Am. J. Hum. Genet. 50:411–421, 1992

Interethnic Genetic Differentiation in Africa: HLA Class I Antigens in The Gambia

Catherine E. M. Allsopp,* Rosalind M. Harding,† Craig Taylor,‡ Michael Bunce,‡ Dominic Kwiatkowski,§ Nicholas Anstey,§ David Brewster,‖ Andrew J. McMichael,* Brian M. Greenwood,§ and Adrian V. S. Hill*

*Molecular Immunology Group, †Medical Research Council Molecular Haematology Unit, Institute of Molecular Medicine, and ‡Tissue Typing Laboratory, Nuffield Department of Surgery, University of Oxford, John Radcliffe Hospital, Oxford; §Medical Research Council Laboratories, Fajara; and ‖Royal Victoria Hospital, Banjul, The Gambia

Summary

A total of 752 individuals from The Gambia, west Africa who are representative of the major ethnic groups in the capital, Banjul, were serologically typed for HLA-A, -B, and -C antigens. Although all were typically "African" in their antigenic profiles, some marked frequency differences were found between the ethnic groups. Genetic distance comparisons with several other African populations showed that, although these west African populations clustered closely together, the positions of the various ethnic groups in The Gambia were consistent with historical and linguistic evidence of their affinities with one another and with other African populations. Despite the potential confounding effects both of selection by infectious diseases and of genetic drift caused by local differences in population structure, HLA frequencies appear to be of value in measuring inter- and intraregional population affinities in sub-Saharan Africa.

Introduction

Human leukocyte antigens were among the first genetic markers to be compared in numerous human populations (Dausset and Colombani 1972) and have been widely studied. Their power in defining population affinities derives from their unique degree of sequence diversity, with almost 50 alleles identified for each of the most polymorphic HLA class I and class II genes (Bodmer et al. 1991). The discovery of population-specific antigens, haplotypes, and, more recently, subtypes has increased the usefulness of this system in population comparisons. For example, using HLA frequencies alone, Serjeantson et al. (1982) were able to construct a picture of prehistoric colonization routes in the Pacific that was compatible with evidence from several independent disciplines. However, although only a few years ago it could be argued that most alleles were effectively neutral (Klein 1987; Serjeantson 1989), there is now convincing evidence of several types to indicate that HLA alleles are subject to significant balancing selection pressures (Hedrick and Thomson 1983; Hughes and Nei 1988; Hill et al. 1991*b*). This violation of the neutrality assumption raises concerns about the interpretation of population comparisons using HLA markers subject to ongoing selection pressures (Jorde 1985).

Interest in the population genetics of sub-Saharan Africa has increased in recent years because of both accumulating evidence that Africans may display more genetic diversity than do other racial groups and the somewhat controversial proposal that modern man originated in Africa (Wainscoat et al. 1986; Cann et al. 1987; Excoffier et al. 1987; Cavalli-Sforza et al. 1988; Nei and Livshits 1989). Unfortunately, the logistic difficulties of obtaining viable lymphocytes from large numbers of individuals in this region mean that relatively few data sets of HLA frequencies are available. Sample sizes have been small, and there are no available comparisons of black ethnic groups within a country. In the context of a large case-control study

Received August 1, 1991; revision received September 25, 1991.

Address for correspondence and reprints: Dr. Adrian V. S. Hill, Molecular Immunology Group, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, England.

0002-9297/92/5002-0018$02.00

of malaria susceptibility in the west African state of The Gambia (Hill et al. 1991*b*), we had the opportunity to compare HLA class I frequencies in the ethnic groups living around the capital city, Banjul.

We have documented their HLA class I antigen frequencies and have performed a genetic distance comparison (of derived gene frequencies) with other African populations. The objectives were to assess the extent of genetic differentiation among ethnic groups and to determine to what degree the known population history in this region might be reflected in present-day HLA frequencies. We find that HLA class I frequencies, despite the confounding effects of selection and possibly different population structures among different ethnic groups, provide a measure of population affinities that is quite compatible with known population histories.

Subjects and Methods

Population Samples

Seven hundred fifty-two blood samples were collected from 641 unrelated children and 111 unrelated healthy adult male blood donors resident within a 25-mile radius around the capital city, Banjul. The samples from children were collected as part of a case-control study of *Plasmodium falciparum* malaria, details of which are reported elsewhere (Hill et al. 1991*b*). Although matching of cases and controls for ethnic group was not performed, the proportions of each ethnic group in the different clinical groups were very similar, perhaps related to matching for area of residence. As reported elsewhere, the only HLA class I antigen frequency found to be significantly affected by malaria was HLA-Bw53 (Hill et al. 1991*b*). Hence, we felt that it was valid in the interethnic comparisons to use the pooled data from malaria cases and the control children for all antigens except HLA-Bw53; for the latter the frequencies reported in tables 1 and 2 are from control children without malaria and from healthy adults. To further exclude an influence of malaria on our general conclusion concerning population relationships, genetic distances were calculated using only the data on individuals (a total of 429) without malaria, yielding very similar results. The analysis of the pooled data is presented because the sampling-variation effect due to small sample size appeared to be more significant than the influence of malaria on antigen frequencies.

Data on the other populations used for comparison were from published sources. The black South Africans were healthy adult Xhosa (du Toit et al. 1987); the San Bushmen were families of the southern !Kung San group (du Toit et al. 1990*b*); the Zimbabweans, more than 95% of whom were Shona, were unrelated blood donors living near Harare (du Toit et al. 1990*a*); the Nigerians were healthy volunteers living in or visiting London (Okoye et al. 1985); and the Malians were healthy individuals from around Bamako (Kalidi et al. 1988). The American blacks were from the United States (HLA-A and -B frequencies were from the very large data set of Lee et al. [1991], but, in the absence of HLA-C frequencies from that report, those of Johnson [1986] were used), and the Japanese and European frequencies are from the Eighth Histocompatibility Workshop study (Terasaki 1980). Although different sets of antisera have been used in these various studies, the standardization of specificities, as judged by International Histocompatibility Workshops in the 1980s, is sufficient to allow reliable frequency comparisons.

History and Demography

Our knowledge of west African prehistory is limited by a paucity of archaeological evidence, and it was only with the establishment of the trans-Saharan trade routes that written history became available (Phillipson 1985). Arab traders to this area between the 8th and 16th centuries A.D. documented the presence of large and powerful kingdoms: before 1450 A.D., those of Mali, Ghana, and Kanem-Borno and, during 1450–1600, the Songhai empire (Crowder 1977). Although there were several ethnic groups within each kingdom, the extent of admixture may have been limited (Sonko-Godwin 1985).

Today, the four largest ethnic groups in the Kombo area of the western region of The Gambia, which surrounds the capital, Banjul, are the Mandinkas (42%), Jolas (14%), Wolofs (14%), and Fulas (12%). Sources of information on their population history are limited archaeological finds (Phillipson 1985), linguistic analysis (Greenberg 1963), and, for more recent times, oral history and written records. The predominant Mandinka group is said to originate from Kangaba, a state of the ancient empire of Mali, and they expanded widely in the Senegambia region some 700 years ago. The Wolofs, who are now most numerous to the north of The Gambia, in Senegal, are thought to have been forced southward by the Berber expansion some 1,300 years ago. The Fula are the most phenotypically distinct of the four ethnic groups. They are relatively light skinned with slender build and, occasionally, some-

Table I

Percentage Antigen Frequencies of HLA-A, -B, and -C Antigens in Various Ethnic Groups in Area of The Gambia Studied

	HLA Class I Antigen Frequencies in The Gambia[a]						
	Mandinka (n = 318)	Jola (n = 123)	Wolof (n = 109)	Fula (n = 90)	Manjago (n = 41)	Serere (n = 25)	Other (n = 46)
A1	13.2	19.5	1.1	13.3	9.8	4.0	8.7
A2	21.4	26.0	35.8	24.4	31.7	44.0	26.1
A3	7.9	6.5	7.3	14.4	12.2	12.0	4.3
A11	.0	.0	.0	.0	.0	.0	.0
A23	25.8	30.9	20.2	23.3	29.3	.80	26.1
A24	.3	.8	.9	3.3	.0	.0	.0
A25	.0	.0	.0	.0	.0	.0	.0
A26	16.0	22.0	11.9	8.9	12.2	20.0	6.5
A28	21.1	22.8	14.7	18.9	7.3	12.0	15.2
A29	4.4	.0	2.8	4.4	2.4	4.0	8.7
A30	28.0	11.4	22.9	17.8	19.5	24.0	28.3
A31	.9	.8	.0	.0	2.4	.0	.0
A32	5.3	4.9	10.1	13.3	4.9	4.0	8.7
Aw33	23.3	24.4	18.3	23.3	24.4	36.0	19.6
Aw34	3.1	1.6	4.6	2.2	4.9	8.0	6.5
Aw36	.0	1.6	.0	.0	.0	.0	.0
Aw43	.0	.0	.0	.0	.0	.0	.0
Aw74	.0	.0	.0	.0	.0	.0	.0
B7	11.0	7.3	18.3	22.2	12.2	8.0	13.0
B8	19.2	23.6	23.9	13.3	19.5	16.0	8.7
B13	1.6	.0	.9	.0	2.4	.0	.0
B14	9.1	5.7	4.6	4.4	7.3	16.0	6.5
B15	2.8	3.3	.9	2.2	2.4	.0	2.2
B17	19.8	28.5	25.7	23.3	12.2	20.0	8.7
B18	4.1	5.7	10.1	5.6	2.4	.0	6.5
B22	1.9	3.3	.9	7.8	7.3	.0	.0
B27	2.2	4.1	3.7	7.8	.0	.0	2.2
B35	31.4	22.8	27.5	27.8	31.7	28.0	43.5
B37	.3	.0	.0	4.4	2.4	.0	2.2
B38	.0	.0	.0	.0	.0	.0	.0
B39	1.6	5.7	2.8	1.1	4.9	4.0	2.2
B40	.6	.0	.9	4.4	2.4	.0	2.2
Bw41	3.5	2.4	4.6	2.2	.0	4.0	.0
Bw42	5.3	1.6	8.3	3.3	4.9	.0	4.3
B44	3.1	4.1	2.8	5.6	4.9	.0	4.3
B45	3.1	.8	2.8	5.6	.0	.0	10.9
B49	8.5	17.1	2.8	2.2	7.3	28.0	4.3
Bw50	5.7	3.3	4.6	4.4	2.4	.0	4.3
B51	4.7	5.7	2.8	4.4	4.9	12.0	4.3
Bw52	.6	.0	.9	.0	.0	.0	.0
Bw53	27.4	18.9	11.1	18.5	13.6	42.9	48.0
Bw70	14.8	17.1	13.8	11.1	22.0	16.0	10.9
Cw1	5.0	8.9	2.8	4.4	12.2	.0	.0
Cw2	15.7	11.4	15.6	18.9	12.2	16.0	17.4
Cw3	29.2	47.2	31.2	33.3	22.0	24.0	8.7
Cw4	26.1	26.0	19.3	33.3	34.1	28.0	50.0
Cw5	4.4	4.9	9.2	5.6	2.4	.0	2.2
Cw6	15.7	10.6	13.8	16.7	12.2	16.0	8.7
Cw7	17.3	13.0	22.0	17.8	17.1	24.0	15.2
Cw8	8.2	4.9	4.6	3.3	7.3	12.0	4.3

[a] Sample sizes (n) are as shown, except for HLA-Bw53, for which the antigen frequencies for only the individuals without malaria (total number = 429) are given. The other samples include members of the following ethnic groups: Serahuli, Aku, and Tukulor. The following antigens were also tested for but not found: Bw47, Bw48, and Bw73.

Table 2

HLA Class I Allele Frequencies in Various African Populations, American Blacks, Europeans, and Japanese

	HLA Class 1 Allele Frequency in[a]								
	Gambians (n = 1,504)	Malians (n = 164)	Nigerians (n = 228)	Zimbabweians (n = 238)	South Africans (n = 2,054)	San (n = 580)	American Blacks (n = 3,004)	Europeans (n = 5,300)	Japanese (n = 1,900)
A1	.069	.037	.022	.026	.035	.021	.059	.149	.005
A2	.141	.159	.147	.155	.130	.139	.188	.260	.246
A3	.044	.037	.097	.030	.070	.147	.088	.116	.006
A11	.000	.006	.000	.000	.002	.000	.018	.059	.090
A23	.135	.096	.097	.079	.080	.108	.105	.023	.006
A24	.004	.012	.005	.021	.027	.057	.022	.096	.356
A25	.000	.006	.000	.000	.001	.000	.007	.019	.001
A26	.078	.018	.009	.039	.064	.002	.030	.037	.098
A28	.099	.116	.102	.116	.149	.102	.100	.039	.010
A29	.018	.063	.027	.070	.075	.009	.031	.038	.002
A30	.121	.103	.107	.267	.182	.193	.131	.024	.002
A31	.003	.018	.036	.004	.003	.002	.015	.027	.080
A32	.036	.037	.009	.004	.017	.079	.015	.045	.001
Aw33	.123	.050	.132	.052	.031	.004	.068	.017	.068
Aw34	.018	.024	.059	.021	.035	.000	.045	.006	.010
Aw36	.001	.000	.082	.034	.002	.002	.025	.004	.003
Aw43	.000	NT	NT	.000	.029	.109	NT	.000	.000
Aw74	NT	.063	.040	.039	.044	.019	.001	NT	NT
A blank	.113	.155	.032	.043	.027	.006	.058	.043	.024
B7	.067	.089	.073	.088	.105	.127	.097	.088	.059
B8	.101	.018	.000	.052	.062	.125	.046	.082	.001
B13	.005	.000	.009	.008	.013	.012	.007	.028	.020
B14	.038	.037	.000	.043	.033	.024	.038	.029	.001
B15	.012	.012	.040	.004	.016	.049	.031	.016	.090
B17	.114	.037	.178	.177	.227	.320	.111	.043	.009
B18	.027	.012	.054	.079	.039	.012	.031	.058	.000

B22	.014	.012	.000	.000	.000	.000	.008	.028	.118
B27	.016	.050	.000	.000	.000	.000	.016	.039	.004
B35	.161	.082	.059	.021	.021	.000	.082	.096	.073
B37	.005	.012	.009	.000	.001	.000	.006	.015	.006
B38	.000	.012	.000	.000	.001	.000	.004	.025	.002
B39	.013	.012	.022	.017	.021	.002	.014	.021	.029
B40	.006	.000	.005	.000	.002	.002	.022	.051	.160
Bw41	.015	.006	.009	.013	.016	.092	.004	.010	.004
Bw42	.024	.012	.068	.116	.113	.004	.063	.003	.006
B44	.018	.006	.054	.070	.065	.042	.076	.110	.065
B45	.016	.012	.022	.056	.050	.002	.047	.011	.002
B47	.000	.006	.000	.004	.000	.016	.001	.005	.002
B48	.000	.000	.000	.000	.000	.000	.003	.005	.023
B49	.044	.024	.040	.013	.009	.002	.025	.023	.003
Bw50	.023	.006	.000	.004	.001	.002	.012	.013	.000
B51	.024	.110	.009	.008	.008	.014	.019	.072	.083
Bw52	.003	.089	.013	.004	.002	.000	.016	.015	.108
Bw53	.147	.043	.222	.083	.010	.004	.115	.009	.001
Bw70	.077	.043	.111	.111	.158	.137	.077	.000	.000
B blank	.063	.263	.004	.029	.031	.014	.025	.107	.134
Cw1	.026	.012	.000	.000	.002	.000	.017	.041	.176
Cw2	.080	.152	.078	.093	.125	.142	.053	.051	.004
Cw3	.170	.103	.082	.097	.069	.028	.084	.101	.269
Cw4	.151	.137	.287	.140	.102	.162	.129	.121	.048
Cw5	.027	.000	.045	.004	.005	.000	.047	.060	.001
Cw6	.073	.006	.112	.111	.220	.327	.123	.079	.007
Cw7	.084	.089	.121	.165	.220	.181	.123	.023	.011
Cw8	.032	.012	.027	.039	.041	.025	.059	.019	.001
C blank	.357	.489	.249	.351	.218	.137	.366	.505	.485

[a] Sample sizes (n) are shown as allele numbers—i.e., twice the number of individuals typed.

[b] Frequencies are from total data set including all ethnic groups. The ethnic distribution of blank allele frequencies in The Gambia, for Mandinka, Jola, Wolof, Fula, Manjago and Serere, was as follows: A blank—.197, .080, .146, .118, .147, and .048; B blank—.078, .059, .077, .056, .094, and .046; and C blank—.357, .312, .360, .286, .367, and .363.

what caucasoid facial features. The Gambian Fula are a subgroup of the nomadic pastoral Fulani people who are widely dispersed in west Africa. They are believed to be of north African or east African origin. Least is known of the Jolas, for whom there is no written history, but they are considered to be descendants of the original inhabitants of this area and are particularly numerous to the south in the Cassamance region of Senegal (Sonko-Godwin 1985). Although the ethnic groups are now largely intermingled in the conurbation around Banjul, there is believed to have been relatively little intermarriage between the ethnic groups.

HLA Class I Typing

Peripheral blood lymphocytes (PBL) were separated using Ficoll-hypaque and were either HLA typed in The Gambia or cryopreserved in liquid nitrogen and transported to Oxford for subsequent analysis. PBL were serologically typed for HLA-A, -B, and -C antigens by using the standard NIH microlymphocytotoxicity test and 180 well-characterized antisera.

Analysis

Gene frequencies were calculated from the antigen frequencies by using the square-root method: gene frequency = $1 - (1 - \text{antigen frequency})^{1/2}$. Because, for each population, the summed gene frequencies were less than 1, their complements to 1 are considered to represent unidentified alleles and are categorized as blanks. These blanks are, arbitrarily, treated as an additional single allele at each locus in the subsequent analyses. The gene frequency data were analyzed using the statistical package for numerical taxonomy, NTSYS-pc (Rohlf 1990). We chose to calculate chord distances (Cavalli-Sforza and Edwards 1967), rather than alternative distance measures, because of their previous use in the HLA literature. As in previous population comparisons using HLA frequencies, independence of loci is assumed in the genetic distance comparisons, even though linkage disequilibrium between certain alleles of the various loci, particularly HLA-B and -C, is observed. This probably leads to a small reduction in the information content of the overall frequencies. The structure in the resulting genetic distance matrix was investigated by UPGMA (unweighted pair-group method, arithmetic average) clustering (Sokal and Rohlf 1981) and principal components analysis.

Information different from that presented by a phenogram (fig. 1) of the genetic distance matrix is preserved in a plot of the principal components of the allele frequency correlation matrix. Whereas phenograms ideally display divergence, in time, of descendant populations from a common ancestor, a principal components plot can represent population affinities resulting from patterns of gene flow. Although the total genetic variation can be presented in as many dimensions as there are allelic variables (here, 55 class I alleles), principal components analysis representation allows a substantial proportion of this variation to be represented in a small number of dimensions. Three dimensions are used here because this allows convenient graphical representation. The positions of populations in three-dimensional space, with axes at 90° to each other, are computed so as to preserve the largest possible amount of information on their affinities. The first, second, and third component axes (labeled I, II, and III, respectively, in figs. 2 and 3) preserve the first-, second-, and third-ranked percentages of the total variation. In addition, in figures 2 and 3, superimposed on the three-dimensional representation are lines connecting a minimum-spanning tree. In this tree, a line joins a population to its nearest genetic neighbor.

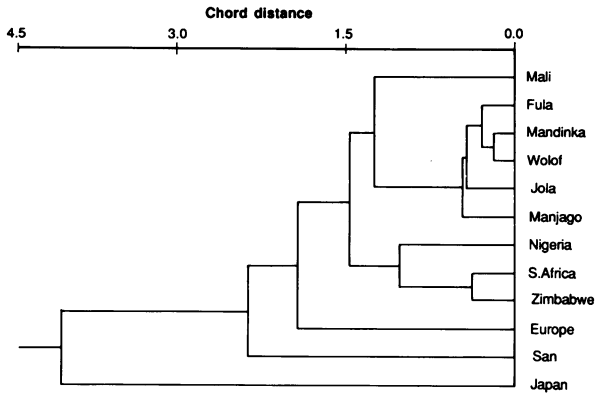

Figure 1 Dendrogram produced by UPGMA method, based on genetic (chord) distances between various populations.

Results

Antigen Frequencies

HLA class I antigen frequencies of the different ethnic groups are shown in table 1, and the derived allele frequencies of the entire Gambian sample compared with those of other African and world populations are shown in table 2. Overall, the Gambian frequencies are generally similar to those found in other African populations, particularly those of Nigeria and Mali. HLA-A antigens that are considered relatively "Afri-

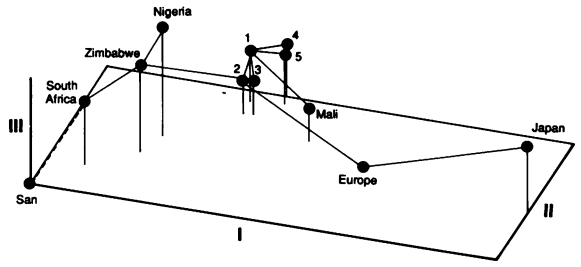

Figure 2 Principal components analysis. The relationships between the 12 populations are presented as a three-dimensional diagram plotting the three principal axes, I–III (see Subjects and Methods), which preserve 64% of the genetic variation. Superimposed on the three-dimensional representation are lines connecting a minimum-spanning tree. In this tree a line joins a population to its nearest genetic neighbor. The Gambian ethnic groups are numbered as follows: 1 = Mandinka; 2 = Fula; 3 = Wolof; 4 = Jola; 5 = Manjago.

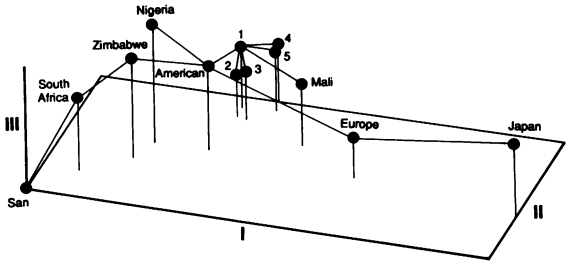

Figure 3 Principal components analysis as in fig. 2, with addition of the American black population.

can specific" (Johnson 1986) include A23, A30, and Aw33, and these have high antigen frequencies of 25%, 23%, and 23%, respectively, in The Gambia. Another "black" allele, Aw36, is rare in The Gambia and Mali compared with Nigeria. As in most African populations, A11 is absent in The Gambia, and A43, which is considered to be characteristic of southern African Bushmen (Nurse et al. 1985), is also absent.

At the HLA-B locus, Bw42, well recognized as a common African antigen (Johnson 1986), is most frequent in Nigerian and South African populations, with antigen frequencies of 13% and 21%, respectively, but in The Gambia it is found in 4.7% of the population, a low frequency comparable with that found in Mali. In The Gambia it is, nonetheless, in strong linkage disequilibrium with HLA-Aw30 (Δ = .029), as has been reported for South African and American blacks (du Toit et al. 1987; Lee et al. 1991). The African antigens, HLA-Bw53 and B70, are both common in The Gambia (27.3% and 14.7%, respectively), Nigeria (40% and 21%) and Zimbabwe (16% and 21%) but less so in Mali (both 8.5%). The most common B locus antigen in The Gambia, B35, is relatively rare in the southern African populations and is completely absent in the San Bushmen. As with other African and non-African populations, a very high frequency (.36) of HLA-C alleles in the Gambia were not typable, or "blank."

Comparison of antigen frequencies of the ethnic groups within The Gambia showed fewer differences. Most notable is the A1 frequency difference between the Wolofs (1.1%) and the Jolas (19.5%), but several frequencies appear somewhat characteristic of particular groups: B49 is high among Jolas, A24 and B40 are found mainly among the Fulas, and there are low frequencies of Bw53 and A1 among the Wolofs.

Genetic Distance Analysis

Genetic distances between the various populations were calculated from the allele frequencies for the five largest Gambian ethnic groups, the other African populations, American blacks, and European and Japanese reference populations, and these are shown in the genetic distance matrix (table 3). UPGMA clustering on the full matrix indicates that there are three major branches separating the European, the Japanese, and the San from the other African populations (fig. 1). This branching indicates the distinctiveness of these three populations compared with the other Africans. The latter form two main clusters; the first of these is composed of South Africa, Zimbabwe, and Nigeria, all of which are Bantu populations (Greenberg 1963), and the second is of the ethnic groups of The Gambia and the Malians, all non-Bantu west Africans.

Population affinities were also displayed using principal components analysis (see Subjects and Methods). In this application the populations are mapped into a three-dimensional genetic space (fig. 2), defined by computing the three principle axes of a 55 × 55 HLA allele frequency correlation matrix. These principle axes here preserve 64% of the genetic variation among the 55 HLA alleles. Superimposed on this three-dimensional representation are lines connecting a minimum-spanning tree, in which tree a line joins a population to its nearest genetic neighbor. To show both the minimum-spanning tree of African populations without the American blacks and the position of the American blacks relative to the other populations, two figures—figures 2 and 3—are presented.

The Japanese are clearly separated from the Europeans, west and south Africans, and the San. The Zimbabweans are centrally placed between the South

Table 3

Genetic Distances between Various Population Groups, calculated by Chord Method of Cavalli-Sforza and Edwards (1967), for Five Largest Ethnic Groups in Table 1 and for Other African and Non-African Populations in Table 2

	Genetic Distance												
	Malians	Nigerians	South Africans	Europeans	Japanese	Fula	Manjago	Jola	Mandinka	Wolof	Zimbabweans	San	American Blacks
Malians	. 000												
Nigerians	2.158	.000											
South Africans	2.028	1.270	.000										
Europeans	1.665	2.756	2.495	.000									
Japanese	3.288	4.861	5.043	2.036	.000								
Fula	1.222	1.417	1.389	1.432	3.732	.000							
Manjago	1.299	1.559	1.702	1.736	3.612	.520	.000						
Jola	1.508	1.644	1.757	1.819	3.770	.550	.557	.000					
Mandinka	1.130	1.379	1.258	1.685	3.970	.301	.307	.306	.000				
Wolof	1.223	1.510	1.202	1.553	3.988	.306	.510	.461	.210	.000			
Zimbabweans	1.685	.843	.396	2.417	4.858	1.238	1.436	1.462	.974	.970	.000		
San	3.246	2.381	.972	3.251	6.081	2.275	2.822	2.753	2.416	2.305	1.672	.000	
American blacks	1.397	.787	.905	1.201	3.464	.621	.803	.939	.567	.628	.669	2.104	.000

Africans and the Nigerians, consistent with their geographical position. The San are more informatively represented in this analysis, because, although figure 1 emphasizes their divergence from all other populations (as further indicated by their corner location in figs. 2 and 3), the minimum-spanning tree indicates that they are, nonetheless, most closely related to the Xhosa, with whom there is known to have been admixture (Nurse et al. 1985). The lack of affinity between the San and the European Caucasian population, evident in both figure 2 and table 3, is not supportive of the suggestion of Cavalli-Sforza et al. (1988) that the San are a hybrid of Caucasian and other African populations.

The Gambian ethnic groups cluster together in the diagram and, with Mali, are separated from the Bantu African populations and from the San. The minimum-spanning tree shows that, of the Gambian ethnic groups, the Fulas are most closely related to Europeans, which appears consistent with their proposed north African origin and more caucasoid appearance. Furthermore, the Mandinkas are the ethnic group closest to the Malian population, compatible with historical evidence of their origin from a province in the ancient kingdom of Mali. Mali is positioned close to the Europeans, perhaps reflecting gene flow from north African Caucasian populations. In figure 3 the American black population has been added into the analysis. As might have been expected from their known origins in west Africa, they are placed intermediately between the coastal west Africans (from Mali and The Gambia) and the Nigerian population. Furthermore, they are the black population most closely related to Europeans, presumably reflecting genetic admixture, estimated at 20% (Reed 1969), from Caucasians.

Discussion

The present analysis of HLA class I antigens shows that the major tribes within The Gambia are all more closely related both to each other and to the nearby Malians than to other African populations. Furthermore, although the interethnic differences are small, the directions in which these differences occur in the distance analysis are consistent with available evidence on the population history of these ethnic groups (Sonko-Godwin 1985). For instance, the Mandinkas are the group most closely related to the population of Mali, and the Fulas are the group most closely related to Caucasians. Also, of the four large ethnic groups surveyed, the Fulas and Jolas are historically and culturally the most distant and distinctive, and this is reflected by their large interethnic genetic distance (table 3).

At a continental level, the representation of the genetic distances among national groups is also consistent with their known geographical, historical, and linguistic relationships. A well-established classification of ethnic groups in Africa is that based on linguistic data. Greenberg (1963) grouped African languages into four phyla: Niger-Congo, Afroasiatic, Nilo-Saharan, and Khoisan. While Afroasiatic and Nilo-Saharan languages are mainly north African, Khoisan is confined almost exclusively to parts of southern Africa. In west Africa, west Atlantic and Mande languages, including most of those in The Gambia, are separated from the Bantu group within the large Niger-Congo phylum. Bantu languages, despite their enormous area of distribution encompassing the Nigerian, Zimbabwean, and South African populations analyzed here, show a relatively strong degree of similarity with one another. The HLA analysis we present agrees with these broad linguistic divisions, by grouping the San, the Bantu-speakers, and the non-Bantu west Africans separately. In contrast, an earlier distance analysis, performed on the data available to Excoffier et al. (1987, fig. 9), found the HLA system "not very informative." The present analysis, using 55 HLA class I allele frequencies compared with the 26 of the earlier study, provides an account of population relationships that is highly plausible both when one compares African populations in different countries and when one compares African populations within a country.

Genetic markers for which numerous sub-Saharan Africa populations have been adequately surveyed are relatively few (reviewed by Excoffier et al. 1987, 1991). The clinal variation in hemoglobinopathies in Africa (Lefevre-Witier 1985) reflects strong malarial selection and is of limited value in comparing populations. However, β-globin gene haplotypes linked to the sickle mutation have been used to separate Atlantic coastal west African populations from those of Benin and Nigeria, as well as from those of central Africa (Pagnier et al. 1984). The clustering, in this analysis, of the west African populations of The Gambia together with Mali and separate from Nigeria is concordant with this view. The most informative comparison of African populations for a single genetic system has been reported recently by Blanc et al. (1990). They reported frequencies of immunoglobulin (GM) allo-

types in three ethnic groups (Bendink, Fulani, and Mandenkalu) of eastern Senegal and compared these with 15 other sub-Saharan African populations by using principal components analysis (the Fulani and Mandenkalu of eastern Senegal are very closely related to the Fula and Mandinka of The Gambia, respectively). Both the close clustering of the local ethnic groups in the Senegambia in the GM analysis and their more distant relationship to the Bantu populations of Africa are similar to the HLA genetic distance comparison, with the San again being the most distinctive of the African populations. Unfortunately, there is no serologically detailed set of HLA frequencies for any east African population available: in the immunoglobulin study, east Africans were substantially different from west Africans, Bantu, and San. Our HLA data are, however, supportive of the general conclusion of Blanc et al. (1990)—i.e., that the sociocultural differences of ethnic groups in this region of west Africa have had relatively little effect on local genetic diversity. Despite linguistic differences and the apparent cultural segregation of the various ethnic groups, the genetic distances between neighboring populations of different ethnic groups are very small.

There are several theoretical problems in using serological HLA frequencies as an indicator of population affinities. One of these is the presence of antigenic subtypes which are not serologically distinguishable together with some serologically undetectable alleles (blanks). Although, as in other regions, HLA-C blanks were very common, blank HLA-A and -B frequencies in most Africans are fairly low, so that definition of these should not alter genetic distances substantially. The existence of subtypes presents a different problem, in that different alleles are grouped as one in the serological classification. There is preliminary evidence that distinctive prevalences of antigenic subtypes may be a feature of African HLA class I specificities. For example, we have found, using molecular methods, that most Gambians with HLA-B27 have a subtype, B*2703, which is not found in nonblack populations (Hill et al. 1991*a*). The extent to which this compromises serological frequency comparisons remains to be seen. It seems possible that grouping of molecular subtypes as one serological specificity may be more justifiable if the subtypes represent recent local differentiation from a common ancestral allele. Hence, although definition of HLA class I subtypes will probably show further genetic diversity within Africans and allow better resolution of genetic affinities, this may not change substantially the overall pattern of genetic relationships.

A second potential problem is that the four largest Gambian ethnic groups in the present study are known to have had different population structures even in recent times (Sonko-Godwin 1985), which of itself may lead to differential changes in HLA frequencies. It is very difficult to assess the importance of this factor in producing genetic differentiation. However, the same problem exists to an even greater degree in the island populations of the Pacific, which have experienced several bottlenecks during island colonizations and which exhibit a variety of population structures. Nonetheless, HLA was the first genetic locus to provide a detailed picture of their genetic affinities (reviewed by Serjeantson 1989) which is concordant with a variety of independent evidence.

Third, the employment of HLA markers in population comparisons has frequently been criticized because of the probable importance of natural selection in altering HLA allele frequencies. Indeed, recent studies have confirmed the importance of selection in maintaining MHC diversity (Hedrick and Thomson 1983; Hughes and Nei 1988), but they also indicate that the magnitude of most of these selective effects may be very small in comparison with selection for sickle hemoglobin (Hill et al. 1991*b*). So, just as several globin alleles which are subject to selection by malaria have been useful in anthropological comparisons, it appears that the magnitudes of geographical differences in selection pressures on particular HLA alleles are insufficient to obscure the value of HLA frequencies in measuring population affinities. Hence, despite the potential confounding effects of natural selection and differences in population structure, HLA class I frequencies are able to provide a highly plausible picture of population affinities in sub-Saharan Africa, and analysis of further populations in this region should be valuable.

Acknowledgments

We thank A. Hughes, P. Twumasi, L. Bayo, L. Manneh, and E.G. Sarr for assistance with sample collections; I. Sambou, F. Sisay, and H. Sillah for laboratory assistance; and A. Ting for facilities. A.V.S.H. is a Wellcome Trust Senior Clinical Fellow.

References

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

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

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Hum Genet 19:233–257

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

Crowder M (1977) West Africa, an introduction to its history. Longman, London

Dausset J, Colombani J (1972) Histocompatibility testing. Munksgaard, Copenhagen

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

du Toit ED, MacGregor KJ, Taljaard DG, Oudshoorn M (1987) HLA-A, -B, -C, -DR and -DQ polymorphisms in three South African population groups: South African negroes, cape coloureds and South African caucasoids. Tissue Antigens 31:109–125

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

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

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

Greenberg JH (1963) The languages of Africa. Mouton, The Hague

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

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Greenwood BM, McMichael AJ (1991*a*) HLA class I typing by PCR: HLA-B27 and an African B27 subtype. Lancet 337:640–642

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

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

Johnson A (1986) Joint report: American Blacks. In: Aizawa M (ed) HLA in Asia-Oceania 1986. Hokkaido University Press, Sapporo, pp 320–323

Jorde LB (1985) Human genetic distance studies: present status and future prospects. Annu Rev Anthropol 24: 343–373

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

Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum Immunol 19:155–162

Lee TD, Lee A, Shi WX (1991) HLA-A, -B, -DR, -DQ antigens in black North Americans. Tissue Antigens 37:79–83

Lefevre-Witier PH (1985) An anthropological perspective on the epidemiology of hemoglobin defects and G-6-PD deficiencies in the northern half of the African continent. In: Chakraborty R, Szthmary EJE (eds) Diseases of complex etiology in small populations: ethnic differences and research approaches. Alan R Liss, New York, pp 365–397

Nei M, Livshits G (1989) Genetic relationships of Europeans, Asians and Africans and the origins of modern *Homo sapiens*. Hum Hered 39:276–281

Nurse GT, Weiner JS, Jenkins T (1985) The peoples of Southern Africa and their affinities. Clarendon, Oxford

Okoye RC, Williams E, Alonso A, Doyle P, Awad J, Navarrete C, Jaraquemada D, et al (1985) HLA polymorphism in Nigerians. Tissue Antigens 25:142–155

Pagnier J, Mears JG, Dunda-Belkhodja O, Schaefer-Rego KE, Beldjord C, Nagel RL, Labie D (1984) Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. Proc Natl Acad Sci USA 81:1771–1773

Phillipson DW (1985) African archaeology. Cambridge University Press, Cambridge

Reed TE (1969) Caucasian genes in American Negroes. Science 165:762–768

Rohlf FJ (1990) NTSYS-pc, numerical taxonomy and multivariate analysis system: technical report. State University of New York, Stony Brook

Serjeantson SW (1989) HLA genes and antigens. In: Hill AVS, Serjeantson SW The colonization of the Pacific: a genetic trail. Oxford University Press, Oxford, pp 120–173

Serjeantson SW, Ryan DP, Thompson AR (1982) The Colonization of the Pacific: the story according to human leukocyte antigens. Am J Hum Genet 34:904–918

Sokal RR, Rohlf FJ (1981) Biometry, 2d ed. Freeman, San Francisco

Sonko-Godwin P (1985) Ethnic groups of the Senegambia: a brief history. Book Production and Material Resources Unit, Banjul, The Gambia

Terasaki PI (ed) (1980) Histocompatibility testing 1980. UCLA Tissue Typing Laboratory, Los Angeles

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