer JM, Roesli D, Excoffier L. Arlequin: A software package for population genetics, Version 2.0, Geneva: Genetics and Biometry Laboratory, Department of Anthorpology, University of Geneva 1999.
73. Watterson GA. The homozygosity test of neutrality. *Genetics* 1978: **88**: 405–17.
74. Ewens WJ. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 1972: **3**: 87–112.
75. Slatkin M. An exact test for neutrality based on the Ewens sampling distribution. *Genet Res* 1994: **64**: 71–4.
76. Liau SW, Mickey MR, Romano P, Lee TD. Study of HLA system in the Haitian Population. *Tissue Antigens* 1984: **23**: 308–13.
77. Cavalli-Sforza LL, Bodmer WF. *The Genetics of Human Populations*. San Francisco: W.H. Freeman, 1971.
78. Sanchez-Velasco P, Escribano-Dediego J, Paz-Miguel JE, Ocejo-Vinyals G, Leyua-Cobian F. HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (pasvalley, Northen Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations. *Tissue Antigens* 1999: **53**: 65–73.
79. Grimaldi MC, Cronan-Roy B, Amoros JP, Cambon-Thomsen A, Carcassi C, Orru S et al. West Mediterranean Islands (Corcica, Balearic islands, Sardinia) and the Basque population: contribution of HLA class I molecular markers to their evolutionary history. *Tissue Antigens* 2001: **58**: 281–92.
80. Martinez-Laso F, Dejuan D, Martinez-Quiles N, Gomez-Casado E, Cuadrado E, Arnaiz-Villena A. The contribution of the HLA-A, B, C and DR-DQ-DNA testing to the study of the origins of Spaniards and Basques. *Tissue Antigens* 1995: **45**: 137–45.
81. Arnaiz-Villena A, Karin M, Bendikuze N, Gomez-Casado E, Moscosio J, Silvera C. HLA alleles and haplotypes in the Turkish population: relatedness to kurds, Armenians and other Mediterraneans. *Tissue Antigens* 2001: **57**: 308–12.
82. Williams RC, Troup GM, Nelson JL, Kostyu DD, McAuley JE, Pettitt DJ et al. Report North American Indian anthropology section. In: Tsuji K, Aizawa M, Sasazuki T, eds. *HLA 1991 – Proceedings of the of the 11th International Histocompatibility Workshop*. New York: Oxford University Press, 1992: 683–5.
83. Petzel-Erler ML, Gorodezky C, Layrisse Z, Klintz W, Fainboim L, Vullo C et al. Anthropology component report for region Latin-America: Amerindian and admixed populations. In: Charron D, ed. *Genetic Diversity of HLA: Functional and Medical* Implications, Vol 1. Paris: EDK Press, 1997: 337–45.
84. Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R et al. Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature* 1992: **357**: 326–9.
85. Watkins DI, McAdam SN, Liu X, Strang CR, Milford EL, Levine CG et al. New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature* 1992: **357**: 329–33.
86. Cerna M, Falco M, Friedman H, Raimondi E, Maccagno A, Fernandez Vina MA et al. Differences in HLA class II alleles of isolated South American Indian populations from Brazil and Argentina. *Hum Immunol* 1993: **37**: 213–20.
87. Fernandez Vina MA, Moraes JR, Moraes ME, Miller S, Stastny P. HLA class II haplotypes in Amerindians and in black North and South Americans. *Tissue Antigens* 1991: **38**: 235–40.
88. Titus Trachtenberg EA, Rickards O, De Stefano GF, Erlich HA. Analysis of HLA class II haplotypes in the Cayapa Indians of Ecuador: a novel DRB1 allele reveals evidence for convergent evolution and balancing selection at position 86. *Am J Hum Genet* 1994: **55**: 160–7.
89. Yunis JJ, Ossa H, Salazar M, Delgardo MB, Deulofeut R, de la Hoz A et al. Major histocompatibility complex class II alleles and haplotypes and blood groups of four Amerindian tribes of northern Colombia. *Hum Immunol* 1994: **41**: 248–58.
90. Vullo CM, Delfino L, Angelini G, Ferrara GB. HLA polymorphism in a Mataco South American Indian tribe: serology of class I and II antigens. Molecular analysis of class II polymorphic variants. *Hum Immunol* 1992: **35**: 209–14.
91. Parham P, Arnett KL, Adams EJ, Little AM, Tees K, Barber LD et al. Episodic evolution and turnover of HLA-B in the indigenous human populations of the Americas. *Tissue Antigens* 1997: **50**: 219–32.
92. Sathiamurthy M, Hickman HD, Cavett JW, Zahoor A, Prilliman K, Metcalf S et al. Population of the HLA ligand database. *Tissue Antigens* 2003: **61**: 12–9.

93. Clavijo OP, Delgado JC, Yu N, Fraser PA, Yunis EJ. HLA-Cw*1701 is associated with two sub-Saharan African-derived HLA haplotypes. HLA-B*4201, DRB1*03 and HLA-B*4202 without DRB1*03. *Tissue Antigens* 1999: **54**: 303–6.
94. Tian W, Boggs DA, Ding WZ, Chen DF, Fraser PA. MICA genetic polymorphism and linkage disequilibrium with HLA-B in 29 African-American families. *Immunogenetics* 2001: **53**: 724–8.
95. Barber LD, Percival L, Arnett KL, Gumperz JE, Chen L, Parham P. Polymorphism in the alpha 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. *J Immunol* 1997: **158**: 1660–9.
96. Jorde LB, Bamshad M, Rogers AR. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 1998: **20**: 126–36.
97. Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M. Mismatch distributions of mtDNA reveal recent human population expansions. *Hum Biol* 1994: **66**: 761–75.
98. Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB. Signatures of population expansion in microsatellite repeat data. *Genetics* 1998: **148**: 1921–30.
99. Reich DE, Goldstein DB. Genetic evidence for a Paleolithic human population expansion in Africa. *Proc Natl Acad Sci USA* 1998: **95**: 8119–23. (Erratum In: *Proc Natl Acad Sci USA* 1998: 95: 11026).
100. Relethford JH, Jorde LB. Genetic evidence for larger African population size during recent human evolution. *Am J Phys Anthropol* 1999: **108**: 251–60.
101. Liu C, Carrington M, Kaslow RA, Gao X, Rinaldo CR, Jacobson LP et al. Association of polymorphisms in human leukocyte antigen class I and transporter associated with antigen processing genes with resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 2003: **187**: 1404–10.
102. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R et al. Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* 2001: **344**: 1668–75.
103. Hendel H, Caillat-Zucman S, Lebuanec H, Carrington M, O'Brien S, Andrieu JM et al. New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *J Immunol* 1999: **162**: 6942–6.
104. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ et al. HLA and HIV-1: heterozygote advantage and B*35–Cw*04 disadvantage. *Science* 1999: **283**: 1748–52.
105. Gillespie GM, Kaul R, Dong T, Yang HB, Rostron T, Bwayo JJ et al. Cross-reactive cytotoxic T lymphocytes against a HIV-1 p24 epitope in slow progressors with B*57. *AIDS* 2002: **16**: 961–72.
106. Costello C, Tang J, Rivers C, Karita E, Meizen-Derr J, Allen S et al. HLA-B*5703 independently associated with slower HIV-1 disease progression in Rwandan women. *AIDS* 1999: **13**: 1990–1.
107. Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* 1996: **2**: 405–11.
108. Begovich AB, McClure GR, Suraj VC, Helmuth RC, Fildes N, Bugawan TL et al. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *J Immunol* 1992: **148**: 249–58.
109. Awdeh ZL, Raum D, Yunis EJ, Alper CA. Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man. *Proc Natl Acad Sci USA* 1983: **80**: 259–63.
110. Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K et al. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 1996: **271**: 1380–7.
111. Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B et al. A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. *Am J Hum Genet* 1998: **62**: 1389–402.
112. Tishkoff SA, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A et al. Short tandem-repeat polymorphism/alu haplotype variation at the PLAT locus: implications for modern human origins. *Am J Hum Genet* 2000: **67**: 901–25.
113. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG et al. Population genetic structure of variable drug response. *Nat Genet* 2001: **29**: 265–9.
114. Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK. Short tandem repeat polymorphism evolution in humans. *Eur J Hum Genet* 1998: **6**: 38–49.
115. Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS. Genetic evidence of an early exit of Homo sapiens from Africa through eastern Africa. *Nat Genet* 1999: **23**: 437–41.
116. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH et al. Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000: **26**: 358–61.
117. Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H et al. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 2001: **18**: 1189–203.
118. Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T et al. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 1998: **15**: 427–41.
119. Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 2002: **70**: 265–8.
120. Middleton D, Williams F, Hamill MA, Meenagh A. Frequency of HLA-B alleles in a Caucasoid population determined by a two-stage PCR-SSOP typing strategy. *Hum Immunol* 2000: **61**: 1285–97.
121. Williams F, Meenagh A, Maxwell AP, Middleton D. Allele resolution of HLA-A using oligonucleotide probes in a two-stage typing strategy. *Tissue Antigens* 1999: **54**: 59–68.
122. Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 1998: **62**: 420–34.
123. Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T. mtDNA control-region sequence variation suggests multiple independent origins of an 'Asian-specific' 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* 1996: **58**: 595–608.