Terry Melton,[1,2] *Ph.D.; Stephanie Clifford,*[2,3] *B.A.; Manfred Kayser,*[2,4] *Ph.D.; Ivane Nasidze,*[2,4] *Ph.D.; Mark Batzer,*[5] *Ph.D.; and Mark Stoneking,*[2,4] *Ph.D.*

Diversity and Heterogeneity in Mitochondrial DNA of North American Populations*

REFERENCE: Melton T, Clifford S, Kayser M, Nasidze I, Batzer M, Stoneking M. Diversity and heterogeneity in mitochondrial DNA of North American populations. J Forensic Sci 2001; 46(1):46–52.

ABSTRACT: Variation in the mitochondrial DNA (mtDNA) control region as detected by sequence-specific oligonucleotide (SSO) probes is described for 2282 individuals from African-American, European-American, and Hispanic subpopulations from five broadly defined regions of North America (Northeast, Southeast, Central, Northwest, Southwest). Population diversity estimates were uniformly high for all subpopulations and for each major ethnic group. Only the Pennsylvania Hispanic group was remarkable with respect to its mitochondrial DNA types, having both six low frequency population specific types (ranging from 1.2–8.6%) and three high frequency shared types (10–20% each). There was no statistically significant subpopulation heterogeneity present within any of the three major groups at either the subpopulation level or the regional level ($p > 0.01$). However, statistically significant heterogeneity was measured when comparing the three major groups to each other, with the variance component attributable to this large division accounting for 18.60% of the total variance ($p < 0.001$). Overall mtDNA is a satisfactory forensic typing locus within broadly defined African-American, European-American, and Hispanic groups from North America, based on the high diversity estimates and absence of heterogeneity, as characterized by SSO typing.

KEYWORDS: forensic science, DNA typing, human mitochondrial DNA, sequence specific oligonucleotide typing, population genetics, North American populations, African-American, European-American, Hispanic

We previously described the population genetics of the human mitochondrial DNA (mtDNA) control region for three continental groups from around the world (1–3). In those reports, we calculated diversity statistics and quantified subpopulation heterogeneity for over 2600 individuals from 39 African, Asian, European, and European-derived populations. The purpose of those studies was to determine the extent of variability of mtDNA and assess whether it was both sufficiently high in diversity and sufficiently lacking in subpopulation heterogeneity to serve as a satisfactory forensic DNA typing tool (4).

For our analyses we used sequence-specific oligonucleotide (SSO) typing, which captures a portion of the total sequence variation in the hypervariable control region and can be a useful population screening tool compared to time-consuming and technically-demanding DNA sequencing. In several studies, SSO typing has been found to perform well as a substitute for DNA sequencing when evaluating the population genetics of large numbers of samples (3,5,6). Overall, we found high diversity in all three continental groups, with the highest diversity estimate in Africans and the lowest in European or European-derived populations. Africans overall had very high levels of subpopulation heterogeneity, whereas Europeans had no statistically-significant heterogeneity. Asians were intermediate between Africans and Europeans, having minimal statistically-significant heterogeneity that followed a west-east geographic pattern. The conclusion of these studies was that mtDNA should be a satisfactory forensic marker for most broadly defined continental groups because of the high diversity within populations and the low heterogeneity between populations from the same geographic region. The exceptional group was that comprised of African subpopulations, which have marked continental heterogeneity and high numbers of subpopulation specific types, although they do display high diversity.

These earlier studies raised the question of how similar analyses of forensically significant North American populations would compare with respect to their ancestral source populations. This study aims to address exactly the same issues of diversity and heterogeneity of mtDNA SSO-types in African-American, European-American, and Hispanic individuals from the Northeast, Southeast, Central, Northwest, and Southwest areas of North America. We also performed heterogeneity measurements on the entire dataset, comprising all subpopulations in this study, to determine what fraction of the total mtDNA variation is attributable to the African-American, European-American, and Hispanic population division.

Methods

The data in this study were obtained from 805 African-Americans (10 populations), 922 European-Americans (11 populations), and 555 Hispanics (seven populations). The ethnic origin of each individual was largely self-reported. All samples came from crime laboratories, with the exception of the Louisiana and Cajun samples (provided by M. Batzer) and the Mexican samples (provided

[1] Mitotyping Technologies, State College, PA.
[2] Department of Anthropology, The Pennsylvania State University, University Park, PA.
[3] Present address: Department of Biology, Trinity University, San Antonio, TX.
[4] Present address: Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
[5] Departments of Pathology, Biometry and Genetics, and Biochemistry and Molecular Biology, Stanley S. Scott Cancer Center, Neuroscience Center of Excellence, Louisiana State University Health Sciences Center, New Orleans, LA.

* This project was supported by NIJ grant 92-IJ-CX-K040 (S-1) to M.S. Portions of this work were presented at the American Academy of Forensic Sciences meeting in San Francisco, 9–14 February 1998.
Received 3 Dec. 1999; accepted 11 Feb. 2000.

by C. Gorodezky). The data for the Mexico and California samples were previously published (5). The remaining samples were received as either purified DNA or dried bloodstains; DNA was extracted from the latter using the IsoQuick extraction kit (Orca Research) according to the supplied protocol.

PCR of the mtDNA control region, immobilization of PCR products on nylon membranes, hybridization of the membrane with SSO probes, and chemiluminescent detection of bound probe were all performed as described in detail previously (5,7). We used 21 SSO probes, of which 17 were identical in sequence to previously-used probes (5), while the sequences of four probes were slightly modified to increase specificity, as described elsewhere (7). These 21 probes detect variation in eight regions of the mtDNA control region, four in the first hypervariable region (IA–ID), and four in the second hypervariable region (IIA–IID), as designated previously (5). The probes are designed to detect variation at either one or two polymorphic sites in each of the eight regions, leading to either two or three probes per region, respectively. Table 1 shows the nucleotide substitutions associated with each of the probe variants.

MtDNA SSO types are designated by an 8-digit number; for example, type 12112301 indicates that the sample typed positive for probes IA1 (and negative for probes IA2 and IA3), IB2, IC1, etc. The "0" for the IIC region indicates that none of the probes for the IIC region hybridized to this sample. Such blank results occur because of additional nucleotide substitutions in the probe region that prevent hybridization. While different individuals typing as a blank for the same probe region could have different substitutions, for the purposes of data analysis the blanks are considered to represent the same variant. The full dataset is available upon request from T. Melton or M. Stoneking.

Each SSO type is therefore treated as an allele at a single haploid locus, with type frequencies equivalent to allele frequencies. Data analyses, including unbiased estimates of diversity, and an analysis of molecular variance (AMOVA) to detect among-population heterogeneity, were carried out as described previously (1–3). For the AMOVA analysis, the Φ-statistics to measure population heterogeneity were calculated by incorporating a genetic distance between each pair of SSO types, and the statistical significance of the Φ-statistics was ascertained by a permutational testing procedure, as described in detail elsewhere (1). The AMOVA for the large data set including all three groups was carried out using the ARLEQUIN computer program (8).

Results

Table 2 summarizes the sample sizes, number of SSO types, and diversity statistics for each of the subpopulations in the three sampled major population groups. Diversity values are uniformly high for all three groups, with the lowest overall values observed in the European-American (0.964) and Hispanic (0.963) groups. Both the European-American subpopulations and Hispanic subpopulations display diversity values in the range of approximately 0.92 to 0.97.

TABLE 1—*Probe variants used in SSO typing and their associated nucleotide substitutions with respect to the published reference sequence (9). For all probe regions, probe 1 is identical to the reference sequence.*

	Hypervariable Region 1				Hypervariable Region 2		
IA	IB	IC	ID	IIA	IIB	IIC	IID
2 = 16126 C 3 = 16129 A	2 = 16223 T 3 = 16217 C	2 = 16304 C 3 = 16311 C	2 = 16362 C	2 = 73 G	2 = 146 C 3 = 152 C	2 = 195 C 3 = 199 C	2 = 247 A

TABLE 2—*MtDNA SSO type diversity (h) for North American populations. Overall total: N = 2282, Number of SSO types = 502, h ± s.e. = 0.998 ± 0.000.*

	African-American Populations			European-American Populations			Hispanic Populations		
	N	Number of SSO Types	h ± s.e.	*N*	Number of SSO Types	h ± s.e.	*N*	Number of SSO Types	h ± s.e.
CENTRAL									
Illinois	68	55	0.993 ± 0.002	42	30	0.974 ± 0.009	…	…	…
Missouri	126	64	0.977 ± 0.003	90	51	0.97 ± 0.006	…	…	…
NORTHEAST									
Maryland	38	27	0.98 ± 0.005	38	26	0.962 ± 0.013	…	…	…
Pennsylvania	93	64	0.988 ± 0.002	105	53	0.958 ± 0.008	115	40	0.925 ± 0.01
Vermont	…	…	…	94	47	0.923 ± 0.016	…	…	…
NORTHWEST									
Oregon	86	55	0.986 ± 0.002	98	54	0.973 ± 0.005	74	45	0.959 ± 0.01
Washington	49	33	0.974 ± 0.007	…	…	…	52	36	0.978 ± 0.005
SOUTHEAST									
Cajun	…	…	…	58	30	0.953 ± 0.01	…	…	…
Louisiana	55	36	0.986 ± 0.008	57	35	0.948 ± 0.014	58	27	0.929 ± 0.014
Virginia	86	58	0.984 ± 0.004	109	56	0.973 ± 0.004	102	52	0.967 ± 0.005
SOUTHWEST									
California	119	62	0.975 ± 0.004	128	72	0.972 ± 0.006	…	…	…
Mexican	…	…	…	…	…	…	94	38	0.938 ± 0.009
Texas	85	51	0.981 ± 0.003	103	56	0.968 ± 0.006	60	34	0.969 ± 0.006
Total	805	251	0.983 ± 0.001	922	226	0.964 ± 0.002	555	170	0.963 ± 0.003

The African-American subpopulations display the most diversity, with individual values falling in a narrow range between 0.974 and 0.993, and an overall value of 0.983. For all subpopulations together, diversity is very high, with a value of 0.998.

The numbers of SSO types that are present in only a single population within each major group are shown in the first line of Table 3. This value includes the number of types that are found multiple times within a single population (population specific types) and the number of types that are observed only once in the group database (unique types). In the remainder of Table 3, the distribution of type sharing among two or more populations (public types) is shown; very few types are shared by most or all of the subpopulations within each large ethnic group. Figure 1 shows the frequency of unique, population specific, and public types for each of the three databases. Unique types are very common, accounting for at least 57% of the types in any population, while shared types are relatively common. A high number of population specific types in any single population was not observed except within Pennsylvania Hispanics. There are six population specific types in this group, occurring from two to seven times each. Within all three databases, no other subpopulations possess more than three population specific types, and no population specific type occurs in more than 10% of any subpopulation.

Figure 2 shows the distributions of SSO types within each large ethnic group. For example, in the African-American database, 150 types are observed only once (unique). Thirty-six types occur twice, and 13 types occur three times. The most common African-American type (12112021) occurs 55 times in this database (6.8%). In European-Americans, the most common type (11111111) occurs 142 times, or in approximately 15% of individuals. The frequency of this type, which is identical to the published reference sequence (9), ranges from approximately 8 to 27% in all of the European-American subpopulations. The next most common type (21112111) occurs in 50 individuals (5.4%). In Hispanics, the most common type (12122011) occurs in 65 individuals (11.8%), while another type occurs in 54 individuals (12112110; 9.7%). Table 4 shows those SSO types that occur in more than 10% of any single subpopulation in each of the three databases. All these types are also found in most, if not all, of the other subpopulations in their respective databases, an indication that there is no heterogeneity for these common types. These common types do however differ among these large groups, such that there is no overlap among them.

TABLE 3—*MtDNA SSO type sharing among populations: For example, in the African-American sample, 42 different types are shared by two populations.*

Number of Populations Sharing	Number of SSO Types		
	African-American	European-American	Hispanic
1*	159	144	120
2	42	28	30
3	11	13	6
4	10	7	4
5	11	8	5
6	7	6	2
7	4	8	3
8	1	3	
9	4	4	
10	2	4	
11		1	

* Includes unique types and population specific types.

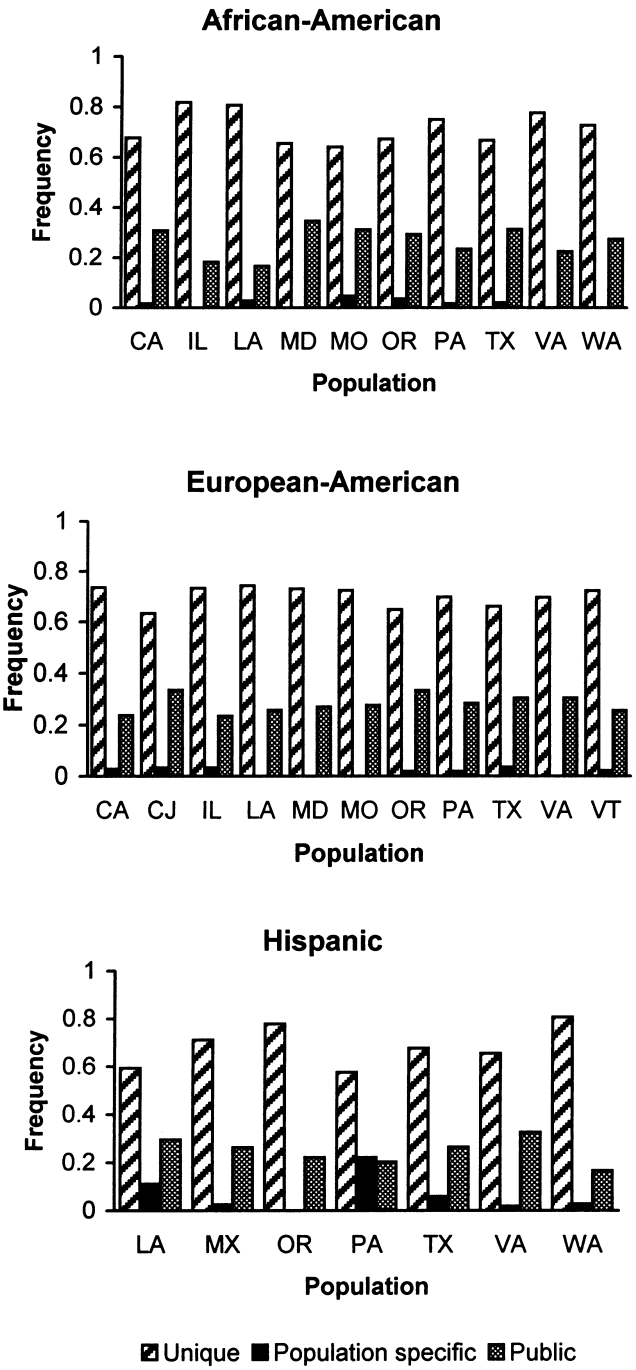

FIG. 1—*Frequency distribution of unique types, population specific types, and shared types among populations.*

Table 5 shows three matrices of within-ethnic group interpopulation Φ_{ST} genetic distances and the associated *p*-values (probability of obtaining a Φ_{ST} value equal to or greater than the observed value, based on 1000 random permutations of the data) for all subpopulation pairwise comparisons. To assess the statistical significance of the *p*-values, they must be adjusted to compensate for the multiple comparisons; since the multiple comparisons consist of pairwise values and hence are not independent, it is not obvious how this should be done (10). As a conservative procedure, we consider *p*-values less than 0.01 to be statistically significant. Within the three databases, although there is a total of 121 different pair-

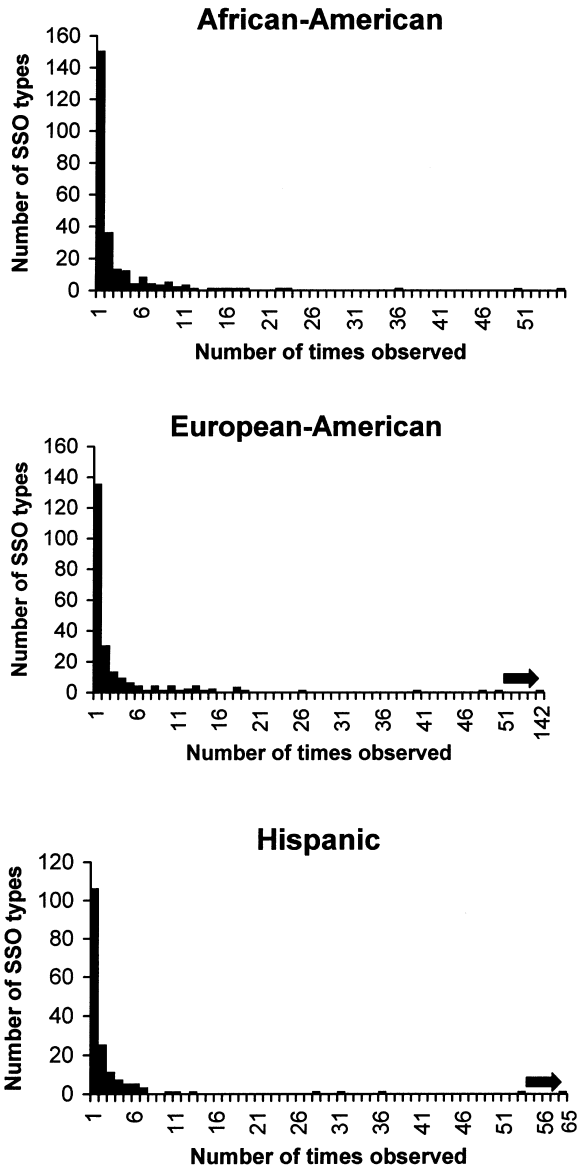

FIG. 2—*Distribution of mtDNA SSO types.*

wise comparisons, only 13 of these comparisons indicate any population pairs are significantly different from each other. For example, within the African-Americans, only the Pennsylvania and California subpopulations are significantly different. Within the European-Americans, individuals from Vermont appear to stand out with respect to four other state populations (IL, MO, OR, TX), but there is no geographic pattern to suggest this result represents any regional heterogeneity. Individuals from Texas and Missouri also appear to be statistically different. Within the Hispanics, Pennsylvanians are significantly different from all six other populations. This is attributable to a higher frequency of several population specific types in this group and perhaps to several high frequency shared types (Fig. 1, Table 4). Mexican Hispanics are statistically different from Washington Hispanics. When comparisons between all 28 subpopulations are made across major group boundaries, however, African-American subpopulations, European-American subpopulations, and Hispanic subpopulations are statistically different with respect to each other in every pairwise comparison (257 comparisons, $p < 0.001$, data not shown).

Table 6 shows the extent of heterogeneity in the three ethnic databases that were evaluated separately and together in an analysis of molecular variance (AMOVA). Within all three databases, the variance within subpopulations accounts for the vast majority of the total variance. At least 98.5% of the observed total variance, regardless of the database, is accounted for by the variance within the subpopulations themselves. The σ_c^2 variance components (within subpopulations) are 99.59, 99.49, and 98.5% in the African-American database, European-American database, and Hispanic database respectively. Hispanics show the most substructure, with a σ_a^2 value of 0.95%, while European-Americans have a value of -0.10%. (A negative value indicates that some subpopulations in different regional groups may actually be slightly more similar to each other than to subpopulations within their own regional group.) None of these values, nor any of the other tests for heterogeneity among subpopulations within the ethnic groups, are statistically significant (p always > 0.01). Within-region variances (σ_b^2) are small and approximately equal for all the ethnic groups, ranging from 0.30% to 0.55%. This indicates that the amount of variance accounted for by differences among subpopulations within their own geographic regions (Central, Northeast, Northwest, Southeast, Southwest) of North America is also minimal. An AMOVA on the entire data set indicates that statistically significant substructuring or heterogeneity is present along ethnic group lines ($p < 0.001$), with a variance among groups of 18.60%. The variance within populations accounts for 80.86% of the total variance, while the variance among subpopulations within each ethnic group is only 0.54% and is not statistically significant ($p > 0.01$), confirming the lack of within ethnic group substructuring.

The historical or geographic boundaries between Cajuns and other European-American groups, and between Mexicans and other Hispanic groups of North America, suggests that both of these groups might be significantly different from the other subpopulations in their respective ethnic groups. However, the genetic distance matrices (Table 5) do not indicate that either group is remarkably different, at least with respect to mtDNA SSO-types. An AMOVA testing this impression was done by first calculating variance components without designating any regional substructure in the European-American and Hispanic databases, and then removing the Cajuns and Mexicans from their respective databases and re-running the analyses. The variance among subpopulations either did not change or changed very little in each test, with σ_a^2 staying the same for the Cajun/European-American calculation (0.49%) and actually rising slightly for the Mexican/Hispanic test (1.39 to 1.50%). Therefore, neither Cajuns nor Mexicans differ significantly from the other subpopulations in their respective large groups.

Discussion

This study surveys baseline mtDNA variation in regional African-American populations, European-American populations, and Hispanics resident in North America. Overall there is high diversity within populations and no significant heterogeneity among populations. While this is not an exhaustive study of all variation in the control region, the results are meaningful when compared with our previous studies of African, Asian, and European populations (1–3), and with other studies of mtDNA variation in North America (6). SSO typing detects a subset of variation in the control region that is typically chosen for DNA sequencing. Comparisons of DNA sequences with their inferred SSO types indicate that SSO typing underestimates the variation present in the entire control re-

TABLE 4—*SSO types occurring in greater than 10% of any subpopulation in the three large databases. States are labeled using conventional postal abbreviations; CJ-Cajun, MX-Mexican.*

African-American Populations										
SSO Type	CA	IL	LA	MD	MO	OR	PA	TX	VA	WA
12112001	0.10	0.04	0.13	0.03	0.07	0.05	0.02	0.05	0.05	0.08
12112021	0.07	0.01	0.09	0.08	0.10	0.02	0.04	0.07	0.09	0.12

European-American Populations											
SSO Type	CA	IL	LA	MD	MO	OR	PA	TX	VA	VT	CJ
11111111	0.14	0.14	0.21	0.18	0.12	0.11	0.17	0.15	0.08	0.27	0.17

Hispanic Origin Populations							
SSO Type	LA	OR	PA	TX	VA	WA	MX
12122111	0.07	0.18	0.02	0.05	0.03	0.0	0.04
12122211	0.0	0.03	0.14	0.05	0.08	0.0	0.03
13112111	0.16	0.08	0.01	0.10	0.03	0.08	0.09
12112110	0.05	0.05	0.20	0.05	0.10	0.02	0.11
12122011	0.21	0.03	0.10	0.10	0.12	0.08	0.18

TABLE 5—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.*

African-American										
	CA	IL	LA	MD	MO	OR	PA	TX	VA	WA
CA	…	0.049	0.325	0.156	0.136	0.020	0.003	0.058	0.225	0.059
IL	0.010	…	0.236	0.667	0.087	0.491	0.977	0.250	0.600	0.240
LA	0.001	0.003	…	0.097	0.810	0.309	0.217	0.163	0.602	0.367
MD	0.008	−0.005	0.013	…	0.111	0.181	0.262	0.328	0.378	0.270
MO	0.004	0.008	−0.005	0.011	…	0.308	0.074	0.608	0.124	0.450
OR	0.010	−0.001	0.003	0.008	0.001	…	0.369	0.557	0.094	0.223
PA	0.015	−0.008	0.004	0.003	0.009	0.001	…	0.073	0.710	0.286
TX	0.009	0.003	0.006	0.003	−0.002	−0.001	0.008	…	0.060	0.269
VA	0.003	−0.002	−0.002	0.001	0.005	0.007	−0.003	0.010	…	0.375
WA	0.011	0.005	0.002	0.005	0.000	0.005	0.002	0.005	0.002	…

European-American											
	CA	IL	LA	MD	MO	OR	PA	TX	VA	VT	CJ
CA	…	0.408	0.954	0.899	0.211	0.594	0.360	0.109	0.921	0.026	0.513
IL	0.001	…	0.391	0.414	0.156	0.474	0.059	0.101	0.211	0.000	0.218
LA	−0.007	0.001	…	0.987	0.584	0.656	0.734	0.068	0.555	0.638	0.433
MD	−0.008	0.000	−0.015	…	0.585	0.485	0.587	0.089	0.594	0.472	0.573
MO	0.003	0.007	−0.003	−0.003	…	0.825	0.010	0.000	0.022	0.000	0.215
OR	−0.002	−0.001	−0.004	−0.001	−0.004	…	0.207	0.341	0.428	0.000	0.679
PA	0.001	0.012	−0.005	−0.003	0.012	0.003	…	0.030	0.838	0.080	0.123
TX	0.004	0.007	0.009	0.017	0.017	0.001	0.010	…	0.110	0.000	0.304
VA	−0.004	0.006	−0.002	−0.003	0.011	0.000	−0.004	0.006	…	0.010	0.177
VT	0.010	0.039	−0.003	−0.002	0.022	0.021	0.008	0.032	0.015	…	0.000
CJ	0.000	0.005	−0.001	−0.003	0.004	−0.004	0.008	0.002	0.004	0.014	…

Hispanic						
	LA	OR	PA	TX	VA	WA	MX
LA	…	0.531	0.000	0.339	0.303	0.250	0.550
OR	−0.003	…	0.000	0.425	0.370	0.100	0.281
PA	0.026	0.028	…	0.000	0.005	0.000	0.008
TX	0.000	−0.001	0.027	…	0.181	0.028	0.222
VA	0.004	0.001	0.014	0.006	…	0.036	0.087
WA	0.005	0.009	0.047	0.019	0.013	…	0.009
MX	−0.004	0.002	0.026	0.005	0.008	0.025	…

TABLE 6—*AMOVA results for North American populations: variance components and Φ-statistics.*

	Variance Components %*			Φ-Statistics^		
Comparison	σ_a^2	σ_b^2	σ_c^2	Φ_{CT}	Φ_{SC}	Φ_{ST}
African-American	−0.10	0.51	99.59	−0.001	0.005	0.004
European-American	0.21	0.30	99.49	0.002	0.003	0.005
Hispanic	0.95	0.55	98.50	0.010	0.006	0.015
All 3 ethnic groups	18.60	0.54	80.86	0.186	0.007	0.191

* σ_a^2 = variance among regions, where regions are the Central, Northeast, Northwest, Southeast, and Southwest, as defined in Table 2, σ_b^2 = variance among populations/within region, and σ_c^2 = variance within populations.

^ Φ_{CT}-correlation of random SSO types within a regional group of populations, relative to that of random pairs of types drawn from the entire data set, Φ_{SC}-correlation between random pairs of SSO types within populations, relative to that of random pairs of SSO types drawn from the region, Φ_{ST}-correlation of random SSO types within populations, relative to that of random pairs of types drawn from the entire data set.

TABLE 7—*The most common SSO types and their frequencies within the databases in this study and comparison to their frequencies in previously published data. The bolded values show the categories where self-identified individuals "cross over" into a new major population group.*

	Frequency, This Study			Frequency, Previous Studies (1–3)		
Most Common Types, This Study	African-American n = 805	European-American n = 922	Hispanic n = 555	African n = 381	European n = 595	Asian n = 993
African-American						
12112021	6.83%	**1.63%**	**0.90%**	4.99%	**0.84%**	**0.20%**
12112001	6.21%	**0.65%**	**0.90%**	14.96%	**0.67%**	**0.30%**
12012021	4.47%	**0.22%**	**0.54%**	2.62%	**0.17%**	**0%**
European-American						
11111111	**1.37%**	15.40%	**1.98%**	**0%**	22.02%	**0.70%**
21112111	**0.12%**	5.42%	**0.18%**	**0%**	3.19%	**0.10%**
11112111	**0.25%**	5.21%	**0%**	**0%**	6.39%	**3.73%**
Hispanic						
12122011	**1.99%**	**0.65%**	11.71%	**0%**	**0%**	1.71%
12112110	**0.25%**	**0.22%**	9.73%	**0%**	**0.37%**	1.01%
13112111	**0.75%**	**0.11%**	6.67%	**0%**	**0%**	4.93%

gion (3,5). Therefore diversity, a highly desirable attribute of a forensic marker, while high for SSO typing, will be even higher at the DNA sequence level for the populations sampled here than the values in Table 2 would suggest.

Heterogeneity is problematic only when particular subpopulations have a number of high frequency, population specific types, or extremely high frequencies of shared types relative to each other. In this study, only Hispanics from Pennsylvania appear to be consistently different from all other populations in their large group with respect to their SSO types. Without sequencing we do not know whether these individuals have identical control regions associated with these SSO types or whether there is additional variation at the sequence level that further discriminates among them. The net effect of additional diverse substitutions at the sequence level would be to break down the population specific types into more clusters of sequences that share most but not all sites. If there is additional variation at the sequence level, then the heterogeneity should diminish, because the population-specific types will decrease in frequency. Because only Pennsylvanians were sampled from the Northeast we are unable to determine if this is a state-specific or a regional phenomenon; additional data might inform us as to whether this is a founder-type effect or a sampling artifact. However, the six population specific types observed in this sample bear no similarity to each other, suggesting that a sampling effect may be the explanation (especially as all but one of these occur only 2–3 times).

Diversity estimates and measures of heterogeneity indicate that mtDNA should be a satisfactory forensic tool for all the North American populations sampled here. The previous study of European or European-derived subpopulations suggested that there is no measurable heterogeneity in any groups sampled from either North America or Europe (2). These new data extend that conclusion by showing that substantial portions of North America contain very homogeneous European-American subpopulations. At both an individual population level and a regional level, there are no high frequency population specific types whatsoever. While it is not desirable to have a widespread common type such as the "reference" type that is often observed in European-Americans, this type is fortunately neither regional nor population specific. Additional control region sequence variation is known to occur in samples with this SSO type (2), so the diversity is greater than SSO typing would detect.

In contrast to an earlier study showing substantial heterogeneity in continental African populations (3), African-Americans are quite homogeneous. This suggests that these individuals are a subset of those populations sampled in the previous study; such a conclusion would be consistent with studies of the origins of North American African populations (11). Table 7 lists the three most frequent SSO types observed in the African-American samples in this study and shows that they were found at similar frequencies in previously published data on continental African populations. In those data, each of these three types was prevalent in the Sierra Leone and Mandenka populations from Western Africa, but was not observed in other African populations at notable frequencies (3). It appears, therefore, that many African-Americans have at least strong maternal ties with coastal West African populations. The

three West African populations in the previous study (Sierra Leone, Mandenka, and Yoruban) were also not significantly different from one another (3).

Hispanics in this study do not display any regional heterogeneity which could have been anticipated, based both on their diverse origins in South America, Central America, and the Caribbean as well as their sizeable migrations to either the Southeast U.S. or Southwest U.S. Given the history of admixture of indigenous American populations with European-derived and African-derived populations, it is also somewhat surprising that there are no European/Hispanic or African/Hispanic subpopulation comparisons suggesting similarity among these groups, although such admixture might be primarily male-mediated, which would not be reflected in this data. We speculate therefore that the majority of mtDNA types in Hispanics are derived from Asian ancestral types, and that these types might also be shared substantially with Native Americans. This is supported by the fact that the three most common and widely distributed Hispanic types in this data set are also three of the higher frequency types reported for Asians previously (Table 7). However, without sequencing the control region we are unable to determine whether these sequences are closely related or identical which would suggest identity by descent from common Asian ancestors.

Table 7 also demonstrates an interesting phenomenon that may relate to the self-identification of those included in this study. Whereas the frequency of each most-common type in this study in its own database has comparable frequencies in its respective continental group of origin (African-American with African, European-American with European, and Hispanic with Asian), there are often either zero or extremely low frequencies of these types in the other two continental databases from the original three studies (1–3). For example, while the most common European type ("reference") in this study was observed at 15.40 and 22.02% in the North American and European databases respectively, it was virtually absent in the continental African and Asian databases. However, when looking at the new study data, the common-type frequencies have often risen in the other two large groups. Using the European "reference" type as an example, we see that its frequency has now risen in the African-American and Hispanic origin databases to 1.37 and 1.98% respectively. It is likely that the commingling within North America of diverse African-derived, European-derived, Hispanic, Asian, and Native American populations in recent history has led to changes in the way individuals self-identify, probably brought about by the blending of families across ethnic boundaries. For example, a number of people sampled (self-identified) as African-American and Hispanic actually have the most common European type ($n = 11$ in each case). This is consistent across almost all categories with only a few exceptions. With this data, we may be seeing the beginning of homogenization of continental mtDNA types in North America. Again, the caveat applies that identical SSO types do not guarantee identical sequences, and identical sequences do not imply identity by descent. However, this intriguing phenomenon will ultimately be quantifiable using mtDNA sequence data as well. Meanwhile, we can roughly quantitate the amount of mtDNA gene flow from European-Americans or Europeans to African/Hispanic individuals, as follows: Assuming a frequency of the reference sequence of 1.4% in African-Americans, 2.0% in Hispanics, and 15% in European-Americans (Table 7), then an estimate of the proportion of European mtDNA in African-Americans is (0.014/0.15) = 9.3%, while in Hispanics it is (0.02/0.15) = 13.3%. The value for African-Americans is comparable to previous estimates by other investigators (12).

A recent comprehensive analysis of 1393 human mtDNA sequences of forensic significance in North America indicated that high diversity is present for each of eight subpopulations at the sequence level ($h > 0.99$, (6)). These estimates of diversity at the sequence level suggest that there are few population specific or high frequency sequence types. In that study, DNA sequence data were also converted to SSO types for ease of comparing populations at individual SSO variant regions. Tests of these data indicated that the primary differences among individuals were present along ethnic lines, with few significant differences among populations within ethnic groups (13 out of 96 tests). The present study augments that data by adding another 2282 individuals in regional subpopulations of forensic interest from around North America. In both studies, diversity is high, heterogeneity is low to nonexistent, and there is therefore no indication for a need for population-specific databases of mtDNA types. With the accumulation of these data, forensic practitioners may feel confident that mitochondrial DNA continues to prove itself a useful forensic typing locus in North American populations (4).

Acknowledgments

We thank Elizabeth Benzinger, Bruce Budowle, Eric Buel, Thomas Grant, George Herrin, Barbara Llewellyn, Teresa Long, Don MacLaren, Chris Tomsey, and Cecilia von Beroldingen for providing samples for this study.

References

1. Melton T, Stoneking M. Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. J Forensic Sci 1996;41(4):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(3): 437–46.
3. 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):582–92.
4. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis—validation and use for forensic casework. Forensic Sci Rev 1999;11:21–50.
5. 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 Hum Genet 1991;48:370–82.
6. Budowle B, Wilson M, DiZinno J, Stauffer C, Fasano M, Holland M, et al. Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 1999;103:23–35.
7. Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, et al. Polynesian genetic affinities with southeast Asian populations as identified by mtDNA analysis. Am J Hu Genet 1995;57:403–14.
8. Schneider S, Kueffer J-M, Roessli D, Excofier L. Arlequin: a software package for population genetic data analysis. Ver 1.1.Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, 1997.
9. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290:457–65.
10. Weir BS. Genetic data analysis. Sunderland, Mass: Sinauer Associates, 1996.
11. Curtin PD. The Atlantic slave trade: a census. University of Wisconsin Press, Madison, 1969.
12. Parra ES, Marcini A, Akey J, Martinson J, Batzer M, Cooper R, et al. Estimating African American admixture proportions by use of population-specific alleles. Am J Hu Genet 1998;63:1839–51.

Addition information and reprint requests:
Terry Melton, Ph.D.
Mitotyping Technologies, LLC
1981 Pine Hall Drive
State College, PA, 16801
Phone: (814) 861-0676
Fax: (814) 861-0576
Email: twm107@mitotyping.com.