CORE
Metadata, citation and similar papers at core.ac.uk
Provided by NSU Works

NSU Florida
NOVA SOUTHEASTERN UNIVERSITY

Nova Southeastern University
NSUWorks

Biology Faculty Articles

Department of Biological Sciences

2-2004

D1S80 Single-Locus Discrimination Among African Populations

Rene J. Herrera
Florida International University

Leslie R. Adrien
Florida International University

Luis M. Ruiz
Florida International University

Nahir Y. Sanabria
Florida International University

George Duncan
Broward County Sheriff's Office, gduncan@nova.edu

Follow this and additional works at: https://nsuworks.nova.edu/cnso_bio_facarticles

 Part of the Biology Commons

NSUWorks Citation

Herrera, Rene J.; Leslie R. Adrien; Luis M. Ruiz; Nahir Y. Sanabria; and George Duncan. 2004. "D1S80 Single-Locus Discrimination Among African Populations." *Human Biology* 76, (1): 87-108. doi:10.1353/hub.2004.0022.

This Article is brought to you for free and open access by the Department of Biological Sciences at NSUWorks. It has been accepted for inclusion in Biology Faculty Articles by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

PROJECT MUSE®

D1S80 Single-Locus Discrimination Among African Populations

Rene J. Herrera, Leslie R. Adrien, Luis M. Ruiz, Nahir Y. Sanabria, George Duncan

Human Biology, Volume 76, Number 1, February 2004, pp. 87-108 (Article)

Published by Wayne State University Press
DOI: *https://doi.org/10.1353/hub.2004.0022*

For additional information about this article
https://muse.jhu.edu/article/53386

Access provided by Nova Southeastern University (15 Mar 2018 19:28 GMT)

Project MUSE®
http://muse.jhu.edu

D1S80 Single-Locus Discrimination Among African Populations

RENE J. HERRERA,[1] LESLIE R. ADRIEN,[1] LUIS M. RUIZ,[1] NAHIR Y. SANABRIA,[1] AND GEORGE DUNCAN[2]

Abstract The highly polymorphic D1S80 locus has no known genetic function. However, this variable number of tandem repeats (VNTR) locus has been highly valuable in forensic identification. In this study we report the allele and genotype frequencies of five African populations (Benin, Cameroon, Egypt, Kenya, and Rwanda), which can be used as databases to help characterize populations and identify individuals. The allele frequencies were used to infer genetic associations through phylogenetic, principal component, and *G* test statistical analyses. Compliance with Hardy–Weinberg equilibrium expectations was determined as were F_{ST} estimates, theta *p* values, and power of discrimination assessment for each population. Our analyses of 28 additional populations demonstrate that the D1S80 locus alone can be used to discriminate geographic and ethnic groups. We have generated databases useful for human identification and phylogenetic studies.

The D1S80 locus is found in the telomeric region of the p arm of chromosome 1 at position 1p36–p35. Since it was first characterized (Nakamura et al. 1988), the D1S80 locus has been widely used in phylogenetic studies, forensic analysis, and paternity testing. It has been described as the most characterized amplified fragment length polymorphism (AmpliFLP) locus (Budowle et al. 1996). The D1S80 marker locus has a variable number of tandem repeats (VNTR) of a 16-nucleotide unit and shows a high degree of polymorphism. The observed heterozygosity for this locus has been reported to be as high as 87.6% (Budowle et al. 1997; Duncan 1996). A previous study has shown that the D1S80 locus alone allows for the discrimination of geographic and ethnic groups (Duncan et al. 1996).

The patterns of genetic variation in African populations are multifaceted and complex. Interactions between the Arab world and African civilizations in North, East, and West Africa as well as the Bantu expansion, which changed the demographic and language map of central and southern Africa, are partially responsible for the genetic diversity of these groups (Holden 2002). In this study

[1]Department of Biological Sciences, Florida International University, University Park, Miami, FL 33199.
[2]Broward County Sheriff's Office, Crime Laboratory, Fort Lauderdale, FL.

Human Biology, February 2004, v. 76, no. 1, pp. 87–108.
Copyright © 2004 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: D1S80, AMPLIFLPS, PCR, AFRICAN POPULATIONS, BENIN, CAMEROON, EGYPT, KENYA, RWANDA, PHYLOGENETIC ANALYSIS, POPULATION GENETICS, VARIABLE NUMBER OF TANDEM REPEATS (VNTR)

we examined five African populations and their phylogenetic relationships in relation to non-African populations according to the D1S80 locus. DNA samples were collected from individuals of Bantu tribes in Cameroon, Kenya, and Rwanda (Hutu) and the Fon ethnic group of Benin. Collectively, sub-Saharan populations are known for their high levels of genetic diversity (Cavalli-Sforza et al. 1994). The fifth African group sampled is from the Tanta region of Egypt and is mainly made up of Arabs and Berbers (Harich et al. 2002). The majority of all sub-Saharan Africans belong to Bantu-speaking populations (Williamson and Blench 2000). Bantu is a large group of about 450 related languages (Guthrie 1971).

The current location of Bantu-speaking populations reflects the spread of farming during the agricultural revolution, which started relatively late in sub-Saharan Africa, about 3000 B.C., and extended to A.D. 500 (Huffman 1982). Archeological and linguistic data support the notion that the major subgroups of modern Bantu populations and their geographic location stem from Neolithic times (Holden 2002). Likely locations for the ancestors of Bantu-speaking populations include present-day Benin and Cameroon (Vansina 1990). One branch of the Bantu dispersal moved in a southeastern direction and reached the Lake Victoria area, located in present-day Kenya and Rwanda, during the last century B.C. (Schmidt 1975). Limited genetic information mirrors the more extensive linguistic and archeological data in terms of the origin and migration patterns of this major human diaspora (Holden 2002; Pereira et al. 2002). The populations examined in this study were sampled from geographic regions representing different stages of the Bantu expansion.

The geographic regions sampled in this study are ethnically heterogeneous. For example, Benin is currently populated by 42 separate ethnic groups (Levinson 1998), and Cameroon has more ethnic groups (approximately 200) than any other country in Africa (Levinson 1998). Kenya, on the east coast of Africa, has been influenced historically by several Arab groups, especially along its coastline. There are more than 100 different ethnic groups established in Kenya (Levinson 1998). In addition, Arabs, Asians, and Europeans make up less than 1% of the total population of Kenya (Maxon 2002). The Egyptian Arab and Berber samples were collected from the Nile delta in the region of Tanta. Berbers were the earliest known inhabitants of North Africa (Clancy-Smith 1997). Subsequently, different groups of people, including Phoenicians, Greeks, and Romans, were attracted to the area and populated the fertile Nile delta region (Arnaiz-Villena et al. 2002). Since A.D. 711, Arabs have occupied the region and have admixed with the Berbers (Clancy-Smith 1997).

In this study we report for the first time the allele frequencies of the VNTR D1S80 locus of African groups from Benin, Cameroon, Egypt, Kenya, and Rwanda and compare them with 28 other worldwide reference populations. Allele and genotype frequencies were determined. Compliance with Hardy–Weinberg equilibrium expectations was ascertained. F_{ST} estimates and power of

discrimination values were calculated for each of the five populations. In addition, the 28 reference populations were incorporated into our study to ascertain genetic similarities. Allele frequencies were used to infer genetic associations through phylogenetic, principal components, and *G* test statistical analyses. Although a single locus may not be representative of genomewide characteristics of populations, the highly polymorphic D1S80 marker generated robust, statistically significant information consistent with data based on numerous and varied genetic systems. In addition, the D1S80 system provides for both rapid and inexpensive phenetic information.

Materials and Methods

Collection of Samples and DNA Isolation. All samples were procured from unrelated individuals and consisted of whole blood collected in Vacutainer tubes containing EDTA. The individuals were identified by biographical information traced back at least two generations. Each collection was arranged through the leaders of the regions and supervised by the same. Samples were collected according to ethical guidelines, as outlined by the Florida International University Institutional Review Board. The blood samples were lysed, and leukocyte nuclei were separated from the rest of the blood components using a previously reported method (Antunez-de-Manolo et al. 2002). The DNA was then purified using proteinase-K digestion and standard organic phenol–chloroform extraction (Novick et al. 1995). All samples were stored at $-$ 80°C when not in use.

Amplification of DNA. PCR was performed in a Perkin-Elmer 480 thermal cycler. Amplification parameters and the sequence of the forward and reverse primers have been previously described (Kasai et al. 1990). PCR amplifications were carried out in 25-μl reactions containing 16.3 μl of water, 2.5 μl of 10X buffer with 15 m*M* $MgCl_2$, 2.5 μl of 0.15 m*M* dNTPs, 1.25 μl of 10 μM of each primer, 0.2 μl (5 U) Taq DNA, and 10–50 ng of DNA sample. The PCR parameters were previously described by Duncan et al. (1996). The Benin samples were pre-amplified with DOP (degenerate oligonucleotide primer), as previously described (Buchanan et al. 2000). This method of PCR enriches high-molecular-weight DNA and is particularly useful when a DNA sample is of poor quality or low yield. To perform the D1S80 amplification reaction, we used 6 μl of the DOP PCR product.

Vertical Polyacrylamide Gel Electrophoresis. A 39:1 acrylamide/bis acrylamide stock solution was used to make 8% polyacrylamide 1X TBE gels. Gels were 35 $\times$ 45 cm in size and 0.8 mm thick. An Applied Biosystems AmpliFLP D1S80 allelic ladder was included every four samples as a size marker to allow the genotyping of samples. The gels were electrophoresed at 1000–1200 V for 5–6 hr, depending on the migration of the xylene–cyanol tracking dye. All gels

were silver stained to visualize the alleles. Silver staining was performed according to methods previously reported (Allen et al. 1989).

Sampled Populations. All the sampled sub-Saharan populations belong to the Niger–Congo language family (Cavalli-Sforza et al. 1994). The population sample from Benin is made up of 100 individuals from the Fon ethnic group. They were collected in the southern part of the country from the Zagnanado population. The Kenya group consists of 106 Bantu individuals from small villages 100 km northeast of Nairobi. The Rwanda population is composed of 100 individuals of the Hutu tribe, a Bantu group. The Egyptian group was collected in the region of Tanta, located at the center of the Nile River delta, and it is made up of 100 individuals. This population is primarily Arab and Berber in origin. Arabian and North African populations, like the Egyptians, speak languages that belong to the Afro-Asiatic family (Murdock 1959). Also, this North African Egyptian group may share ancestry, to a lesser extent, with Greeks, Turks, and other Mediterranean basin people. The Nile River delta population is mainly Caucasian in origin. The Cameroon group consists of 16 Bantu individuals from villages 50 km southwest of the city of Yaounde, in the southern region of the country. Another 28 worldwide populations were analyzed along with our 5 African populations to ascertain genetic relationships. Allele frequencies from these additional 28 groups were obtained from the literature, as indicated in Table 1.

Statistical Analysis. The gene-counting method (Li 1976) was used to generate allele and genotype frequencies. Allele frequencies were analyzed using PHYLIP 3.5 (Felsenstein 1993), the Numerical Taxonomy and Multivariate Analysis System (NTSYS) Principal Component program (Rohlf 2002), and a *G* test program (Carmody 1990). The maximum-likelihood tree option in PHYLIP 3.5 was used to visualize the general phylogenetic relationships among populations. A large number (1000) of bootstrapped data sets generated by Seqboot was analyzed using the CONTML program. Bootstrap analysis provides a conservative test of the level of support for each node on the tree (Holden 2002). The CONTML and Consense programs configured the best-fitting tree directly from the allele frequencies.

The principal-components (PC) test was performed to generate two-dimensional plots of PC 1 and PC 2. The principal-components analyses, based on the relevant population frequencies, were performed by means of the NTSYS program. The *G* test generates a 2×2 contingency table based on the observed and expected values for each pairwise comparison (Rangel-Villalobos et al. 1999). One thousand simulations were done for each pairwise comparison. The *G* test ascertains which pairs of populations are homogeneous with each other.

Expected heterozygosity values were compared to those reported in the literature for other populations. Conformance with Hardy–Weinberg equilibrium expectations was tested by using the Fisher exact test and the chi-square test within the Genetic Data Analysis (GDA) program (Lewis and Zaykin 2001). The

Table 1. Populations Studied

Population	*Code*[a]	*N*	*Location*	*Reference*
Andalucia	Anc	120	Southern province of Spain	Lorente et al. (1997)
Andalucia2	An2	147	Southern province of Spain	Flores et al. (2001)
Arab Moslems	Ara	94	Gaza Strip, Judea, Samaria, Israel	Peterson et al. (2000)
Australia	Aus	250	Victoria	Gutowski et al. (1995)
Bahama	Bah	88	Throughout Bahamas	Duncan et al. (1996)
Bari	Bar	24	Northeast Colombia, South America	Duncan et al. (1996)
Basque	Bas	257	North-central Spain	Peterson et al. (2000)
Benin	Ben	100	South Benin, West Africa	Present study
BWH Alaska	BWH	109	Bethel-Wade Hampton, Alaska	Walkinshaw et al. (1996)
Cameroon	Cam	34	Yaounde City	Araújo Da Silva et al. (1999)
Cameroon2	Ca2	16	Southwest of Yaounde City, West Africa	Present study
Canary Islands	Can	123	General population of Canary Islands	Flores et al. (2001)
Chimila	Chm	46	Northeast Colombia, South America	Duncan et al. (1996)
China (Han)	Chn	216	Xian and Shijiazhuang districts of China	Peterson et al. (2000)
Congo	Con	34	Lubumbashi City	Araújo Da Silva et al. (1999)
Denmark	Den	210	General population of Denmark	Peterson et al. (2000)
Dubai Arabs	Dub	93	United Arab Emirates	Alkhayat et al. (1996)
Egypt (Tanta)	Tan	100	Tanta, North Africa	Present study
Galicia	Gal	149	Northwest province of Spain	Peterson et al. (2000)
Greece	Gre	107	Cyprus	Cariolou et al. (1998)
Haiti	Hai	83	Caribbean	Peterson et al. (2000)
Kenya	Ken	106	Northeast of Nairobi, East Africa	Present study
Korea	Kor	116	Seoul	Peterson et al. (2000)
Mapuche, Argentina	Map	61	Argentina, South America	Hutz et al. (1997)
Navajo	Nav	28	New Mexico, USA	Duncan et al. (1996)
Nigeria	Nig	67	West Africa	Peterson et al. (2000)
North Slope Alaska	NSA	92	North Slope Borough, Alaska	Walkinshaw et al. (1996)
Philippines	Phi	103	Metro Manila	Halos et al. (1999)
Rwanda	Rwa	100	East Africa	Present study
Saudi Arabia	Sau	220	Riyadh	Tahir et al. (2000)
Taiwan (Han)	Tai	105	General population of Taiwan	Peterson et al. (2000)
Turkey	Tur	112	General population of Turkey	Cakir et al. (2001)
Zimbabwe	Zim	101	Mashonaland Province	Peterson et al. (2000)
Total (33 populations)		3611		

a. These codes are used in Table 5.

Table 2. D1S80 Allele Frequencies

D1S80 Allele	*Benin (N = 100)*	*Kenya (N = 106)*	*Rwanda (N = 100)*	*Egypt (N = 100)*	*Cameroon2 (N = 16)*
14	–	–	–	–	–
15	–	–	–	–	–
16	–	–	0.0100	–	–
17	0.0550	0.0283	0.1050	0.0150	0.0625
18	0.0550	0.0613	0.0900	0.1400	–
19	–	–	0.0050	0.0050	–
20	0.0550	0.0236	0.0100	0.0150	0.0313
21	0.1100	0.0991	0.1000	0.0300	0.1563
22	0.1000	0.0896	0.1200	0.0450	–
23	0.0300	0.0283	0.0050	0.0350	0.0313
24	0.2150	0.2028	0.2200	0.4150	0.2188
25	0.0650	0.0377	0.0300	0.0600	0.0625
26	0.0050	0.0142	–	0.0300	–
27	0.0100	0.0236	0.0550	0.0200	–
28	0.1100	0.1981	0.0700	0.0650	0.0313
29	0.0200	0.0330	0.0100	0.0500	–
30	0.0100	–	–	0.0050	–
31	0.0800	0.0330	0.0400	0.0400	0.0313
32	0.0050	0.0283	0.0200	0.0050	–
33	–	0.0047	–	0.0050	0.0313
34	0.0750	0.0802	0.1000	0.0200	0.3125
35	–	–	–	–	–
36	–	–	–	–	0.0313
37	–	–	–	–	–
38	–	–	–	–	–
39	–	–	–	–	–
40	–	0.0142	0.0100	–	–
41	–	–	–	–	–

"–" indicates allele not detected.
N is total number of individuals.

GDA program was also used to calculate theta *p* values, which are equivalent to F_{ST} values (Lewis and Zaykin 2001). Theta *p* values have been used to correct for the effects of subpopulation structure (Monson and Budowle 1998). This test assumes no mutation rate and symmetric migration. If these assumptions are met, the values range between 0 and 1. Power of discrimination values were estimated according to the method previously described by Saferstein (1982).

Results

The populations examined in this study are listed in Table 1. Table 1 also includes the geographic locations and references for each population. Table 2 lists the allele frequencies for the five African populations sampled and analyzed in this study. *D1S80*24* is the most frequent allele in the Benin, Kenya, Rwanda,

Table 3. Theta *p* Values and Observed and Expected Heterozygosity Values

Population	*N*	*Theta p Values*	*Observed Heterozygosity (%)*	*Expected Heterozygosity, Unbiased (%)*
Benin	100	0.032	85	89.72
Kenya	106	0.027	93.4	88.82
Rwanda	100	0.029	85	89.13
Egypt	100	0.009	72	79.38
Cameroon	16	0.021	81.3	84.27

and Egypt populations. In contrast, in the Cameroon population the **34* allele is the most common (0.313) followed by allele **24* (0.219), which has a frequency similar to the other four African populations studied. Previous studies have shown that alleles **24, *28,* and **34* are the most frequent among African groups (Duran and Ruiz-García 2001).

The observed heterozygosities for Benin, Cameroon, Kenya, Rwanda, and Egypt (Tanta) are 0.8500, 0.8130, 0.9340, 0.8500, and 0.7200, respectively (Table 3). These high levels of heterozygosity are consistent with the reported elevated genetic diversity within sub-Saharan African populations at other loci (Watkins et al. 2001). The level of observed heterozygosity in non-sub-Saharan African populations previously examined for the D1S80 locus ranges from 0.458 in Bari Native Americans to 0.806 in southeastern Hispanics (Budowle et al. 1996; Duncan 1996); these values are lower than those in the sub-Saharan African groups. The theta *p* values for the Benin, Cameroon, Kenya, Rwanda, and Egypt populations are 0.032, 0.021, 0.027, 0.029, and 0.009, respectively.

Figure 1 shows a radial representation of a maximum-likelihood tree. All of the African groups and all the populations of African descent (Bahamas and Haiti) cluster together in a major clade. Populations of African descent and the Tanta collection from Egypt, which is known to have experienced gene flow from sub-Saharan regions, mainly by way of the Nile waterway, group closer to the two Caucasian clusters (i.e., the Middle East–Southwest Asian clade and the European group of populations). The East Asian and Native American groups make up the next major clade. In this major clade the East Asian groups segregate away from the Native Americans and the insular East Asians segregate from the continental East Asian populations. The bootstrap values, the results of a conservative test of support, indicate the likelihood that the depicted relationships are statistically significant. Bootstrap values greater than 50% (indicated as 500 or greater in Figure 1) are considered indicators of significant phylogenetic relationships. All bootstrap values, except two 48% bifurcations, are above 50%. Most of the bootstrap values are above 90%.

Figures 2–7 illustrate a series of principal-components analyses. In Figure 2, which includes all worldwide populations, PC 1 (depicted along the *x*-axis) represents 21% of the variability, whereas PC 2 (reflected along the *y*-axis) accounts for 14% of the variability. PC 1 shows delineation between African populations and all other populations (see Figure 2). PC 2 separates East African

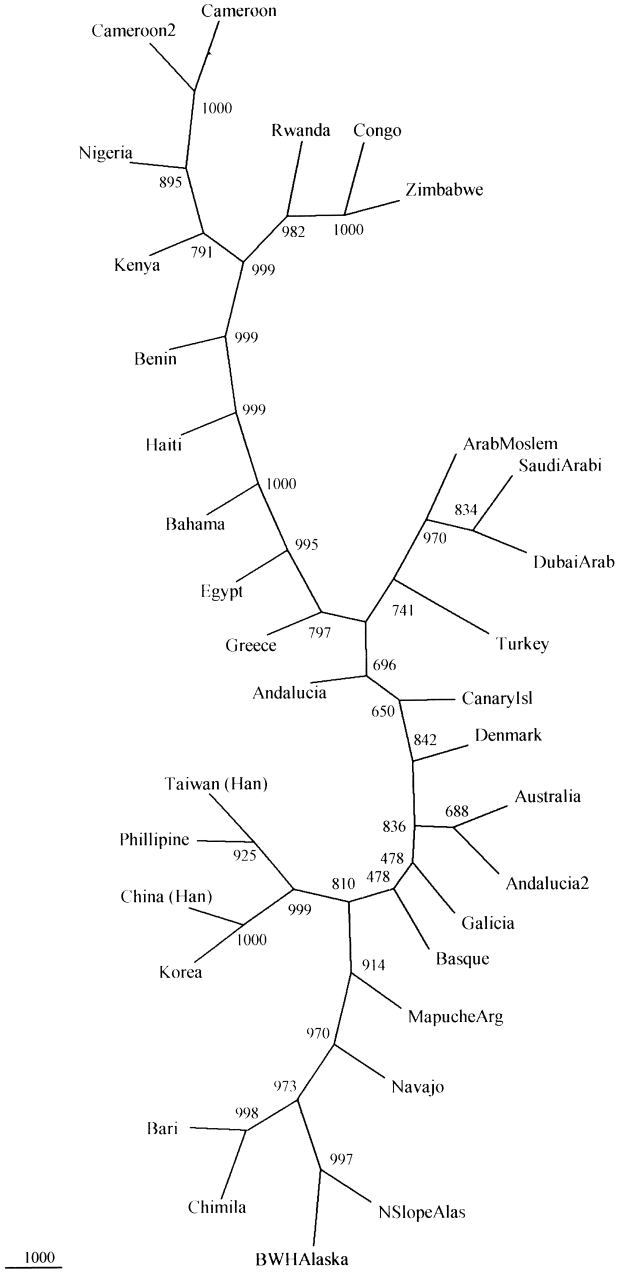

Figure 1. Maximum-likelihood tree with bootstrap values. This tree was generated directly from the allele frequency data of the D1S80 locus from 33 populations using PHYLIP 3.5c (Felsenstein 1993). The percentages at each bifurcation represent bootstrap values determined with the Seqboot and Consense options in the PHYLIP 3.5c program.

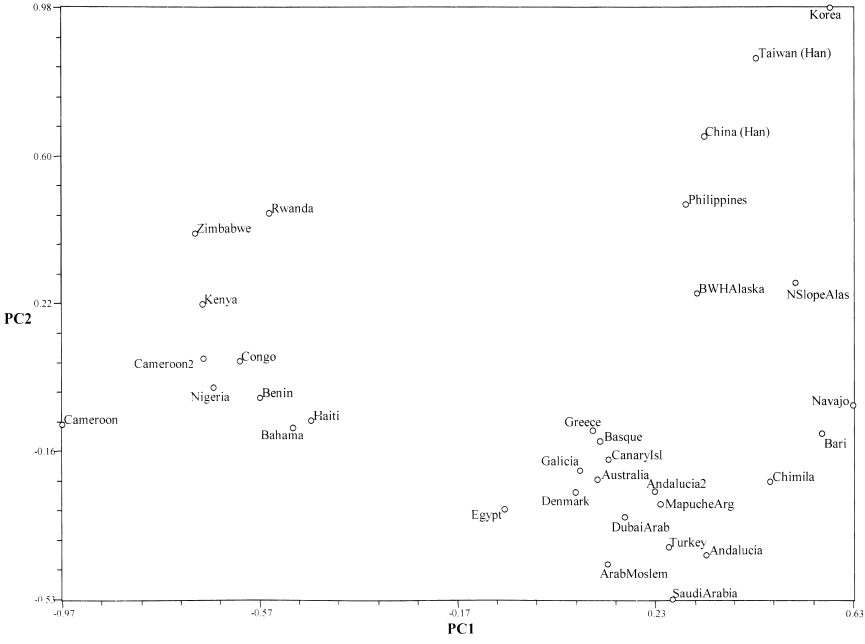

Figure 2. Principal-components plot from an analysis performed on 33 worldwide populations. PC 1 and PC 2 represent the first and second principal-component values, respectively, for each population.

populations from Central and West African groups. Admixed populations of African descent group at the fringes of the African cluster that is closer to the West African groups in the direction of the Caucasian populations. This is expected because populations of African descent in America are of West African ancestry, with contributions from populations of European descent. PC 2 also segregates the East Asians from the Native American groups. Caucasians separate from the other groups along PC 1 and PC 2. Although the Caucasian populations cluster closer together than any of the major geographic groups, PC 2 allows for the segregation of the Middle East–Southwest Asian groups from the Europeans. Of the three major geographic groups, the Orientals, represented by East Asian and Native American groups, form a more diffuse assembly than the Caucasian and sub-Saharan African clusters. The Egyptian population plots away from the Caucasian cluster in the direction of the sub-Saharan African populations, as expected because of gene flow from the south.

Figure 3 illustrates a principal-components analysis of the worldwide populations depicted in Figure 2 minus the East Asian and Native American groups. Here, PC 1 represents 28% and PC 2 14% of the variability. The subtraction of these two population groups allows for a greater separation between the sub-Saharan groups, although the geographic partitioning observed in the totality of the populations is compromised; the West African populations of Congo and

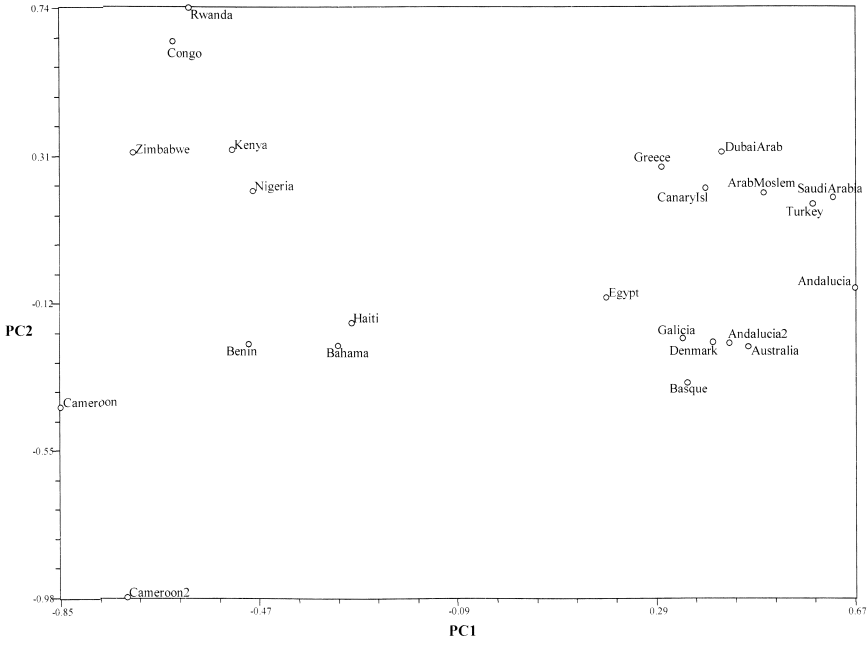

Figure 3. Principal-components plot from an analysis excluding Oriental groups, that is, Native American (Bari, Chimila, Mapuche, Navajo, and Alaskan Eskimos) and East Asian (China, Korea, Philippines, and Taiwan) populations.

Nigeria now segregate with the East African populations (Figure 3). On the other hand, within the Caucasian cluster a clearer segregation of Middle East–Southwest Asian groups from European populations can now be observed.

When only the sub-Saharan African populations are considered, PC 1 and PC 2 represent 28% and 23% of the variability, respectively (Figure 4). In Figure 4 an even greater separation between the populations is observed. As in Figure 3, the geographic partitioning of East from West African groups is not as well delineated as in the principal-components analysis with all the populations (Figure 2). Deleting the Caucasian groups from the principal-components analysis has the effect of aggregating all sub-Saharan populations into a compact cluster in which East and West African populations are not separated (Figure 5). The variability provided by Figure 5 is 26% for PC 1 and 19% for PC 2.

In Figure 6 the Native American groups were excluded from the analysis. The variability reflected in PC 1 and PC 2 is 26% and 17%, respectively. In this figure the separation among sub-Saharan populations is comparable to the one provided by the principal-components analysis of all populations (Figure 2). The one exception is that the East African population from Kenya is within the West African cluster. When the East Asian populations are excluded from the principal-components analysis, the sub-Saharan populations do not segregate into geographic zones (Figure 7). The East African groups from Rwanda and Zimbabwe

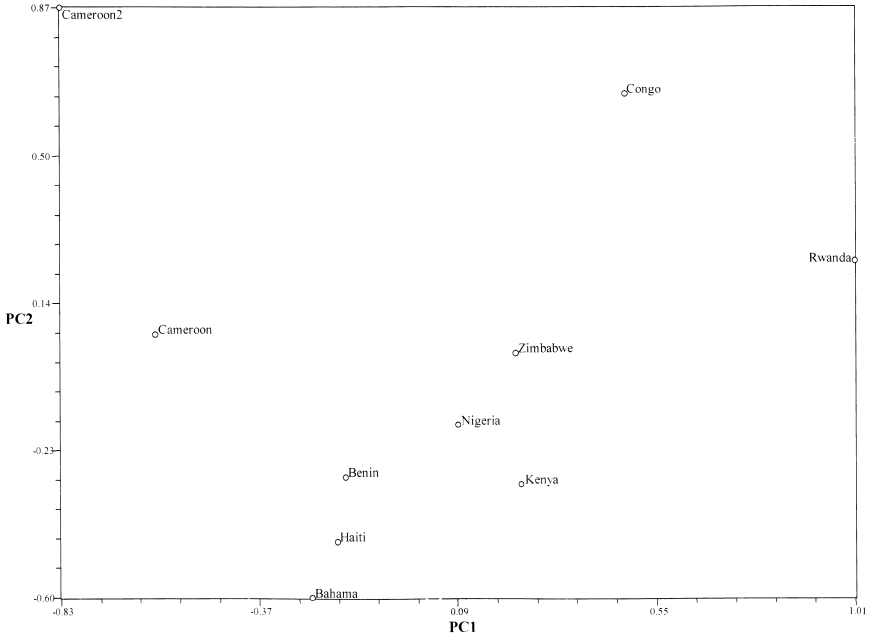

Figure 4. Principal-components plot from an analysis of African populations (Benin, Cameroon, Cameroon2, Congo, Kenya, Nigeria, Rwanda, and Zimbabwe) and groups of African descent (Bahamas and Haiti).

segregate in between the two Cameroon populations from West Africa, and the East African Kenyans cluster with the West African populations. Similarly, among the Caucasian groups, the Middle Eastern population from Dubai groups away from the other geographically close populations and with the Europeans.

The data presented in Table 4 indicate the results of tests for adherence to Hardy–Weinberg equilibrium expectations. The Benin population did not conform to Hardy–Weinberg equilibrium according to the Fisher exact test ($p = 0.019$). The high theta p value for Benin may indicate subpopulation structure, which could also be responsible for the population's violation of Hardy–Weinberg equilibrium. The Rwanda population was also shown to be marginally out of Hardy–Weinberg equilibrium by the chi-square test, with a p value of 0.036.

The G test performed on the 33 different populations (Table 5) provided data consistent with the other statistical tests previously mentioned. No significant differences were seen between the Egyptian and other Caucasian groups, and, in general, populations within the major geographic groups were not found to be significantly different from each other. Notable exceptions are the Mapuche and Navajo populations compared to other Native Americans. The Mapuche and Navajo populations were heterogeneous with respect to other Native Americans

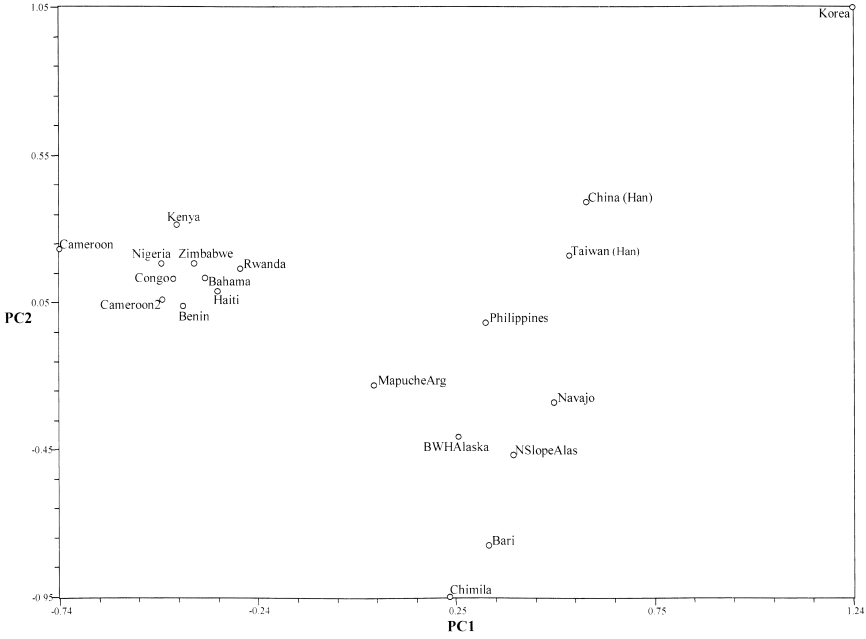

Figure 5. Principal-components plot from an analysis excluding all Caucasian groups (Andalucia, Andalucia2, Arab Moslems, Australia, Basque, Canary Islands, Denmark, Dubai Arabs, Egypt, Galicia, Greece, Saudi Arabia, and Turkey).

but not significantly different from the European groups. This might be attributed to admixture with groups of European descent.

Discussion

The observed heterozygosity for each of the five African populations in our study is greater than 70%, yet the sub-Saharan populations exhibit higher heterozygosity values from 81% to 93%. Of the five African populations examined in this study, the Kenya group exhibits the highest level of heterozygosity, 93.4%. These heterozygosity values are indicative of substantial genetic diversity in the D1S80 locus. A heterozygosity of 87.6% was previously reported in an African American population (Budowle et al. 1997). The power of discrimination value of the D1S80 locus is high, with values above 90% for all five populations. Previous reports have indicated that the discrimination power for this locus ranges between 94% and 98% for worldwide populations (Applied Biosystems 2001).

Of the five populations genotyped in our study, only the Benin population was found not to be in Hardy–Weinberg equilibrium ($p = 0.019$). Theta values ranged from 0.009 in the Egyptian population to 0.032 in the Benin samples. High theta values indicate greater genetic substructure within a population. The

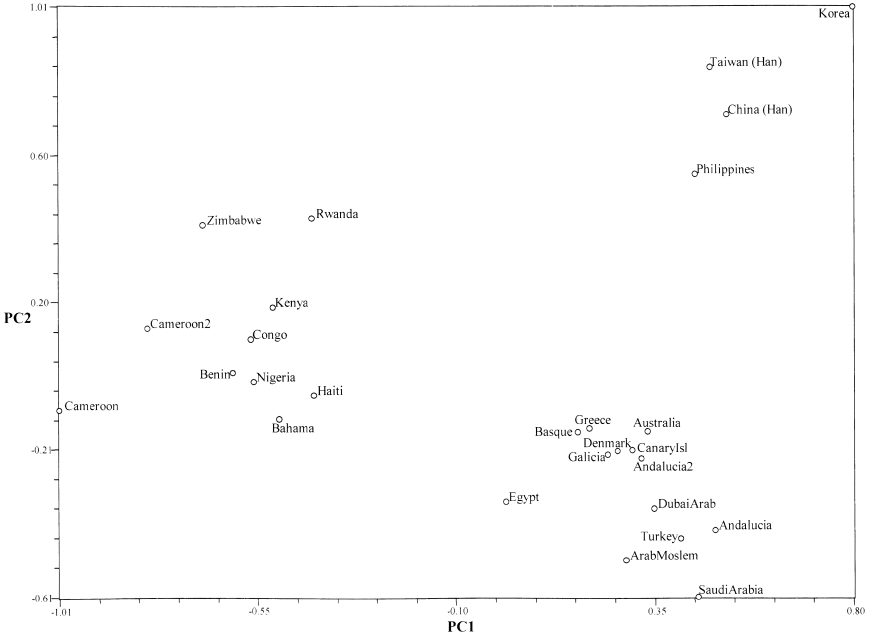

Figure 6. Principal-components plot from an analysis excluding all Native American groups.

elevated theta value in Benin could be related to the lack of genetic equilibrium, which in turn could be due to subpopulation structure. The overall theta *p* value, reflecting the allele variance in the D1S80 locus in the five African populations, was 0.023. This is comparable to the 0.019 value reported for other populations (Klitz et al. 2000). The National Research Council II on forensic DNA fingerprinting recommends the use of such a factor in probability calculations of all forensic cases. Recommendation 4.1 of this council's 1996 report suggests that a theta value of 0.01 should be used for United States populations, whereas a value of 0.03 would be appropriate for more isolated populations to correct for the effects of subpopulation structure (Committee on DNA Forensic Science 1996).

In the maximum-likelihood dendrogram presented in Figure 1, the Middle Eastern groups (i.e., Saudi Arabia, Dubai Arabs, and Arab Moslems) cluster by themselves into a subclade separate from the other Caucasian populations. The East Asians are also shown to be distinct and segregate into a subclade separate from the Native American populations, even though they group in the same cluster. Furthermore, the East Asians separate into an insular subclade (Taiwan and Philippine) and a continental (China and Korea) subclade. Similarly, the two South American native populations (the Bari and Chimila Amerindians) cluster together distinctly from the North American groups (the two Alaskan Eskimos populations and the Navajo). It is interesting that, within the African cluster,

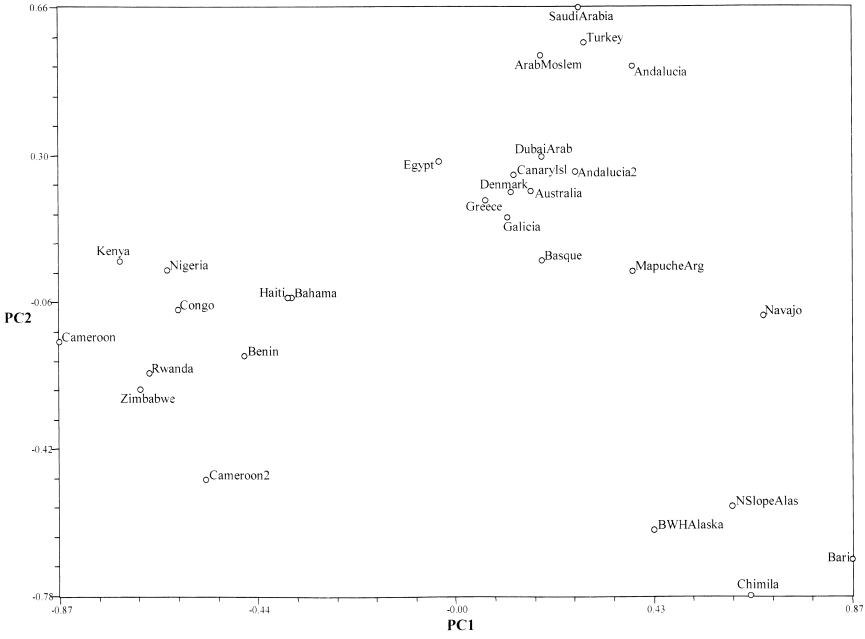

Figure 7. Principal-components plot from an analysis excluding all East Asian groups.

admixed populations of sub-Saharan African descent (Haiti and Bahamas) are located in the fringe of the clade closer to other geographic groups. Considering that the bootstrap values were derived from a single locus, it is remarkable that few nodes score below 90% probability. The overall topology of the maximum-likelihood tree based on D1S80 allele frequencies conforms closely to the accepted segregation of populations along geographic and ethnic groups based on classic genetic markers as well as protein and immunological polymorphisms (Cavalli-Sforza et al. 1994). D1S80 is not only capable of accurately segregating populations into geographic groups but is also able to discriminate between groups belonging to different biogeographic areas within the same ethnic group (e.g., Middle Easterners and Southwest Asians from Europeans, East Asians from

Table 4. Test for Hardy–Weinberg Equilibrium and Power of Discrimination

Population	*N*	*Fisher Exact (3200 Shufflings) (p)*	*Chi-Square (3200 Shufflings) (p)*	*Power of Discrimination (%)*
Benin	100	0.019	0.109	96.98
Kenya	106	0.704	0.212	97.03
Rwanda	100	0.247	0.036	97.06
Egypt	100	0.049	0.377	93.48
Cameroon	16	0.879	0.492	92.19

Table 5. *G* Scores and Probability Values for All Pairwise Combinations Involving the 33 Populations[a]

Population	*Anc*	*An2*	*Ara*	*Aus*	*Bah*	*Bar*	*Bas*	*Ben*	*BWH*	*Cam*	*Ca2*
Andalucia		39.607	37.366	38.200	113.841	110.568	44.118	143.180	152.635	130.007	100.841
Andalucia2	0.904		47.932	25.949	93.360	117.457	49.165	113.126	144.274	111.033	77.863
Arab Moslems	0.895	0.732		58.012	75.265	129.990	78.713	102.207	179.643	100.351	74.700
Australia	0.937	1.000	0.322		130.877	118.601	38.410	172.272	162.712	150.267	108.843
Bahama	0.000	0.000	0.008	0.000		151.487	140.302	17.390	162.303	31.186	43.138
Bari	0.000	0.000	0.000	0.000	0.000		103.104	167.631	65.014	141.848	106.582
Basque	0.803	0.744	0.012	0.971	0.000	0.001		178.190	141.540	152.011	110.707
Benin	0.000	0.000	0.000	0.000	1.000	0.000	0.000		176.910	31.453	37.415
Bethel-Wade	0.000	0.000	0.000	0.000	0.000	0.009	0.001	0.000		162.327	111.761
Cameroon	0.000	0.000	0.000	0.000	0.967	0.000	0.000	0.958	0.000		21.447
Cameroon2	0.000	0.000	0.000	0.000	0.677	0.000	0.000	0.812	0.000	1.000	
Canary Islands	0.996	0.998	0.963	0.995	0.000	0.000	0.888	0.000	0.000	0.000	0.000
Chimila	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
China	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
Congo	0.000	0.000	0.000	0.000	0.176	0.000	0.000	0.425	0.000	0.934	0.993
Denmark	0.994	0.995	0.294	1.000	0.000	0.000	0.993	0.000	0.000	0.000	0.000
Dubai Arabs	0.674	0.723	0.975	0.575	0.000	0.000	0.157	0.000	0.000	0.000	0.000
Galicia	0.902	0.990	0.146	0.976	0.000	0.000	1.000	0.000	0.000	0.000	0.000
Greece	0.814	0.629	0.830	0.112	0.000	0.000	0.424	0.000	0.000	0.000	0.000
Haiti	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.992	0.773
Kenya	0.000	0.000	0.000	0.000	0.995	0.000	0.000	0.957	0.000	0.975	0.645
Korea	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
Mapuche	0.714	0.773	0.661	0.827	0.000	0.000	0.689	0.000	0.001	0.000	0.000
Navajo	0.366	0.697	0.112	0.753	0.001	0.004	0.588	0.000	0.166	0.000	0.004
Nigeria	0.000	0.000	0.009	0.000	1.000	0.000	0.000	1.000	0.000	1.000	0.960
North Slope	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.000	0.319	0.000	0.000
Philippines	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rwanda	0.000	0.000	0.000	0.000	0.061	0.000	0.000	0.566	0.000	0.349	0.678
Saudi Arabia	0.054	0.001	0.329	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
Taiwan	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
Tanta (Egypt)	0.944	0.978	0.995	0.843	0.134	0.000	0.799	0.005	0.000	0.002	0.032
Turkey	0.999	0.694	0.999	0.751	0.000	0.000	0.247	0.000	0.000	0.000	0.000
Zimbabwe	0.000	0.000	0.000	0.000	0.227	0.000	0.000	0.546	0.000	0.922	0.891

Table 5. Continued.

Population	*Can*	*Chm*	*Chn*	*Con*	*Den*	*Dub*	*Gal*	*Gre*	*Hai*	*Ken*	*Kor*
Andalucia	26.710	242.916	123.671	146.468	30.276	45.073	40.870	40.924	125.557	153.487	126.254
Andalucia2	28.462	212.149	131.290	111.572	31.112	45.934	30.928	47.290	106.217	135.584	138.957
Arab Moslems	33.556	230.491	146.251	97.702	56.502	31.172	62.444	39.725	89.025	99.273	139.397
Australia	30.595	275.305	156.095	171.063	20.131	50.561	34.893	65.816	155.570	191.049	168.103
Bahama	98.566	231.274	163.705	56.588	121.455	97.947	115.624	91.023	9.918	26.481	130.172
Bari	113.677	65.327	117.950	142.303	123.785	112.306	109.924	120.636	170.148	181.735	120.155
Basque	42.610	274.587	152.081	179.871	32.600	62.875	25.339	53.222	167.673	209.024	158.468
Benin	121.300	251.940	218.647	46.469	148.110	128.501	140.188	110.725	18.432	32.126	183.707
Bethel-Wade	160.778	171.059	144.319	187.132	156.341	150.532	165.211	151.856	173.671	221.464	147.171
Cameroon	119.943	180.598	163.019	33.435	136.752	113.046	128.385	105.311	27.216	30.811	137.238
Cameroon2	94.847	121.522	120.043	29.011	101.393	81.840	84.684	93.787	37.972	42.955	100.573
Canary Islands		233.634	109.141	119.099	26.296	35.715	29.656	26.248	112.040	126.744	119.247
Chimila	0.000		279.437	179.004	280.336	228.222	256.023	232.589	249.148	276.121	254.587
China	0.000	0.000		186.581	152.396	131.487	148.493	117.489	180.447	221.648	32.800
Congo	0.000	0.000	0.000		161.252	117.265	130.518	108.194	55.783	52.436	152.371
Denmark	0.997	0.000	0.001	0.000		48.415	29.092	57.369	141.032	179.522	166.526
Dubai Arabs	0.935	0.000	0.000	0.000	0.604		40.832	45.101	110.938	126.084	130.129
Galicia	0.998	0.000	0.001	0.000	1.000	0.851		52.790	135.106	161.797	137.335
Greece	0.999	0.000	0.000	0.000	0.260	0.595	0.384		101.656	113.479	123.019
Haiti	0.000	0.000	0.000	0.127	0.000	0.000	0.000	0.000		21.703	142.693
Kenya	0.000	0.000	0.000	0.294	0.000	0.000	0.000	0.000	1.000		177.234
Korea	0.000	0.000	0.998	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Mapuche	0.565	0.000	0.034	0.000	0.468	0.611	0.121	0.365	0.000	0.000	0.000
Navajo	0.289	0.000	0.079	0.000	0.547	0.132	0.544	0.020	0.000	0.000	0.040
Nigeria	0.000	0.000	0.000	0.926	0.000	0.000	0.000	0.000	1.000	1.000	0.000
North Slope	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Philippines	0.002	0.000	0.243	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rwanda	0.000	0.000	0.000	0.850	0.000	0.000	0.000	0.000	0.047	0.425	0.000
Saudi Arabia	0.005	0.000	0.001	0.000	0.001	0.361	0.001	0.027	0.000	0.000	0.000
Taiwan	0.002	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.693
Tanta (Egypt)	0.966	0.000	0.000	0.000	0.971	0.894	0.918	0.989	0.010	0.001	0.000
Turkey	0.983	0.000	0.000	0.000	0.687	0.766	0.433	0.973	0.000	0.000	0.000
Zimbabwe	0.000	0.000	0.000	0.969	0.000	0.000	0.000	0.000	0.415	0.646	0.000

Table 5. Continued.

Population	*Map*	*Nav*	*Nig*	*NSA*	*Phi*	*Rwa*	*Sau*	*Tai*	*Tan*	*Tur*	*Zim*
Andalucia	41.104	49.716	131.656	169.899	108.519	168.469	65.720	103.030	36.641	26.075	222.281
Andalucia2	43.099	43.987	109.624	157.981	120.831	142.548	104.940	103.976	33.689	48.917	193.150
Arab Moslems	41.504	56.448	78.158	192.991	120.772	117.153	51.768	107.591	29.973	27.009	159.421
Australia	41.717	42.383	156.745	172.030	147.608	223.569	128.394	120.909	42.397	46.275	285.404
Bahama	84.808	80.533	20.551	227.178	159.925	65.632	175.476	121.110	59.671	147.989	54.610
Bari	76.470	61.784	175.901	59.775	115.429	166.623	176.081	122.218	133.725	119.068	191.230
Basque	45.585	46.337	168.817	159.639	128.989	223.925	137.352	114.901	44.065	59.029	298.696
Benin	107.236	92.984	18.197	229.203	174.025	46.187	218.823	145.400	75.585	152.219	44.397
Bethel-Wade	75.158	51.168	193.508	46.936	160.884	207.418	296.155	126.507	148.347	117.485	245.616
Cameroon	90.682	112.572	22.213	206.225	157.942	50.547	159.837	133.847	80.703	139.143	35.580
Cameroon2	77.951	71.980	32.567	118.487	106.827	41.962	110.536	97.612	65.794	106.624	36.584
Canary Islands	44.023	51.156	107.602	160.159	81.753	132.269	76.757	82.041	33.659	32.125	189.462
Chimila	147.515	93.234	239.257	186.523	223.738	264.791	344.003	230.828	232.619	241.360	273.669
China	72.374	63.737	194.085	159.601	60.732	230.538	308.712	23.733	129.085	137.124	272.791
Congo	125.675	102.051	35.139	191.858	142.699	38.328	173.634	136.409	98.742	137.854	31.104
Denmark	50.298	47.231	140.014	157.979	130.925	196.173	99.333	119.295	33.970	46.820	265.064
Dubai Arabs	45.437	55.877	102.245	145.521	128.256	122.977	47.847	120.199	37.462	45.368	207.168
Galicia	62.296	47.986	134.122	156.447	121.664	157.906	97.748	117.993	38.109	54.235	233.191
Greece	48.039	64.533	95.708	177.959	96.639	107.389	67.310	86.571	27.537	32.431	175.021
Haiti	95.959	93.759	20.347	238.417	171.147	67.002	183.290	133.163	71.018	144.069	48.656
Kenya	119.720	113.156	13.760	282.176	202.894	49.228	228.369	166.715	81.392	156.318	43.631
Korea	96.236	71.044	162.908	171.779	98.165	185.059	277.317	48.672	133.840	149.635	213.142
Mapuche		33.734	109.913	103.612	81.819	136.517	81.088	61.267	45.510	39.556	162.977
Navajo	0.770		101.281	53.518	56.840	105.015	74.869	50.194	56.984	56.483	119.820
Nigeria	0.000	0.000		234.626	166.791	30.837	164.973	143.123	70.595	131.079	34.258
North Slope	0.000	0.135	0.000		138.991	228.179	303.174	140.537	179.921	186.407	296.454
Philippines	0.000	0.103	0.000	0.000		173.564	209.880	27.685	111.666	115.354	208.422
Rwanda	0.000	0.000	0.983	0.000	0.000		236.932	159.017	100.531	163.188	55.842
Saudi Arabia	0.001	0.004	0.000	0.000	0.000	0.000		213.788	46.167	49.613	342.655
Taiwan	0.072	0.304	0.000	0.000	0.995	0.000	0.000		96.118	107.290	179.009
Tanta (Egypt)	0.617	0.211	0.017	0.000	0.000	0.000	0.568	0.000		36.962	147.280
Turkey	0.788	0.145	0.000	0.000	0.000	0.000	0.497	0.000	0.937		225.073
Zimbabwe	0.000	0.000	0.942	0.000	0.000	0.183	0.000	0.000	0.000	0.000	

a. *G* scores are given in the upper triangle of values; probability values are given in the lower triangle of values.

Native Americans, native North Americans from South American Amerindians, and sub-Saharan African populations from admixed groups of sub-Saharan African descent).

The principal-components plot based on all worldwide populations (Figure 2) clearly delineates the sub-Saharan African populations from the other populations along the *x*-axis. Within the sub-Saharan African cluster there is an evident separation of East African groups (Kenya, Rwanda, and Zimbabwe) from West African populations (Benin, Cameroon, Cameroon2, Congo, and Nigeria) and from populations of West African descent (Haiti and the Bahamas) along PC2. The admixed Haitian and Bahamian populations of sub-Saharan West African descent segregate at the fringes of the African cluster in the direction of the Caucasian groups. The North African population from the Nile delta falls between these admixed groups and the Caucasians but nearer to the Caucasians. The sub-Saharan groups had the highest range of variability along the *x*-axis, spanning from the Cameroon population to the Haitians.

Orientals (i.e., East Asians and Native American populations) are the second most variable group along the *x*-axis, yet they are better resolved along the *y*-axis. This geographic group is the most dispersed with respect to the first two principal components. This diffused segregation of East Asians and populations of East Asian descent (Native Americans) may be the result of many factors, such as genetic drift, founder effects, inbreeding, and admixture experienced by Native American groups (Salzano 2002). The East Asian populations group together some distance from Native American populations along PC 2. The principal-components plot also provides the resolution to segregate the North American populations (the two Alaskan Eskimo groups and the Navajo) from the South American groups (Bari and Chimila). The Navajo's geographic location in between the Alaskan Eskimo groups and the South American populations is reflected in their intermediate plot position between the two. The most likely explanation for the Mapuche's clustering with Caucasians is their high level of admixture with Europeans. Also, the close proximity of the Chimila, an Amerindian group, to the Caucasian cluster might represent some degree of gene flow from Europeans. The European populations are clustered together closely, implying more similarity within the group. Limited genetic diversity among European populations has been previously reported using various types of genetic markers (Cavalli-Sforza et al. 1994). Segregation, however, is evident between the Middle East–Southwest Asian groups (Caucasian populations) and European populations or populations of European descent. Although the principal-components plot was able to better resolve the sub-Saharan African populations along the east–west geographic axis of the continent and the maximum-likelihood analysis better delineated the genetic affinities between populations from Greece, Andalucia, and the Canary Islands with African groups, overall there exists a striking parallelism between the maximum-likelihood and principal-components results.

To provide greater detail on the phylogenetic relationships of sub-Saharan African populations, we performed a series of principal-components analyses in

which geographic groups of populations were sequentially subtracted. As expected, the deletion of specific geographic groups outside sub-Saharan Africa generated plots representing a greater percentage of the total diversity. This may indicate a reduction in the distortion introduced by including the rest of the world. The subtraction of the East Asian and Native American groups increased the detected combined PC 1 and PC 2 diversity to 42% of the total (Figure 3). Not surprisingly, this principal-components analysis allowed for a greater separation between populations, including the sub-Saharan Africans, but it is not clear why the segregation of East and West African groups was less well delineated (i.e., West African populations from Congo and Nigeria grouped within the East African cluster) (Figure 3).

A greater separation of the Middle East–Southwest Asian populations from European groups in the absence of the Oriental populations also provided greater resolution, as represented by the analysis in Figure 3. As expected, maximum detected diversity (51%) and separation among sub-Saharan African groups were observed when they were examined by themselves (Figure 4). A lack of segregation of East from West African populations was also observed in the principal-components analyses when sub-Saharan groups were examined alone or when Caucasian, Native American, and East Asian population groups were individually subtracted (Figures 4–7). Geographic delineation between sub-Saharan populations was lowest when the East Asian groups were deleted (Figure 7). Similarly, the previous observed segregation of Middle East–Southwest Asian groups from Europeans deteriorates when the East Asians are removed. It seems that subtracting geographic groups of populations can have multidimensional effects on the outcome of principal-components analyses and that a simple reduction in groups does not necessarily augment discrimination.

The *G* test indicated some genetic homogeneity within the major geographic groups. In general, the results of the *G* test corroborate the phylogenetic data generated from the maximum-likelihood and principal-components analyses. The *G* test allows for pairwise comparison of all possible combinations of populations. This provides the means to focus only on two populations at a time and to ascertain whether their genetic differences, in this case at the D1S80 locus, are statistically significant. The Mapuche group is significantly different from the Philippine, Chinese, and the two Alaskan Eskimo populations but not from the Greeks, whereas the Navajos are not different from the same four Oriental groups but are significantly different from the Greeks. Both the Mapuche and the Navajo are not significantly different from several European groups. These *G* test results might indicate gene flow from European gene pools into these two Native American populations.

As expected, the Mapuche and the Navajo are significantly different from all sub-Saharan African populations. The group from Egypt represents another example of admixture. This population shows no significant difference with the rest of the Caucasian groups. Lack of a significant difference with the Bahamians, a group primarily of sub-Saharan West African descent, and borderline significant distinctness ($p = 0.032$) from one of the Cameroon populations might indicate gene flow from sub-Saharan Africa.

The Bari and Chimila groups are significantly different from each other and from all the Native American and East Asian populations studied. The Bari and Chimila are noninterbreeding groups and are linguistically distinct. They are geographically proximal in northern Colombia, South America. The fact that the Bari and Chimila are significantly different from other Native Americans and East Asians may result from genetic drift resulting from a founder effect, bottleneck, and/or reproductive isolation. Because our data are based on a single locus, the effect of genetic drift may be particularly apparent. Similarly, Native North American populations (Alaskan Eskimos and Navajo) were significantly different from South American groups but not from each other. Although Eskimos and Navajos belong to different major linguistic families, they are geographically closer to each other than to the South American Amerindian groups.

In summary, in this study we report for the first time on D1S80 databases from five African populations. The D1S80 marker system provides the basis for an expeditious and simple procedure to identify individuals and to ascertain phylogenetic relationships. These five African databases will contribute to the field of DNA fingerprinting for forensic, maternity, and paternity determinations. In addition, this locus allowed us to discriminate among 33 worldwide populations at the subgeographic level.

Received 11 March 2003; revision received 10 July 2003.

Acknowledgments This research was partially supported by the National Institutes of Health through grants GM 08205 to R.J. Herrera and GM 61347 to L.M. Ruiz. We would like to thank Barbara Sewerin and Rachel Chow for their contributions in the preparation of this manuscript.

Literature Cited

Alkhayat, A., F. Alshamali, and B. Budowle. 1996. Population data on the PCR-based loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80 from Arabs from Dubai. *Forensic Sci. Int.* 81:29–34.

Allen, R.C., B. Budowle, and G. Graves. 1989. Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver. *Biotechniques* 7:736–744.

Antunez-de-Mayolo, G., A. Antunez-de-Mayolo, P. Antunez-de-Mayolo et al. 2002. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* 23:3346–3356.

Applied Biosystems. 2001. *AmpliFLP D1S80 PCR Amplification Kit Product Detail.* Foster City, CA: Applied Biosystems.

Araújo Da Silva, W., Jr., M.C. Bortolini, D. Meyer et al. 1999. Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. *Am. J. Phys. Anthropol.* 109:425–437.

Arnaiz-Villena, A., E. Gomez-Casado, and J. Martinez-Laso. 2002. Population genetic relationships

between Mediterranean populations determined by HLA allele distribution and a historic perspective. *Tissue Antigens* 60(2):111–121.

Buchanan, A.V., G.M. Risch, M. Robichaux et al. 2000. Long DOP-PCR of rare archival anthropological samples. *Hum. Biol.* 72:911–925.

Budowle, B., B.W. Koons, and J.D. Errera. 1996. Multiplex amplification and typing procedure for the loci D1S80 and amelogenin. *J. Forensic Sci.* 41:660–663.

Budowle, B., J.B. Smerick, K.M. Keys et al. 1997. United States population data on the multiplex short tandem repeat loci—HUMTH01, TPOX, and CSF1PO—and the variable number tandem repeat locus D1S80. *J. Forensic Sci.* 42:846–849.

Cakir, A.H., L. Acik, and T. Kesici. 2001. Distribution of amplified fragment length polymorphism D1S80 alleles in Turkish population. *J. Forensic Sci.* 46:1002.

Cariolou, M.A., P. Maroli, M. Christophorou et al. 1998. Greek Cypriot allele and genotype frequencies for amplitype PM-DQA1 and D1S80 loci. *J. Forensic Sci.* 43:661–664.

Carmody, G. 1990. *G-Test.* Ottawa: Department of Biology, Carleton University (unpublished program distributed by the author).

Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes.* Princeton, NJ: Princeton University Press, 73–83, 162–166.

Clancy-Smith, J. 1997. Berber history. *J. Afr. Hist.* 38(3):497–534.

Committee on DNA Forensic Science. 1996. *The Evaluation of Forensic DNA Evidence.* Washington, DC: National Academy Press, 5.

Duncan, G. 1996. Microvariation at the human D1S80 locus. Miami: Ph.D. dissertation, Florida International University.

Duncan, G., E. Thomas, J.C. Gallo et al. 1996. Human phylogenetic relationships according to the D1S80 locus. *Genetica* 98:277–287.

Duran, R., and M. Ruiz-García. 2001. Genetic population history relationships of the population of Bogota, Columbia, by using the D1S80, VWA, and TH01 molecular markers. *Am. J. Hum. Biol.* 13:374–383.

Felsenstein, J. 1993. *PHYLIP (Phylogeny Inference Package), Version 3.5c.* Seattle: Department of Genetics, University of Washington.

Flores, I., I. Frias, V. Prieto et al. 2001. Population data for southern Spain and Canary Islands of HLA-DQA1, PM, and D1S80 loci. *Forensic Sci. Int.* 119:116–118.

Guthrie, M. 1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages,* v. 2, *Bantu Prehistory, Inventory, and Indexes.* Farnborough, UK: Gregg International.

Gutowski, S., B. Budowle, J. Auer et al. 1995. Statistical analysis of an Australian population for the loci Gc, HLA-DQA1, D1S80, and HUMTH01. *Forensic Sci. Int.* 76:1–6.

Halos, S.C., J.Y. Chu, A.C.M. Ferreon et al. 1999. Philippine population database at nine microsatellite loci for forensic and paternity applications. *Forensic Sci. Int.* 101:27–32.

Harich, N., E. Esteban, A. Chafik et al. 2002. Classical polymorphisms in Berbers from Moyen Atlas (Morocco): Genetics, geography, and historical evidence in the Mediterranean peoples. *Ann. Hum. Biol.* 29(5):473–487.

Holden, C.J. 2002. Bantu language trees reflect the spread of farming across sub-Saharan: A maximum-parsimony analysis. *Proc. R. Soc. Lond.* 269:793–799.

Huffman, T.N. 1982. Archeology and ethnohistory of the African Iron Age. *Ann. Rev. Anthropol.* 11:133–150.

Hutz, M.H., V.S. Mattevi, S.M. Callegari-Jaques et al. 1997. D1S80 locus variability in South American Indians. *Ann. Hum. Biol.* 24:249–255.

Kasai, K., Y. Nakamura, and R. White. 1990. Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. *J. Forensic Sci.* 35:1196–2000.

Klitz, W., R. Reynolds, J. Chen et al. 2000. Analysis of genotype frequencies and interlocus association for the PM, DQA1, and D1S80 loci in four populations. *J. Forensic Sci.* 45:1009–1015.

Lee, J.W., H.-S. Lee, M. Park et al. 2001. Evaluation of DNA match probability in criminal case. *Forensic Sci. Int.* 116:139–148.

Levinson, D. 1998. *Ethnic Groups Worldwide: A Ready Reference Handbook.* Phoenix: Oryx Press, 111, 115, 140, 152, 159–160.

Lewis, P.O., and D. Zaykin. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data, Version 1.0 (d16c). Free program distributed by the authors; available at http://lewis.eeb.uconn.edu/lewishome/software.html.

Li, C.C. 1976. *First Course in Population Genetics.* Pacific Grove, CA: Boxwood Press.

Lorente, M., J.A. Lorente, M.R. Wilson et al. 1997. Spanish population data on seven loci: D1S80, D17S5, HUMTH01, HUMVWA, ACTBP2, D21S11, and HLA-DQA1. *Forensic Sci. Int.* 86:163–171.

Maxon, R.M. 2002. *Kenya*. Microsoft Encarta Online. Available at http://encarta.msn.com. Redmon, WA: Microsoft Corporation.

Monson, K., and B. Budowle. 1998. Effect of reference database on frequency estimates of polymerase chain reaction (PCR)-based DNA profiles. *J. Forensic Sci.* 43:483–488.

Murdock, G.P. 1959. *Africa: Its Peoples and Their Culture History*. New York: McGraw–Hill, 111, 127, 392.

Nakamura, Y., M. Carlson, K. Krapcho et al. 1988. Isolation and mapping of a polymorphic DNA sequence (pMCT118) on chromosome 1p(D1S80). *Nucleic Acids Res.* 16:9364.

Novick, G.E., C.C. Novick, J. Yunis et al. 1995. Polymorphic human specific *Alu* insertions as markers for human identification. *Electrophoresis* 16:1596–1601.

Pereira, L., L. Gusmao, C. Alves et al. 2002. Bantu and European Y-lineages in sub-Saharan Africa. *Ann. Hum. Genet.* 66(6):369–378.

Peterson, B., B. Su, R. Chakraborty et al. 2000. World population data for the HLA-DQA1, PM, and D1S80 loci with least and most common profile frequencies for combinations of loci estimated following NRC II guidelines. *J. Forensic Sci.* 45:118–146.

Rangel-Villalobos, H., F. Rivas, M. Torres-Rodríguez et al. 1999. Allele frequency distributions of six Amp-FLPs (D1S80, APO-B, VWA, TH01, CSF1PO, and HPRTB) in a Mexican population. *Forensic Sci. Int.* 105:125–129.

Rohlf, F.J. 2002. *Numerical Taxonomy System (NTSys pc), Version 2.1.* Setauket, NY: Exeter Publishing.

Saferstein, R. 1982. *Forensic Science Handbook,* v. 1. Englewood Cliffs, NJ: Prentice Hall.

Salzano, F.M. 2002. Molecular variability in Amerindians: Widespread but uneven information. *Ann. Braz. Acad. Sci.* 74(2): 223–263.

Schmidt, P.R. 1975. A new look at interpretations of the Early Iron Age in East Africa. *Hist. Afr.* 2:127–136.

Tahir, M.A., S.K. Sinha, N. A-Al-Kubaidan et al. 2000. Distribution of amplified fragment length polymorphism D1S80 alleles in a Saudi Arabian population. *J. Forensic Sci.* 45:1159.

Vansina, J. 1990. *Paths in the Rainforests: Toward a History of Political Tradition in Equatorial Africa.* London: Elsevier.

Wake Forest University. 2002. Population genetics of gene flow. Available from www.wfu.edu/users/sipet00g/populationgenetics.html.

Walkinshaw, M., L. Strickland, H. Hamilton et al. 1996. DNA profiling in two Alaskan Native populations using HLA-DQA1, PM, and D1S80 loci. *J. Forensic Sci.* 41:478–484.

Watkins, W.S., C.E. Ricker, M.J. Bamshad et al. 2001. Patterns of ancestral human diversity: An analysis of *Alu*-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* 68:738–752.

Williamson, K., and R. Blench. 2000. Niger-Congo. In *African Languages: An Introduction,* B. Heine and D. Nurse, eds. Cambridge: Cambridge University Press, 11–42.