Terry Melton,[1] *Ph.D.; Charles Ginther,*[2] *Ph.D.; George Sensabaugh,*[2] *Ph.D.; Himla Soodyall,*[1,3] *Ph.D.; and Mark Stoneking,*[1] *Ph.D.*

Extent of Heterogeneity in Mitochondrial DNA of Sub-Saharan African Populations*

REFERENCE: Melton T, Ginther C, Sensabaugh G, Soodyall H, Stoneking M. Extent of heterogeneity in mitochondrial DNA of sub-Saharan African populations. J Forensic Sci 1997;42(4):582–592.

ABSTRACT: Variation in the mitochondrial DNA (mtDNA) control region as detected by sequence-specific oligonucleotide (SSO) probes is described for 381 individuals from nine sub-Saharan African populations. Population diversity estimates for SSO types ranged from 0.23 to 0.97, while 102 SSO types were detected, none of these types was shared by more than four populations. Eighteen types occurred in ≥ 10% of individuals in some populations; of these, 11 were population-specific. One type occurred in 15% of the total sample, but was shared among only three populations. African SSO types were characterized by high frequencies of blank variants, indicating that there was additional variation present at the nucleotide sequence level in regions where SSO probes hybridize. Analyses of molecular variance (AMOVA) incorporating genetic distances between SSO types showed that 30% of the total variation was due to differences among populations, indicating that there is statistically significant heterogeneity ($p < 0.001$). An AMOVA on mtDNA control region nucleotide sequence data from 12 populations showed that including all additional variation present at the sequence level increased the variance due to population subdivision to 34% ($p < 0.001$). Overall, when considering both the low diversity within some populations and high heterogeneity among populations, SSO typing of mtDNA may not be a desirable forensic DNA typing method for continental African populations. Further mtDNA sampling of African-derived populations of North America should be carried out to determine how much of the continental African mtDNA variation is of forensic significance. However, the existence of extensive mtDNA control region nucleotide sequence variation in African populations means that control region sequencing is still appropriate in forensic cases requiring mtDNA analysis.

KEYWORDS: forensic science, DNA typing, genetic markers, mitochondrial DNA, population genetics, African populations, sequence-specific oligonucleotide typing, analysis of molecular variance.

The population genetics of mitochondrial DNA (mtDNA), as they apply to its use as a forensic DNA typing locus, have been previously described for Asian, European, and European-derived North American populations (1,2). In these earlier reports, we demonstrated that there was high diversity for mtDNA SSO types in all populations of both continental groups, modest statistically significant heterogeneity among Asian populations, and no significant heterogeneity among European populations. These conclusions supported the use of mtDNA as a forensic typing locus in these populations, including the use of sequence-specific oligonucleotide (SSO) mtDNA typing as an inexpensive exclusionary method which would be followed by confirmatory sequencing of the control region hypervariable segments in the event of a match. MtDNA is readily recoverable from biological evidentiary material such as bone, hair, or blood, where the DNA may be minimal and/or degraded. These qualities, plus its high discriminatory potential and relative lack of subpopulation heterogeneity indicate that mtDNA is a potentially valuable locus for forensic DNA typing. Laboratory guidelines for both SSO typing and control region sequencing are available, and mtDNA is currently being used in forensic investigations (3–9).

The purpose of this report is to conclude our continental survey of mtDNA populations genetics with an analysis of mtDNA types in African populations. We have maintained consistency among the three investigations by using identical methods of SSO typing and data analysis for each; thus far, including this study, a total of 39 populations and more than 2600 individuals have been described with respect to their mtDNA SSO type variation. In this investigation of the variation in the SSO types of 13 sub-Saharan African populations, mtDNA control region sequences were also available for almost every individual in the sample. SSO typing captures only a portion of the total sequence variation and hence may underestimate heterogeneity, as individuals with identical SSO types might nonetheless have different sequences. We therefore examined how the analysis based on SSO typing compares with the analysis based on the complete nucleotide sequence variation present in the control region in African populations.

Methods

For this study, complete sequence-specific oligonucleotide (SSO) types were inferred from the mtDNA control region nucleotide sequence data of 353 individuals from eight sub-Saharan African populations. Complete types, defined as those which include all eight possible sequence variant regions, came from the following ethnic groups: Herero (Botswana, N = 28) (10), Biaka Pygmies (Central African Republic, N = 17) (10), !Kung (Botswana, N = 25) (10), Yorubans (Nigeria, N = 13) (10), Mandenka (Senegal, N = 119) (11), Sierra Leoneans

[1]Department of Anthropology, The Pennsylvania State University, University Park, PA.
[2]Forensic Sciences Department, University of California Berkeley, CA.
[3]Department of Human Genetics, South African Institute for Medical Research and the University of Witwatersrand, Johannesburg, South Africa.

*This project was supported by NIJ grant 92-IJ-CX-K040 to M.S.; Portions of this work were presented at the 48th Annual Meeting, American Academy of Forensic Sciences, February 1996, Nashville TN.
Received 21 June 1996; and in revised form 25 Nov. 1996; accepted 26 Nov. 1996.

Copyright © 1997 by ASTM International

(Bonkolenken/Bagruwa chiefdoms, N = 114) (C. Ginther and G. Sensabaugh, personal communication), Hadza (Tanzania, N = 17) (10), and Mbuti Pygmies (Zaire, N = 20) (10). Individuals in these populations were chosen at random and were maternally unrelated. An additional 28 DNA samples were obtained from the Mukogodo, a pastoralist tribe of Kenya (12), and were typed according to methods reported in a previous paper (1). Some of the Mukogodo samples came from within known matrilines, and therefore the samples were not all from maternally unrelated individuals (13).

An additional 96 individuals for whom partial SSO type data were inferred from unpublished sequences (H. Soodyall, personal communication) were included for the analysis of molecular variance (AMOVA). This smaller data set included three Negroid populations, consisting of southeastern Bantu speakers from South Africa (N = 35), Dama (Namibia, N = 14), and Herero (Botswana, N = 8), and two Khoisan populations, consisting of Nama (Namibia, N = 9), and Sekele (Namibia, N = 30) (14). Partial data were available for all 96 individuals for six of the eight variant regions (excluding ID and IIA) and for 72 individuals for seven of the eight variant regions (excluding ID). Because of the sampling strategy used to collect and select these mtDNAs for sequencing, their SSO types were used only in the AMOVA analyses, and not to describe SSO type frequency, diversity, or distribution. Where possible, the additional Herero samples were combined for AMOVA analyses with those Herero samples from Botswana mentioned previously. The locations of populations included in this study are shown in Fig. 1.

The sequence-specific oligonucleotide SSO type variants and their arrangement into individual SSO types are described in detail elsewhere (1). The data set consisting of complete control region SSO types from African individuals (N = 381) was examined with respect to type frequency distribution both overall and within the nine populations. An unbiased estimate of diversity (h) and its variance were calculated to quantify the amount of both mtDNA SSO type variation and complete control region sequence variation in each population (1).

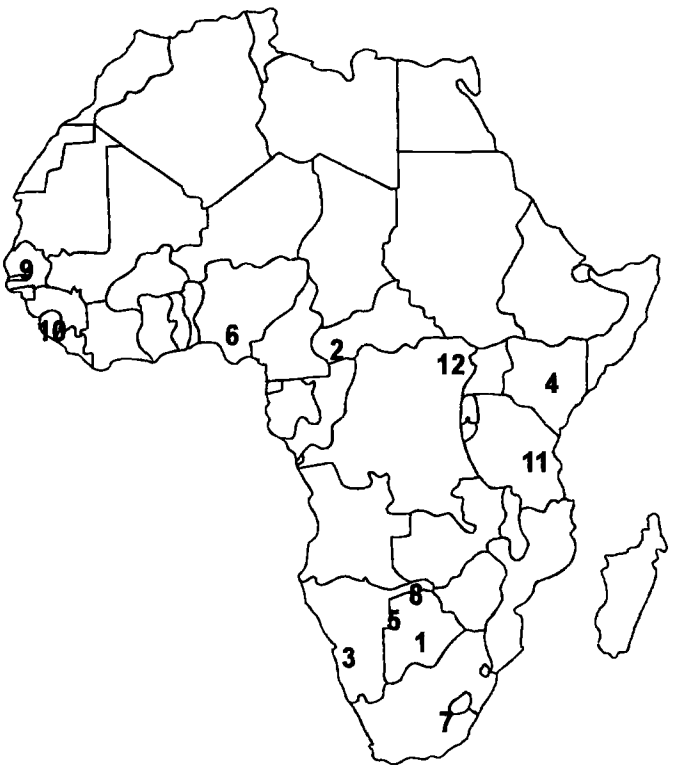

FIG. 1—*Locations of populations in this study: 1-Herero, 2-Mbuti Pygmies, 3-Dama/Nama, 4-Mukogodo, 5-!Kung, 6-Yorubans, 7-Southeastern Bantu, 8-Sekele, 9-Mandenka, 10-Sierra Leone, 11-Hadza, 12-Biaka Pygmies.*

AMOVA analyses were applied to complete and partial data sets of SSO types. AMOVA as used for these analyses is described in detail elsewhere (1,15). In brief, while incorporating additional data about the genetic distances between pairs of SSO types, AMOVA partitions the total variance in the data set into variance among populations (σ^2_a) and variance within populations (σ^2_b), as in a traditional analysis of variance, to reveal the extent of statistically significant population substructure. A correlation of random SSO types within populations, Φ_{ST}, is computed, and permutational procedures are used to test the significance of Φ_{ST} and the variance components. Test statistics, equal to the probability of observing random variance components and statistics greater than those generated in the analysis are reported.

Because SSO typing revealed only a portion of the total variation present in the mtDNA control region in African populations, AMOVA was also applied to available control region nucleotide sequence data to determine the extent of heterogeneity which might be attributable to additional control region variation. The current limit on AMOVA parameters is 256 haplotypes, or, in this case, sequence types (L. Excoffier, personal communication), therefore 40 of the 90 unique Sierra Leone sequences were chosen at random for the analysis, and combined with 271 sequences from the 11 other populations (not including the Mukogodo sample). After excluding missing nucleotide data, insertions, and deletions, and removing any duplicate sequences from within populations, 237 unique sequences remained (up to 753 sites were available for pairwise comparisons). Both the absolute number of nucleotide differences and Tamura-Nei distances (which accounts for different transition and transversion substitution rates in mitochondrial DNA) (16) were computed for each pair of sequences using the software program MEGA (17). The pairwise genetic distance matrices were used as AMOVA input along with population files which contained a roster of the sequence types in each population. Variance components and significance statistics were reported as for the SSO type AMOVA analyses. All AMOVA procedures were carried out with software provided by L. Excoffier.

AMOVA was also used to measure heterogeneity at one of the SSO variant regions for which there were large numbers of blanks (frequency of 0–100%, depending on the population). The total number of different sequence variants at region IIC that were distributed among 449 individuals in 12 populations (excluding Kenya) were tallied and designated as "types" in a standard AMOVA. This analysis allowed testing of three alternative hypotheses—that the nucleotide variation in this region is the result of: 1) high frequency population-specific substitutions which create significant heterogeneity, 2) high frequency substitutions shared by most populations which are not detectable by these SSO probes, or 3) low frequency random substitutions which prevent the detection of overall homogeneity among populations at probe specific sites.

Results

Within the data set of 381 complete SSO types from nine populations, there were 102 different SSO types (Appendix A). Figure 2

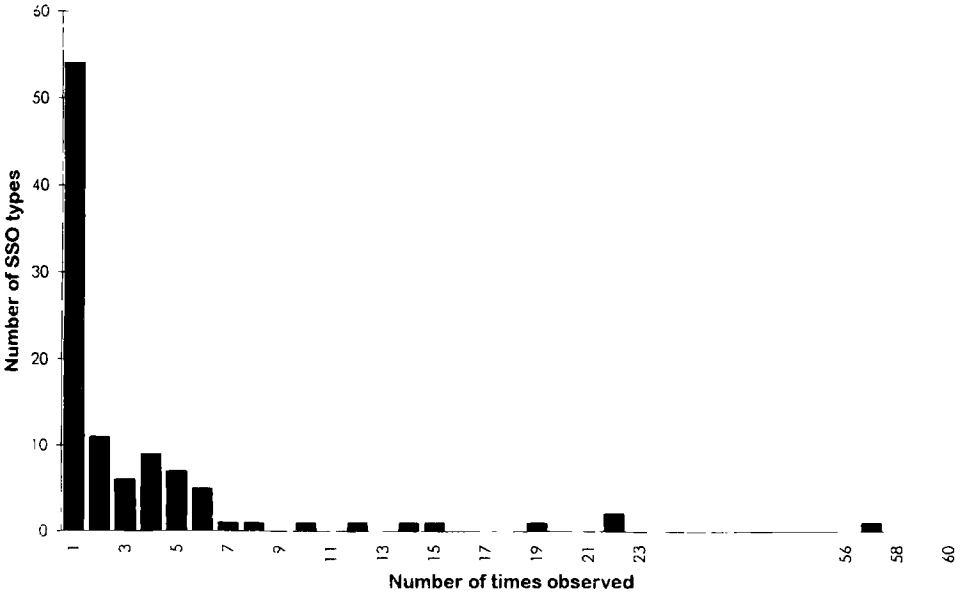

FIG. 2—*Distribution of mtDNA SSO types, showing the count of types occurring one through 57 times.*

shows the distribution of SSO types within the total sample. Sixty-four percent of all types were rare, occurring once (54 types) or twice (11 types). A single SSO type (type 38) occurred 57 times, or in 15% of the overall sample. Two other types (4 and 71) occurred in 22 individuals apiece, or approximately 6% of the total sample. All other types occurred in less than 4% of the total sample.

Fifteen types were shared by two populations, five types were shared by three populations, and two types were shared by four populations. No SSO type was shared by more than four populations. Of the 22 shared, or "public" types, seven occurred in $\geq$ 10% of the samples in at least one population. Of the 80 types which occurred in only one population, 54 were singletons, or "unique" types, occurring only one time. There were 26 types which occurred more than once in a single population, or "population specific" types, and of these, 11 were observed in $\geq$ 10% of the samples in that population. Figure 3 shows the frequency of unique types, population specific types, and public types within each of the nine populations (calculated as frequency of kind of type relative to the total number of types). There was no consistent distribution of kind of type among sub-Saharan African populations; for example, the !Kung and Mbuti Pygmies had mostly unique types, whereas the Biaka Pygmies and Mukogodo had mostly population specific types, and the Herero, Yorubans, and Mandenka had mostly public types.

Table 1 shows the number of individuals and numbers of SSO types and unique sequences observed in each population, along with unbiased estimates of population diversity (h) and their standard errors. Diversity estimates were extremely variable; for example, Yorubans displayed 11 SSO types among 13 individuals (h = 0.974 $\pm$ 0.014), whereas the Hadza had only 3 SSO types among 17 individuals (h = 0.228 $\pm$ 0.091). The mean SSO type diversity for the total sample was 0.961 $\pm$ 0.004. Diversity estimates at the sequence level exceeded diversity for SSO types for all populations, as expected, but were similarly limited in the Herero and Hadza populations. The largest number of identical sequences (N = 20) seen in any population was observed in the Herero population. The overall diversity for sequences was 0.992 $\pm$ 0.001.

The distributions of the most common public SSO types and the two SSO types shared by four populations are illustrated in Fig. 4. Types 33 and 91, shared by four of the nine populations, were present in low frequency where they were observed. The seven public types which occurred in $\geq$ 10% of any population were shared, at most, with two other populations, which displayed them at low frequency. Type 38, for example, occurred in 33% of the Mandenka, or 39 individuals, but the only other population with this type was Sierra Leone.

This data set displayed high frequencies of blank variants across most populations for several of the variant regions and in several populations for particular variant regions. Blank variants occur when none of the SSO probes hybridize, indicating the presence of unknown sequence variants. Table 2 shows these frequencies across all populations and SSO regions. In general, there were high frequencies of blanks for most populations for regions IIB and IIC, and for specific populations, such as the Hadza, at specific regions (e.g., IID). Because sequence data were available for most populations, it was possible to determine the true extent of nucleotide variation at these regions. Region IIC, for example, a stretch of nucleotides 18 bases long, had 33 different sequence variants among 449 individuals in 12 populations, demonstrating the high diversity which is present at the sequence level (data not shown). This pattern of variation was observed for all regions for which the frequency of blanks was high across populations. An AMOVA based on this region is described below.

AMOVA was used initially to generate interpopulation Φ_{ST} genetic distances (analogous to coancestry coefficients) and the significance values for all comparisons. Several SSO type AMOVA population pairwise distance matrices were generated to test whether both the complete and partial data sets, which contained different numbers of individuals and different populations, would yield similar results. The distance matrix for the complete data set, which contained nine populations and 381 individuals with 102 SSO types, is shown in Table 3 (data in Appendix A). Table 4 shows the distance matrix for one of the two partial data sets, which contained 453 individuals from 12 populations with 114 different SSO types (seven SSO variant regions; excludes ID; data

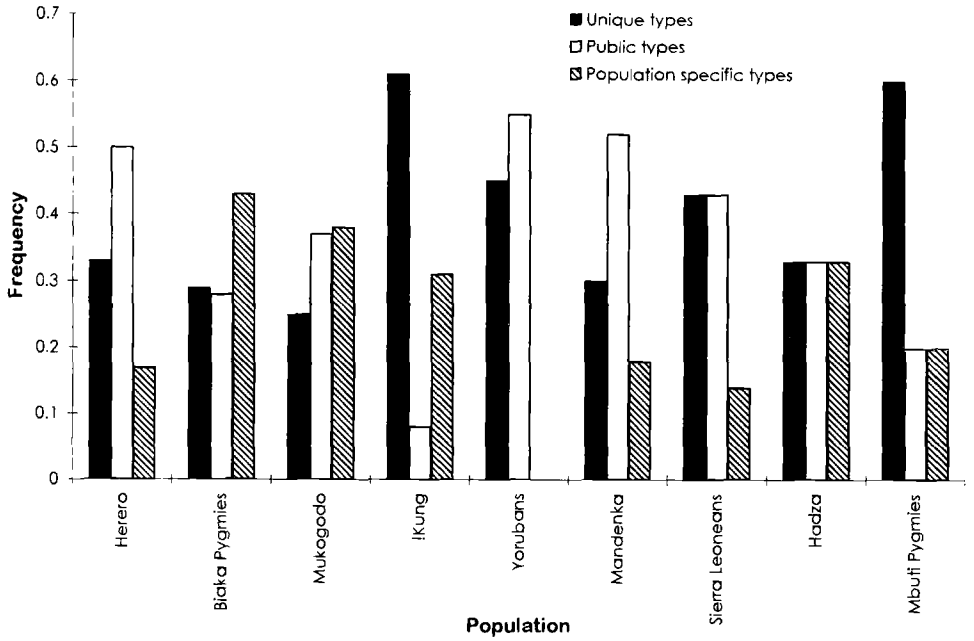

FIG. 3—*Frequency distribution of unique types, population specific types, and shared types among populations. Frequency of each type was calculated relative to the total number of SSO types in the population.*

TABLE 1—*MtDNA SSO type and sequence diversity (*h*) for nine African populations.*

Population	N*	SSO Types N†	SSO Types *h* ± S.E.	Sequences N‡	Sequences *h* ± S.E.
Herero	28	6	0.386 ± 0.081	8	0.495 ± 0.081
Biaka Pygmies	17	7	0.853 ± 0.029	14	0.971 ± 0.016
Mukogodo	28	8	0.778 ± 0.043	—	—
!Kung	25	13	0.880 ± 0.034	16	0.927 ± 0.024
Yorubans	13	11	0.974 ± 0.014	12	0.987 ± 0.011
Mandenka	119	33	0.870 ± 0.018	60	0.979 ± 0.003
Sierra Leoneans	114	42	0.953 ± 0.006	90	0.991 ± 0.002
Hadza	17	3	0.228 ± 0.091	4	0.596 ± 0.065
Mbuti Pygmies	20	10	0.863 ± 0.035	15	0.968 ± 0.012
Total	381	102	0.961 ± 0.004	219	0.992 ± 0.001

*N = number in population sample.
†N = number of different SSO types.
‡N = number of different sequences.

in Appendix B). In a preliminary AMOVA, the Dama and Nama populations were found to be nearly indistinguishable with respect to their SSO types (p = 0.850), and although they are known to be linguistically and biologically different they are geographically close (H. Soodyall, personal communication), so for further analyses they were merged into a single group, resulting in the 12 populations of Table 4. The distance matrix of the third data set, which contained 477 individuals from these same 12 populations with 118 SSO types (six SSO variant regions; excludes ID and IIA; data in Appendix C) is not shown, as the results were not appreciably different from those in Table 4.

Table 5 summarizes the pairs of populations taken from the three distance matrices which were not significantly different from each other ($p > 0.01$). There is consistency across all three data sets for both the pairs which display homogeneity and their significance values (1000 permutations), indicating that even partial results from SSO types are adequate to detect population similarities where they are present. A single homogeneous geographic cluster was formed by the three western African populations (Mandenka, Sierra Leoneans, and Yorubans), which do not appear to be significantly different from each other ($p \geq 0.119$). For all comparisons the significance values were not high, demonstrating that these populations share some, but not appreciable, SSO type identity. Most populations were significantly different (33 of 36 comparisons for the complete data set, and 58 or 59 of 66 for the partial data sets). Heterogeneity for mtDNA SSO types, therefore, predominates for sub-Saharan African populations.

Results from AMOVA analysis partitioning of total variation into portions attributed to variation within and among populations are shown in Table 6. For the complete data set (eight variant regions), approximately 30% of the total variation was due to the differences among populations, and approximately 70% was attributable to the variation within the populations themselves. In other words, on average, 70% of the observed total variation could be found in any single population, and the remaining variation was due to substructure among populations. Results for the partial data sets were very similar. An AMOVA done for the nucleotide sequence variation present for the IIC variant region, one of those for which there was a high frequency of blanks, indicated that sequence variation due to population substructuring accounted for about 16% of the total variation in this stretch of 18 nucleotides. In this region, therefore, large numbers of population-specific substitutions accounted for the high frequency of blanks. A final analysis on control region sequences, using as genetic distances the absolute count of nucleotide differences between each pair of sequences, indicated that when considering the additional variation which is present over SSO type variation, and inclusive of the variation which is present due to SSO blanks, nearly 34% of the total variation is due to differences among populations (a similar analysis using Tamura-Nei genetic distances increased this value to 35.5%). All these analyses had significance values of $p <$

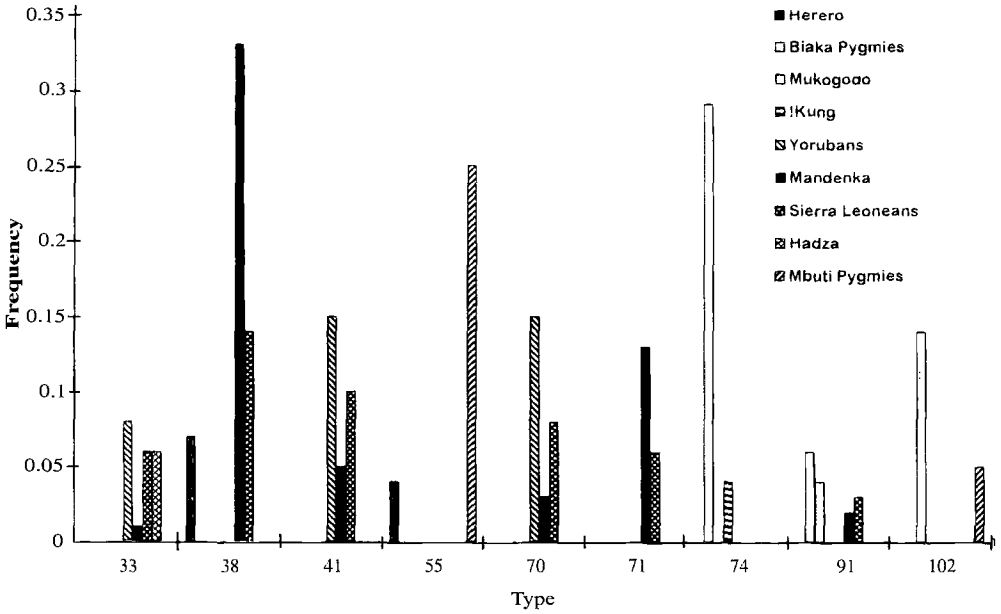

FIG. 4—*Frequencies of the nine most common SSO types in populations.*

TABLE 2—*Frequency of blank variants for 12 African populations.*

Population	N	IA	IB	IC	ID	IIA	IIB	IIC	IID
Herero	36	0.72	0.08	0.64	0.81*	0*	0.17	0.25	0
Biaka Pygmies	17	0	0.29	0.24	0.71	0	0.53	1.00	0.24
Dama/Nama	23	0.65	0.26	0.18	na†	0‡	0.61	0.52	0.04
Mukogodo	28	0	0	0	0	0	0.36	0.21	0.43
!Kung	25	0.04	0.16	0	0.20	0	0.60	0.32	0.12
Yorubans	13	0.15	0	0.15	0	0	0.31	0.46	0
Southeastern Bantu	35	0.43	0.14	0.17	na†	0§	0.54	0.43	0
Sekele	30	0.27	0.33	0.03	na†	0‖	0.87	0.67	0
Mandenka	119	0.13	0.12	0.10	0.11	0	0.63	0.53	0.02
Sierra Leoneans	114	0.13	0.06	0.15	0.17	0	0.61	0.52	0.03
Hadza	17	0	0	0.12	0.88	0	0.12	0	0.88
Mbuti Pygmies	20	0.05	0.05	0	0	0	0.05	0.05	0
Total	477	0.21	0.12	0.15	0.23¶	0**	0.52	0.45	0.08

*N = 31.
†Data not available.
‡N = 13.
§N = 32.
‖N = 24.
¶N = 381.
**N = 453.

TABLE 3—*Complete data set (9 populations, 8 variant regions, N = 381): Φ_{ST} between pairs of populations are shown in lower left-hand matrix; p-values based on 1000 permutations are shown in upper right-hand matrix. 1-Herero, 2-Biaka Pygmies, 4-Mukogodo, 5-!Kung, 6-Yorubans, 9-Mandenka, 10-Sierra Leoneans, 11-Hadza, 12-Mbuti Pygmies.*

	1	2	4	5	6	9	10	11	12
1	—	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.583	—	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.580	0.354	—	0.000	0.000	0.000	0.000	0.000	0.000
5	0.718	0.312	0.312	—	0.000	0.000	0.000	0.000	0.000
6	0.441	0.387	0.218	0.445	—	0.119	0.289	0.000	0.003
9	0.393	0.375	0.257	0.417	0.027	—	0.150	0.000	0.000
10	0.367	0.316	0.201	0.347	0.009	0.005	—	0.000	0.000
11	0.796	0.639	0.323	0.614	0.582	0.515	0.460	—	0.000
12	0.537	0.402	0.259	0.466	0.141	0.243	0.198	0.648	—

TABLE 4—*Partial data set (12 populations, 7 variant regions, N = 453): Φ_{ST} between pairs of populations are shown in lower left-hand matrix; p-values based on 1000 permutations are shown in upper right-hand matrix. 1-Herero, 2-Biaka Pygmies, 3-Dama/Nama, 4-Mukogodo, 5-!Kung, 6-Yorubans, 7-Southeastern Bantu, 8-Sekele, 9-Mandenka, 10-Sierra Leoneans, 11-Hadza, 12-Mbuti Pygmies.*

	1	2	3	4	5	6	7	8	9	10	11	12
1	—	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.588	—	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.396	0.271	—	0.003	0.000	0.096	0.068	0.023	0.001	0.006	0.000	0.000
4	0.541	0.326	0.172	—	0.000	0.006	0.000	0.000	0.000	0.000	0.002	0.002
5	0.705	0.297	0.326	0.288	—	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.299	0.358	0.048	0.181	0.468	—	0.112	0.004	0.142	0.342	0.000	0.000
7	0.386	0.273	0.046	0.186	0.328	0.029	—	0.000	0.002	0.015	0.000	0.000
8	0.519	0.239	0.086	0.196	0.348	0.122	0.114	—	0.001	0.000	0.000	0.000
9	0.316	0.369	0.123	0.243	0.448	0.029	0.059	0.108	—	0.139	0.000	0.000
10	0.303	0.312	0.087	0.179	0.377	0.006	0.030	0.090	0.005	—	0.000	0.000
11	0.793	0.679	0.505	0.221	0.614	0.542	0.476	0.567	0.505	0.451	—	0.000
12	0.448	0.378	0.308	0.241	0.497	0.152	0.240	0.333	0.265	0.216	0.633	—

TABLE 5—*Populations which are not significantly different by AMOVA analyses of complete and partial data sets of SSO types (p > 0.01).*

Comparison	8 variant regions	7 variant regions	6 variant regions
Yorubans vs Mandenka	$p = 0.119$	$p = 0.142$	$p = 0.121$
Yorubans vs Sierra Leoneans	$p = 0.289$	$p = 0.342$	$p = 0.292$
Mandenka vs Sierra Leoneans	$p = 0.150$	$p = 0.139$	$p = 0.174$
Dama/Nama vs Yorubans	*	$p = 0.096$	$p = 0.123$
Dama/Nama vs Southeastern Bantu	*	$p = 0.068$	$p = 0.255$
Dama/Nama vs Sekele	*	$p = 0.023$	$p = 0.006$†
Yorubans vs Southeastern Bantu	*	$p = 0.112$	$p = 0.120$
Southeastern Bantu vs Sierra Leoneans	*	$p = 0.015$	$p = 0.014$

*Data for this comparison not available.
†Significantly different.

TABLE 6—*AMOVA results for complete and partial data sets of SSO types, region IIC sequences, and complete sequences.*

Sites in Control Region	Populations*	Variance Components† %		Φ_{ST}	p (Φ_{ST} and σ^2_a)
		σ^2_a	σ^2_b		
8 variant regions: IA-D, IIA-D	1, 2, 4–6, 9–12	30.44	69.56	0.304	$p < 0.001$
7 variant regions: IA-C, IIA-D	1–12	25.91	74.09	0.259	$p < 0.001$
6 variant regions: IA-C, IIB-D	1–12	24.20	75.80	0.242	$p < 0.001$
Region IIC sequences	1–3, 5–12	16.17	83.83	0.162	$p < 0.001$
Complete sequences	1–3, 5–12	33.90	66.10	0.339	$p < 0.001$

*Populations are: 1-Herero, 2-Biaka Pygmies, 3-Dama/Nama, 4-Mukogodo, 5-!Kung, 6-Yorubans, 7-Southeastern Bantu, 8-Sekele, 9-Mandenka, 10-Sierra Leoneans, 11-Hadza, 12-Mbuti Pygmies.
†σ^2_a = variance among populations, σ^2_b = variance within populations.

0.001 for 1000 permutations of the data, indicating that significant heterogeneity exists for these populations. Figure 5 shows the null distribution of the 1000 variances among populations generated in permutation testing and the observed value ($\sigma^2_a = 0.697$) for the complete SSO type data set.

Discussion

This analysis of mtDNA variation is the last in a three-part continental survey of worldwide variation which encompasses a total of 39 African, Asian, and European populations (1,2). Overall, while continental SSO type diversity was extremely high for all three groups (>0.94), there were obvious differences in the way this variation was apportioned among populations within them. For example, Asian populations displayed statistically significant heterogeneity, but Φ_{ST} values were fairly low (approximately 0.045 for an analysis of 12 Asian populations). Populations geographically close to each other were more likely to be homogeneous, with greater heterogeneity between west and east Asian populations. European populations demonstrated virtually no heterogeneity, with the exception of the contribution of heterogeneity by a single unusual population, the Saami, from Finland. Sub-Saharan African populations, on the other hand, have the greatest heterogeneity thus far observed, and some of the sampled populations display low SSO type diversity, a result not previously observed in the Asian and European populations.

The populations included in this study were both geographically and culturally diverse. Although all the populations were sub-Saharan, they represented the west coast (three populations), central and east Africa (four populations), and southern Africa (six

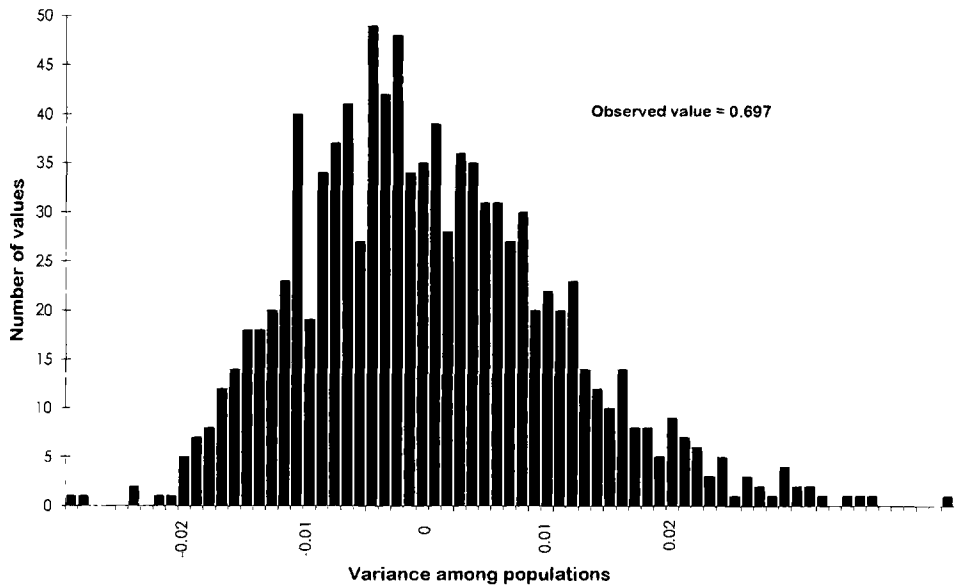

FIG. 5—*Null distribution of the variance component (σ^2_a) obtained through 1000 random permutations of the 381 individuals into nine populations of identical size of those in this study.*

populations). Linguistically, they represented the complete range of major language groupings according to Greenberg's classification system (18): Niger-Kordofanian speakers (Bantu: Biaka Pygmies, Herero, South African Negroid; non-Bantu: Bonkolenken and Bagruwa, Mandenka, Yoruban), Khoisan speakers (Dama, Hadza, !Kung, Nama, Sekele), Nilo-Saharan speakers (Mbuti Pygmies), and Afro-Asiatic speakers (Cushitic: Mukogodo) (19,20).

Diversity estimates for these populations were quite variable. These estimates and the mtDNA type frequency distributions (Figs. 3 and 4) should be viewed cautiously because sample sizes for seven of the nine populations were small which could lead to biased estimates and standard errors (21). In addition, because not all the Mukogodo were maternally unrelated, the estimate of SSO type diversity for this population is probably greater in reality than that reported here. The Herero of Botswana and Hadza of Tanzania had the lowest diversity estimates that we have observed in compiling complete SSO types from 27 populations worldwide. The presence of numerous types which existed at relatively high frequency (> 10%), some of which were also population-specific, was a characteristic of African populations. In the Asian sample, which contained 12 populations and over 900 individuals, only four types were observed in more than 10% of individuals in any population, and these were always shared with other populations. The European data set (6 populations and 595 individuals) displayed no types such as these.

The high frequency of blank variants in the SSO types of Africans overall indicated that additional variation was present at the nucleotide sequence level in regions where SSO probes anneal, either as novel substitution patterns involving probe-specific sites or as nearby substitutions. This pattern has been observed for all populations SSO typed to date for which there have been sequence data available; in addition, there is always further variation at the sequence level for areas in the control region which have not been typed with SSO probes. Therefore, SSO typing has always conservatively estimated control region variation. To maintain consistency in SSO type comparisons among African, Asian, and European populations, we allowed these blanks to represent the baseline variation in Africans, while being cognizant that the undetected variation would increase diversity estimates. Because complete sequence data were available for the majority of individuals in this data set, we were able to observe that out of 449 individuals, there were 287 unique sequences. This set included the 237 sequences deemed unique by MEGA's pairwise comparison as well as 50 additional unique sequences from Sierra Leone which were not used in AMOVA. SSO typing detected 102 different complete SSO types among 381 individuals, for an overall diversity estimate of 0.961 ± 0.004, whereas the comparable sequence diversity for this same group was 0.992 ± 0.001.

All the AMOVA analyses indicated that there was an enormous amount of substructure resulting from the unshared variation present among these populations. Although this was immediately evident in the observation that SSO types are shared among, at most, four populations, AMOVA quantified the degree of population subdivision, indicating that variance among populations always exceeded 24% in analyses using all or part of the control region SSO type data. The net effect of the high frequency of blanks was to slightly increase the appearance of similarity at certain variant regions for some populations, or to underestimate the heterogeneity at these sites, making populations appear more similar than they really were. However, the variance among populations increased only slightly, to about 34% of the total variation, when sequence data were used in the analysis. The reason for this small difference in the variance among populations for complete SSO type data versus sequence data is that, although blanks increase homogeneity at specific sites, the complete SSO type acts as a haplotype, which assorts all eight sites in population-specific patterns. Therefore, the Φ_{ST} values observed for SSO types and sequences both accurately represent the extreme heterogeneity present in the mtDNA of African populations, even though SSO typing underestimates the diversity.

To date, we have found only minor limitations in the use of SSO typing as a forensic DNA typing method. Occasional populations such as the Saami from Finland or aboriginal Taiwanese were observed to be unique with respect to their SSO types, almost certainly a result of an initial bottleneck in variation followed

by population isolation (1,2,22). Heterogeneity was low in Asian populations and was nonexistent in European and European-derived populations. A forensic typing locus, which in this study was the mtDNA control region as described by an individual SSO type, should have both high diversity and low subpopulation heterogeneity. In continental African populations, we have seen widespread exceptions to these traits for SSO types: 1) occasional low population diversity, 2) SSO types which appear at high frequency and are population specific, and 3) substantial subpopulation heterogeneity. At the sequence level, whereas diversity is generally higher (although it remains low in those populations already mentioned), significant substructure is still evident.

These results raise questions about the patterns of mtDNA variation in forensically significant African-derived populations in North America. We observed that the cluster of western African populations in this study (Mandenka, Sierra Leoneans, and Yorubans) had high diversity and no heterogeneity; African-American populations are known to have their sources in part from these regions (23). Extensive sampling of African-American groups will be necessary to determine the degree of similarity with western Africans, and in turn, whether these desirable attributes for use of SSO typing remain present. In addition, we have included in this study a number of populations which would be unlikely to be forensically significant in North America, such as Biaka and Mbuti Pygmies, the hunter-gatherer Herero, Sekele, and Mukogodo, and the Hadza, a Khoisan linguistic isolate tribe. Inclusion of these groups illustrates the range of variation present in Africa, but it is unlikely that their mtDNA types are present at high frequency in North America.

We observed that SSO typing as designed for this worldwide survey from preliminary data on worldwide variation (3) was somewhat inadequate to detect all the nucleotide variation in this sample of Africans. The large numbers of control region nucleotide substitutions are consistent with the antiquity of African populations relative to other continental populations (24,25). Because even within short stretches of nucleotides there were many possible sequence variants, as were seen in the IIC region, many more probes would have to be designed to successfully identify even the more common ones. There may be diminishing returns in doing this, because the object of SSO typing is to avoid time-consuming sequencing unless an exclusion is not obtained. Therefore, control region sequencing would be a good alternative for forensic identifications in African or African-derived populations where there is uncertainty about whether subpopulations are present, at least until further populations are studied. With sequence diversity estimates greater than 0.92 for most single populations, and greater than 0.99 overall, mtDNA sequence typing, where identity can be compared over several hundred nucleotide positions, is appropriate in forensic cases in which other DNA typing methods may not be possible.

Acknowledgments

We are grateful to Sudhir Kumar for computational assistance and Sean Walsh and Roche Molecular Systems for providing SSO probes. We also thank Lee Cronk for his contribution of the Kenyan Mukogodo samples, and appreciate the contribution from Bryan Sykes of facilities of the Institute of Molecular Medicine during the final stages of manuscript preparation.

References

1. Melton T, Stoneking M. Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. J Forensic Sci 1996;41:591–602.
2. Melton T, Wilson M, Batzer M, Stoneking M. Extent of heterogeneity in mitochondrial DNA of European populations. J Forensic Sci 1997;42(4):582–92.
3. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hu Gen 1991;48:370–82.
4. Ginther C, Issel-Tarver L, King MC. Identifying individuals by sequencing mitochondrial DNA from teeth. Nat Genet 1992;2:135–8.
5. Holland MM, Fisher DL, Mitchell LG, Rodgriquez WC, Canik JJ, Merril CR, et al. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. J Forensic Sci 1993;38:542–53.
6. Gill P, Kimpton C, Aliston-Greiner R, Sullivan K, Stoneking M, Melton T, et al. Establishing the identity of Anna Anderson Manahan. Nat Genet 1995;9:9–10.
7. Wilson MR, Stoneking M, Holland MM, DiZinno JA, Budowle B. Guidelines for the use of mitochondrial DNA sequencing in forensic science. Crime Lab Dig 1993;20:68–77.
8. Wilson MR, Polanskey D, Butler J, DiZinno JA, Repogle J, Budowle B. Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. BioTech 1995;18:662–9.
9. Stoneking M, Melton T. Forensic applications of mitochondrial DNA analysis. In: Budowle B, editor, Forensic Applications of PCR. In press.
10. Vigilant L. Control region sequences from African populations and the evolution of human mitochondrial DNA. Ph.D. thesis, University of California, Berkeley, 1986.
11. Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, et al. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 1995;12:334–45.
12. Cronk L. From hunters to herders: subsistence change as a reproductive strategy among the Mukogodo. Current Anthropology 1989;30:224–34.
13. Melton T, Cronk L, Smith J, Stoneking M. Mitochondrial DNA variation in the Mukogodo of central Kenya. Proceedings of the American Anthropological Association Meeting (abstract), Washington DC, November 15, 1995.
14. Soodyall H. Mitochondrial DNA polymorphisms in southern African populations. Ph.D. thesis, University of the Witwatersrand, Johannesburg 1993.
15. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 1992;131:479–91.
16. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 1993;6:447–59.
17. Kumar S, Tamura K, Nei M. MEGA: Molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park.
18. Greenberg JH. The languages of Africa. Indiana University Press, Bloomington, 1963.
19. Ehret C. Southern Nilotic history: linguistic approaches to the study of the past. Northwestern University Press, Evanston, 1971.
20. Ehret C. Ethiopians and East Africans: the problem of contacts. East Africa Publishing House, Nairobi, 1974.
21. Nei M. Molecular evolutionary genetics. Columbia University Press, New York, 1987.
22. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus M-L, et al. Genes and languages in Europe: An analysis of mitochondrial lineages. Genome Research 1995;5:42–52.
23. Curtin PD. The Atlantic slave trade: a census. University of Wisconsin Press, Madison, 1969.
24. Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 1989;86:9350–4.
25. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. African populations and the evolution of human mitochondrial DNA. Science 1991;253:1503–7.

Additional information and reprint requests:
Terry Melton, Ph.D.
Department of Anthropology
409 Carpenter Building
University Park, PA 16802

Appendix A
African SSO Types: Eight Variant Regions

Complete SSO types (eight variant regions) and distribution in nine African populations. The first column shows the designated SSO type number and profile. The numbered columns show the count in each population: 1-Herero, 2-Mbuti Pygmies, 4-Mukogodo, 5-!Kung, 6-Yorubans, 9-Mandenka, 10-Sierra Leone, 11-Hadza, 12-Biaka Pygmies. The final column shows the total across all populations.

		Population									
SSO type		1	2	4	5	6	9	10	11	12	Grand Total
1	00112311	0	0	0	0	0	3	0	0	0	3
2	00312111	0	0	0	1	0	0	0	0	0	1
3	02002301	1	0	0	0	0	0	0	0	0	1
4	02002311	22	0	0	0	0	0	0	0	0	22
5	02012301	0	0	0	0	0	0	1	0	0	1
6	02022311	0	0	0	0	0	2	0	0	0	2
7	02102111	0	0	0	0	0	1	0	0	0	1
8	02102311	0	0	0	0	0	4	1	0	0	5
9	02112011	0	0	0	0	1	1	2	0	0	4
10	02112301	0	0	0	0	0	0	1	0	0	1
11	02112311	0	0	0	0	0	0	1	0	0	1
12	02112321	0	0	0	0	0	2	1	0	0	3
13	02122001	0	0	0	0	0	0	1	0	0	1
14	02122111	0	0	0	0	0	0	4	0	0	4
15	02122121	0	0	0	0	0	2	3	0	0	5
16	02312021	0	0	0	0	0	0	0	0	1	1
17	02312311	0	0	0	0	1	0	0	0	0	1
18	02322111	0	0	0	0	0	1	0	0	0	1
19	03102311	1	0	0	0	0	0	0	0	0	1
20	10012021	0	0	0	0	0	2	0	0	0	2
21	10112001	0	0	0	0	0	1	0	0	0	1
22	10112011	0	0	0	0	0	1	3	0	0	4
23	10112101	0	0	0	0	0	0	1	0	0	1
24	10122111	0	0	0	0	0	1	0	0	0	1
25	10312002	0	0	0	1	0	0	0	0	0	1
26	10312101	0	0	0	0	0	0	1	0	0	1
27	11002021	0	0	0	0	0	0	1	0	0	1
28	11122111	0	0	0	0	0	1	0	0	0	1
29	11312011	0	0	0	0	0	1	0	0	0	1
30	12002021	0	0	0	0	0	1	4	0	0	5
31	12012001	0	0	0	0	0	5	0	0	0	5
32	12012011	0	0	0	0	0	0	0	1	0	1
33	12012021	0	0	0	0	1	1	7	1	0	10
34	12012301	0	0	0	0	0	0	1	0	0	1
35	12022021	0	0	0	0	0	0	1	0	0	1
36	12102001	0	0	0	0	0	1	0	0	0	1
37	12102021	0	0	0	0	0	0	2	0	0	2
38	12112001	2	0	0	0	0	39	16	0	0	57
39	12112011	0	0	0	0	0	4	2	0	0	6
40	12112012	0	0	2	0	0	0	0	0	0	2
41	12112021	0	0	0	0	2	6	11	0	0	19
42	12112311	0	0	1	0	0	0	0	0	0	1
43	12122001	0	0	0	0	0	3	1	0	0	4
44	12122111	0	0	0	0	1	1	1	0	0	3
45	12302001	0	0	0	0	0	0	2	0	0	2
46	12302210	0	0	0	0	0	0	0	15	0	15
47	12311112	0	0	0	0	0	0	0	0	6	6
48	12311302	0	3	0	0	0	0	0	0	0	3
49	12312001	0	0	5	0	0	0	0	0	0	5
50	12312002	0	0	0	1	0	0	0	0	0	1
51	12312021	0	0	1	0	0	0	0	0	0	1
52	12312101	0	0	0	0	0	0	6	0	0	6
53	12312111	0	0	0	0	0	0	0	0	1	1
54	12312301	0	0	0	0	0	0	0	0	1	1
55	12312311	1	0	0	0	0	0	0	0	5	6
56	12312321	0	0	0	0	0	0	0	0	2	2
57	12322011	0	0	2	0	0	0	1	0	0	3
58	12322210	0	0	12	0	0	0	0	0	0	12
59	12322311	0	0	0	0	0	0	0	0	1	1
60	12322322	0	0	0	0	0	1	0	0	0	1
61	20312101	0	0	0	0	0	0	1	0	0	1
62	22012302	0	0	0	0	0	1	0	0	0	1
63	22012322	0	0	0	0	1	0	0	0	0	1
64	22302322	0	0	0	0	0	1	0	0	0	1
65	22312002	0	0	0	0	0	0	1	0	0	1
66	22312022	0	0	0	0	0	4	0	0	0	4
67	22312101	0	0	0	0	1	0	0	0	0	1
68	22312102	0	0	0	0	1	0	1	0	0	2
69	22312122	0	0	0	0	0	0	1	0	0	1
70	22312302	0	0	0	0	2	3	9	0	0	14
71	22312322	0	0	0	0	0	15	7	0	0	22
72	30112001	0	0	0	0	0	0	1	0	0	1
73	30122001	0	0	0	0	0	3	0	0	0	3
74	30302002	0	5	0	1	0	0	0	0	0	6
75	30302302	1	0	0	0	0	3	0	0	0	4
76	30312022	0	0	0	1	0	0	0	0	0	1
77	30322322	0	0	0	0	0	0	0	0	1	1
78	31012021	0	0	0	0	0	0	2	0	0	2
79	31302002	0	0	0	0	0	0	1	0	0	1
80	31312002	0	0	0	1	0	0	0	0	0	1
81	31312101	0	0	0	0	1	0	0	0	0	1
82	32002000	0	4	0	0	0	0	0	0	0	4
83	32102002	0	0	0	0	0	0	1	0	0	1
84	32112001	0	0	0	0	0	0	1	0	0	1
85	32302000	0	0	0	0	0	2	3	0	0	5
86	32302002	0	0	0	0	0	0	4	0	0	4
87	32302022	0	0	0	4	0	0	0	0	0	4
88	32302102	0	2	0	0	0	0	0	0	0	2
89	32302302	0	1	0	0	0	0	0	0	0	1
90	32311102	0	0	0	0	0	0	1	0	0	1
91	32311302	0	1	1	0	0	2	3	0	0	7
92	32312000	0	0	0	1	0	0	0	0	0	1
93	32312001	0	0	0	0	0	0	1	0	0	1
94	32312002	0	0	0	2	0	0	0	0	0	2
95	32312020	0	0	0	2	0	0	0	0	0	2
96	32312022	0	0	0	1	0	0	0	0	0	1
97	32312101	0	0	0	0	1	0	0	0	0	1
98	32312102	0	1	0	0	0	0	0	0	0	1
99	32312202	0	0	0	1	0	0	0	0	0	1
100	32312222	0	0	0	8	0	0	0	0	0	8
101	32312322	0	0	0	0	0	0	0	0	1	1
102	32322322	0	0	4	0	0	0	0	0	1	5
Grand Total		28	17	28	25	13	119	114	17	20	381

Appendix B
African SSO Types: Seven Variant Regions

Partial SSO types (seven variant regions) and distribution in 12 populations. The first column shows the designated SSO type number and profile. The numbered columns show the count in each population: 1-Herero, 2-Mbuti Pygmies, 3-Dama/Nama, 4-Mukogodo, 5-!Kung, 6-Yorubans, 7-Eastern South Africans, 8-Sekele, 9-Mandenka, 10-Sierra Leone, 11-Hadza, 12-Biaka Pygmies. The final column shows the total across all populations.

		Population												
SSO type		1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
1	0012001	0	0	2	0	0	0	0	2	0	0	0	0	4
2	0012311	0	0	0	0	0	0	0	0	3	0	0	0	3
3	0031002	0	0	1	0	0	0	0	0	0	0	0	0	1
4	0032002	0	0	0	0	0	0	1	0	0	0	0	0	1
5	0032022	0	0	0	0	0	0	0	1	0	0	0	0	1
6	0032111	0	0	0	0	1	0	0	0	0	0	0	0	1
7	0032311	0	0	0	0	0	0	0	1	0	0	0	0	1
8	0112222	0	0	0	0	0	0	1	0	0	0	0	0	1
9	0132032	0	0	0	0	0	0	0	1	0	0	0	0	1
10	0202102	0	0	0	0	0	0	1	0	0	0	0	0	1
11	0202111	0	0	1	0	0	0	1	0	0	0	0	0	2
12	0202301	1	0	0	0	0	0	0	0	0	1	0	0	2
13	0202311	22	0	1	0	0	0	1	0	2	0	0	0	26
14	0211102	0	0	0	0	0	0	1	0	0	0	0	0	1
15	0211302	0	0	0	0	0	0	1	0	0	0	0	0	1
16	0212001	0	0	0	0	0	0	0	0	0	1	0	0	1
17	0212002	0	0	0	0	0	0	1	1	0	0	0	0	2
18	0212011	0	0	0	0	0	1	0	0	1	2	0	0	4
19	0212022	0	0	0	0	0	0	1	0	0	0	0	0	1
20	0212111	0	0	0	0	0	0	1	0	1	4	0	0	6
21	0212121	0	0	0	0	0	0	0	0	2	3	0	0	5
22	0212301	0	0	0	0	0	0	0	0	0	1	0	0	1
23	0212311	1	0	0	0	0	0	0	0	4	2	0	0	7
24	0212321	0	0	0	0	0	0	0	0	2	1	0	0	3
25	0231102	0	0	0	0	0	0	2	0	0	0	0	0	2
26	0232002	0	0	0	0	0	0	0	1	0	0	0	0	1
27	0232012	0	0	2	0	0	0	0	0	0	0	0	0	2
28	0232021	0	0	0	0	0	0	0	0	0	0	0	1	1
29	0232022	0	0	2	0	0	0	0	0	0	0	0	0	2
30	0232111	0	0	0	0	0	0	0	0	1	0	0	0	1
31	0232121	0	0	0	0	0	0	1	0	0	0	0	0	1
32	0232210	0	0	1	0	0	0	0	0	0	0	0	0	1
33	0232311	0	0	0	0	0	1	0	0	0	0	0	0	1
34	0312311	1	0	0	0	0	0	0	0	0	0	0	0	1
35	1002021	0	0	0	0	0	0	0	0	2	0	0	0	2
36	1012001	1	0	0	0	0	0	0	0	1	0	0	0	2
37	1012002	0	0	0	0	0	0	0	1	0	0	0	0	1
38	1012011	0	0	0	0	0	0	0	0	1	3	0	0	4
39	1012101	0	0	0	0	0	0	0	0	0	1	0	0	1
40	1012111	0	0	0	0	0	0	0	0	1	0	0	0	1
41	1012202	0	0	1	0	0	0	0	0	0	0	0	0	1
42	1012222	0	0	0	0	0	0	1	0	0	0	0	0	1
43	1032002	0	0	0	0	1	0	0	2	0	0	0	0	3
44	1032022	0	0	0	0	0	0	1	1	0	0	0	0	2
45	1032101	0	0	0	0	0	0	0	0	0	1	0	0	1
46	1032302	0	0	0	0	0	0	0	1	0	0	0	0	1
47	1102021	0	0	0	0	0	0	0	0	0	1	0	0	1
48	1112111	0	0	0	0	0	0	0	0	1	0	0	0	1
49	1132011	0	0	0	0	0	0	0	0	1	0	0	0	1
50	1202001	0	0	0	0	0	0	0	0	5	0	0	0	5
51	1202011	0	0	0	0	0	0	0	0	0	0	1	0	1
52	1202021	0	0	0	0	0	1	2	0	2	12	1	0	18
53	1202022	0	0	0	0	0	0	1	0	0	0	0	0	1
54	1202301	0	0	0	0	0	0	0	0	0	1	0	0	1
55	1211002	0	0	0	0	0	0	1	0	0	0	0	0	1
56	1211021	0	0	0	0	0	0	2	0	0	0	0	0	2
57	1211312	0	0	0	0	0	0	1	0	0	0	0	0	1
58	1212001	2	0	1	0	0	0	2	3	43	17	0	0	68
59	1212011	0	0	0	0	0	0	0	0	4	2	0	0	6
60	1212012	0	0	0	2	0	0	0	0	0	0	0	0	2
61	1212021	0	0	0	0	0	2	3	1	6	13	0	0	25
62	1212022	0	0	0	0	0	0	1	0	0	0	0	0	1
63	1212101	0	0	1	0	0	0	0	0	0	0	0	0	1
64	1212102	0	0	0	0	0	0	1	0	0	0	0	0	1
65	1212111	0	0	0	0	0	1	0	0	1	1	0	0	3
66	1212311	1	0	0	1	0	0	0	0	0	0	0	0	2
67	1231112	0	0	0	0	0	0	0	0	0	0	0	6	6
68	1231302	0	3	0	0	0	0	1	0	0	0	0	0	4
69	1232001	0	0	0	5	0	0	1	2	0	2	0	0	10
70	1232002	0	0	0	0	1	0	0	4	0	0	0	0	5
71	1232011	0	0	0	2	0	0	0	0	0	1	0	0	3
72	1232021	0	0	0	1	0	0	1	0	0	0	0	0	2
73	1232022	0	0	0	0	0	0	0	2	0	0	0	0	2
74	1232101	0	0	0	0	0	0	0	0	0	6	0	0	6
75	1232111	0	0	0	0	0	0	0	0	0	0	0	1	1
76	1232210	0	0	0	12	0	0	0	0	0	0	15	0	27
77	1232301	0	0	0	0	0	0	0	0	0	0	0	1	1
78	1232311	1	0	0	0	0	0	0	0	0	0	0	6	7
79	1232321	0	0	0	0	0	0	0	0	0	0	0	2	2
80	1232322	0	0	0	0	0	0	0	0	0	1	0	0	1
81	2032101	0	0	0	0	0	0	0	0	0	1	0	0	1
82	2202302	0	0	0	0	0	0	0	0	1	0	0	0	1
83	2202322	0	0	0	0	0	1	0	0	0	0	0	0	1
84	2232002	0	0	0	0	0	0	0	0	0	1	0	0	1
85	2232022	0	0	0	0	0	0	0	0	4	0	0	0	4
86	2232101	0	0	0	0	0	1	0	0	0	0	0	0	1
87	2232102	0	0	0	0	0	1	0	0	0	1	0	0	2
88	2232122	0	0	0	0	0	0	0	0	0	1	0	0	1
89	2232302	0	0	0	0	0	2	0	0	3	9	0	0	14
90	2232322	0	0	0	0	0	0	0	0	16	7	0	0	23
91	3012001	0	0	0	0	0	0	0	0	3	1	0	0	4
92	3032002	0	5	0	0	1	0	0	0	0	0	0	0	6
93	3032022	0	0	0	0	1	0	0	0	0	0	0	0	1
94	3032302	1	0	0	0	0	0	0	0	3	0	0	0	4
95	3032322	0	0	0	0	0	0	0	0	0	0	0	1	1
96	3102021	0	0	0	0	0	0	0	0	0	2	0	0	2
97	3132002	0	0	0	0	1	0	0	0	0	1	0	0	2
98	3132101	0	0	0	0	0	1	0	0	0	0	0	0	1
99	3202000	0	4	0	0	0	0	0	0	0	0	0	0	4
100	3212001	0	0	0	0	0	0	0	0	0	0	1	0	1
101	3212002	0	0	0	0	0	0	0	0	0	1	0	0	1
102	3231102	0	0	0	0	0	0	0	0	0	1	0	0	1
103	3231302	0	1	0	1	0	0	0	0	2	3	0	0	7
104	3232000	0	0	0	0	1	0	0	0	2	3	0	0	6
105	3232001	0	0	0	0	0	0	0	0	0	1	0	0	1
106	3232002	0	0	0	0	2	0	0	0	0	4	0	0	6
107	3232020	0	0	0	0	2	0	0	0	0	0	0	0	2
108	3232022	0	0	0	0	5	0	0	0	0	0	0	0	5
109	3232101	0	0	0	0	0	1	0	0	0	0	0	0	1
110	3232102	0	3	0	0	0	0	0	0	0	0	0	0	3
111	3232202	0	0	0	0	1	0	0	0	0	0	0	0	1
112	3232222	0	0	0	0	8	0	0	0	0	0	0	0	8
113	3232302	0	1	0	0	0	0	0	0	0	0	0	0	1
114	3232322	0	0	0	4	0	0	0	0	0	0	0	2	6
Grand Total		31	17	13	28	25	13	32	24	119	114	17	20	453

Appendix C
African SSO Types: Six Variant Regions

Partial SSO types (six variant regions) and distribution in 12 populations. The first column shows the designated SSO type number and profile. The numbered columns show the count in each population: 1-Herero, 2-Mbuti Pygmies, 3-Dama/Nama, 4-Mukogodo, 5-!Kung, 6-Yorubans, 7-Eastern South Africans, 8-Sekele, 9-Mandenka, 10-Sierra Leone, 11-Hadza, 12-Biaka Pygmies. The final column shows the total across all populations.

		Population												
SSO type		1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
1	001001	0	0	2	0	0	0	0	2	0	0	0	0	4
2	001212	0	0	0	0	0	0	1	0	0	0	0	0	1
3	001311	0	0	0	0	0	0	0	0	3	0	0	0	3
4	002002	0	0	1	0	0	0	0	0	0	0	0	0	1
5	003001	0	0	1	0	0	0	0	0	0	0	0	0	1
6	003002	0	0	1	0	0	0	1	0	0	0	0	0	2
7	003022	0	0	0	0	0	0	0	1	0	0	0	0	1
8	003111	0	0	0	0	1	0	0	0	0	0	0	0	1
9	003311	0	0	0	0	0	0	0	1	0	0	0	0	1
10	011202	0	0	0	0	0	0	1	0	0	0	0	0	1
11	011222	0	0	0	0	0	0	1	0	0	0	0	0	1
12	013032	0	0	0	0	0	0	0	1	0	0	0	0	1
13	013301	0	0	1	0	0	0	0	0	0	0	0	0	1
14	020102	0	0	0	0	0	0	1	0	0	0	0	0	1
15	020111	0	0	1	0	0	0	1	0	0	0	0	0	2
16	020301	1	0	0	0	0	0	0	0	0	1	0	0	2
17	020311	22	0	1	0	0	0	1	0	2	0	0	0	26
18	021001	0	0	0	0	0	0	0	0	0	1	0	0	1
19	021002	0	0	1	0	0	0	1	1	0	0	0	0	3
20	021011	0	0	0	0	0	1	0	0	1	2	0	0	4
21	021022	0	0	0	0	0	0	1	0	0	0	0	0	1
22	021102	0	0	0	0	0	0	1	0	0	0	0	0	1
23	021111	0	0	0	0	0	0	1	0	1	4	0	0	6
24	021121	0	0	0	0	0	0	0	0	2	3	0	0	5
25	021301	0	0	0	0	0	0	0	0	0	1	0	0	1
26	021302	0	0	0	0	0	0	1	0	0	0	0	0	1
27	021311	1	0	0	0	0	0	0	0	4	2	0	0	7
28	021321	0	0	0	0	0	0	0	0	2	1	0	0	3
29	023002	0	0	1	0	0	0	0	1	0	0	0	0	2
30	023012	0	0	2	0	0	0	0	0	0	0	0	0	2
31	023021	0	0	0	0	0	0	0	0	0	0	0	1	1
32	023022	0	0	2	0	0	0	0	0	0	0	0	0	2
33	023102	0	0	0	0	0	0	2	0	0	0	0	0	2
34	023111	0	0	0	0	0	0	0	0	1	0	0	0	1
35	023121	0	0	0	0	0	0	1	0	0	0	0	0	1
36	023210	0	0	1	0	0	0	0	0	0	0	0	0	1
37	023311	1	0	0	0	0	1	0	1	0	0	0	0	3
38	031311	1	0	0	0	0	0	0	0	0	0	0	0	1
39	100021	0	0	0	0	0	0	0	0	2	0	0	0	2
40	101001	1	0	0	0	0	0	0	0	1	0	0	0	2
41	101002	1	0	0	0	0	0	0	1	0	0	0	0	2
42	101011	0	0	0	0	0	0	0	0	1	3	0	0	4
43	101101	0	0	0	0	0	0	0	0	0	1	0	0	1
44	101111	0	0	0	0	0	0	0	0	1	0	0	0	1
45	101202	0	0	1	0	0	0	0	0	0	0	0	0	1
46	101222	0	0	0	0	0	0	1	0	0	0	0	0	1
47	103002	0	0	0	0	1	0	1	3	0	0	0	0	5
48	103022	0	0	0	0	0	0	1	1	0	0	0	0	2
49	103101	0	0	0	0	0	0	0	0	0	1	0	0	1
50	103302	0	0	0	0	0	0	0	1	0	0	0	0	1
51	110021	0	0	0	0	0	0	0	0	0	1	0	0	1
52	111111	0	0	0	0	0	0	0	0	1	0	0	0	1
53	112311	0	0	1	0	0	0	0	0	0	0	0	0	1
54	113011	0	0	0	0	0	0	0	0	0	1	0	0	1
55	120001	0	0	0	0	0	0	0	0	5	0	0	0	5
56	120011	0	0	0	0	0	0	0	0	0	0	1	0	1
57	120021	0	0	0	0	0	1	2	0	2	12	1	0	18
58	120022	0	0	0	0	0	0	1	0	0	0	0	0	1
59	120221	0	0	0	0	0	0	0	1	0	0	0	0	1
60	120222	0	0	1	0	0	0	0	0	0	0	0	0	1
61	120301	0	0	0	0	0	0	0	0	0	0	1	0	1
62	121001	3	0	2	0	0	0	2	3	43	17	0	0	70
63	121002	0	0	0	0	0	0	1	0	0	0	0	0	1
64	121011	0	0	0	0	0	0	0	0	4	2	0	0	6
65	121012	0	0	0	2	0	0	0	0	0	0	0	0	2
66	121021	0	0	0	0	0	2	5	1	6	13	0	0	27
67	121022	0	0	0	0	0	0	1	0	0	0	0	0	1
68	121101	0	0	1	0	0	0	0	0	0	0	0	0	1
69	121102	0	0	0	0	0	0	1	0	0	0	0	0	1
70	121111	0	0	0	0	0	1	0	0	1	1	0	0	3
71	121311	1	0	0	1	0	0	0	0	0	0	0	0	2
72	121312	0	0	1	0	0	0	1	0	0	0	0	0	2
73	123001	1	0	0	5	0	0	1	2	0	2	0	0	11
74	123002	0	0	0	0	1	0	0	6	0	0	0	0	7
75	123011	0	0	1	2	0	0	0	0	0	0	1	0	4
76	123021	0	0	0	1	0	0	1	0	0	0	0	0	2
77	123022	0	0	0	0	0	0	0	3	0	0	0	0	3
78	123101	0	0	0	0	0	0	0	0	0	6	0	0	6
79	123111	0	0	0	0	0	0	0	0	0	0	0	1	1
80	123112	0	0	0	0	0	0	0	0	0	0	0	6	6
81	123210	0	0	0	12	0	0	0	0	0	0	15	0	27
82	123301	1	0	0	0	0	0	0	0	0	0	0	1	2
83	123302	0	3	0	0	0	0	1	0	0	0	0	0	4
84	123311	1	0	0	0	0	0	0	0	0	0	0	6	7
85	123321	0	0	0	0	0	0	0	0	0	0	0	2	2
86	123322	0	0	0	0	0	0	0	0	1	0	0	0	1
87	203101	0	0	0	0	0	0	0	0	0	1	0	0	1
88	220302	0	0	0	0	0	0	0	0	1	0	0	0	1
89	220322	0	0	0	0	0	1	0	0	0	0	0	0	1
90	223002	0	0	0	0	0	0	0	0	0	1	0	0	1
91	223022	0	0	0	0	0	0	0	0	4	0	0	0	4
92	223101	0	0	0	0	0	1	0	0	0	0	0	0	1
93	223102	0	0	0	0	0	1	0	0	0	1	0	0	2
94	223122	0	0	0	0	0	0	0	0	0	1	0	0	1
95	223302	0	0	0	0	0	2	0	0	3	9	0	0	14
96	223322	0	0	0	0	0	0	0	0	16	7	0	0	23
97	301001	0	0	0	0	0	0	0	0	3	1	0	0	4
98	303002	0	5	0	0	1	0	0	0	0	0	0	0	6
99	303022	0	0	0	0	1	0	0	0	0	0	0	0	1
100	303302	1	0	0	0	0	0	0	0	3	0	0	0	4
101	303322	0	0	0	0	0	0	0	0	0	0	0	1	1
102	310021	0	0	0	0	0	0	0	0	0	2	0	0	2
103	313002	0	0	0	0	1	0	0	0	0	1	0	0	2
104	313101	0	0	0	0	0	1	0	0	0	0	0	0	1
105	320000	0	4	0	0	0	0	0	0	0	0	0	0	4
106	321001	0	0	0	0	0	0	0	0	0	1	0	0	1
107	321002	0	0	0	0	0	0	0	0	0	1	0	0	1
108	323000	0	0	0	0	1	0	0	0	2	3	0	0	6
109	323001	0	0	0	0	0	0	0	0	0	1	0	0	1
110	323002	0	0	0	0	2	0	0	0	0	4	0	0	6
111	323020	0	0	0	0	2	0	0	0	0	0	0	0	2
112	323022	0	0	0	0	5	0	0	0	0	0	0	0	5
113	323101	0	0	0	0	0	1	0	0	0	0	0	0	1
114	323102	0	3	0	0	0	0	0	0	0	1	0	0	4
115	323202	0	0	0	0	1	0	0	0	0	0	0	0	1
116	323222	0	0	0	0	8	0	0	0	0	0	0	0	8
117	323302	0	2	0	1	0	0	0	0	2	3	0	0	8
118	323322	0	0	0	4	0	0	0	0	0	0	0	2	6
Grand Total		36	17	23	28	25	13	35	30	119	114	17	20	477