J Forensic Sci, May 2006, Vol. 51, No. 3
doi:10.1111/j.1556-4029.2006.00136.x
Available online at: www.blackwell-synergy.com

Marc W. Allard,[1] *Ph.D.; Deborah Polanskey,*[2] *B.S.; Mark R. Wilson,*[3] *Ph.D.; Keith L. Monson,*[4] *Ph.D.; and Bruce Budowle,*[5] *Ph.D.*

Evaluation of Variation in Control Region Sequences for Hispanic Individuals in the SWGDAM mtDNA Data Set

ABSTRACT: The Scientific Working Group on DNA Analysis Methods (SWGDAM) Hispanic data set was analyzed to determine the diversity, phylogeny, and relevant single nucleotide polymorphisms (SNPs) that describe haplogroup patterns for Hispanic Americans ($N = 686$), and to assess the degree of admixture regarding mitochondrial DNA (mtDNA). The largest component of admixture based on mtDNA analysis derives from the four major haplogroups previously observed in Native American ancestry, including A (29.3%), B (15.7%), C (20.6%), and D (4.8%). European (17.8%) and African (11.8%) haplogroups also were observed within this data set. Hispanic SWGDAM samples from the southwest, compared with other SWGDAM Hispanic samples, were observed to have a greater percent of Native American haplogroups present (79.9%), and fewer African American haplogroups (4.5%). A total of 234 SNPs were observed in the data set, including 36 newly reported variable positions. These SWGDAM Hispanic data set SNPs ranged from having 1 to 31 changes (Length = *L*) on the phylogenetic tree, with site 16519 being the most variable. On average, there were 3.9 character changes for each variable position on the tree. The most variable sites (with 13 or more changes each listed from fastest to slowest) observed were 16519 ($L = 31$), 16189 ($L = 23$), 152 ($L = 23$), 16311 ($L = 19$), 146 ($L = 17$), 195 ($L = 17$), 16093 ($L = 15$), 16362 ($L = 14$), 16129 ($L = 13$), 150 ($L = 13$), and 153 ($L = 13$). These sites are consistent with other reports on highly variable positions. A total of 27 SNPs were chosen to identify all clusters containing 1% ($N = 7$) or more individuals in the SWGDAM Hispanic data set. The descriptive analyses revealed that the SWGDAM Hispanic data set is similar to published Native American and Hispanic data sets.

KEYWORDS: forensic science, mitochondrial DNA, SWGDAM forensic mtDNA data set, haplogroup, control region, SNP, Hispanic

The Hispanic population is a geopolitical group defined as descendants from Latin America or other Spanish cultures (1). These cultures are genetically heterogeneous population groups generally comprised of a tri-hybrid structure of European, Native American, and African contributions. Western Hispanics (e.g., Mexicans) are comprised predominantly of European and Native American ancestry and to a lesser extent African; while Eastern Hispanics (e.g., Puerto Ricans and Cubans) are comprised predominantly of European and African ancestry and to a lesser extent Native Americans (1–7). In U.S. Hispanics the largest contribution is from European genes. This ethno-history is supported by analysis with polymorphic autosomal DNA markers. However, there are differences between nuclear autosomal and mitochondrial DNA (mtDNA) markers that can impact the estimate of component contributions in Hispanics. The mtDNA is inherited uniparentally through the maternal lineage. Therefore, because of directional mating, one would expect the tri-hybrid admixture structure to be different based on mtDNA than that estimated using nuclear autosomal markers (6).

Native American mtDNA is defined by four major haplogroups A, B, C, and D, and including the X haplogroup as well (8–43). These genetic subdivisions have been characterized by RFLP analysis (6,9–31) and sequences of the mtDNA control region (CR) (8,32–42). Several studies have suggested that the largest of the maternal admixture components in Hispanics is from Native American ancestry (6,8,28). Thus, any study of Hispanic populations might expect to observe the four major haplogroups of Native Americans and to a lesser extent European and African mtDNA haplogroups. The Scientific Working Group on DNA Analysis Methods (SWGDAM) mtDNA database typically consists of sequence data on the hypervariable regions 1 (HV1) and 2 (HV2) of the CR of the human mtDNA genome (43–47). One of the population groups represented in the data set is Hispanics ($N = 686$). The mtDNA genetic variation in the SWGDAM Hispanic sample is described herein. The data support that the variation observed in the SWGDAM data set is comparable with other studies on Hispanics and that the major component of admixture based on mtDNA analysis derives from Native American ancestry.

Methods

Subjects

Approximately one third of the self-identified Hispanic individuals in the SWGDAM data set ($N = 686$) come from the southwestern ($N = 245$, Texas or New Mexico) United States, with the remainder coming from Connecticut ($N = 148$), Illinois ($N = 5$), or from an unknown location ($N = 288$).

Data Availability

All sequence designations are based on comparisons with the revised Cambridge Reference Sequence (rCRS) (48,49). The

[1]Department of Biological Sciences, George Washington University, Washington, DC 20052.
[2]Federal Bureau of Investigation, DNA Unit 2, Quantico, VA 22135.
[3]Federal Bureau of Investigation, Chem-Bio Sciences Unit, FBI Laboratory, Quantico, VA 22135.
[4]Federal Bureau of Investigation Academy, Counterterrorism and Forensic Science Research Unit, FBI Laboratory, Quantico, VA 22135.
[5]Federal Bureau of Investigation, Laboratory Division, Quantico, VA 22135.

Received 22 Jan. 2005; and in revised form 18 July 2005 and 4 Aug. 2005; accepted 31 Dec. 2005; published 21 April 2006.

mtDNA sequences and available descriptive data can be found at CODIS version 1.2 and in *Forensic Science Communications* (http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm) (44) with corrected sequences as described by Budowle et al. (50) and Polanskey and Budowle (51).

Phylogenetic Methods

The Hispanic data set is composed of samples whose complete CR (positions 16024–16569 and 1–576) has been sequenced ($N = 449$), or those in which only regions HV1 (16024–16365) and HV2 (72–340) were sequenced ($N = 237$). Parsimony analysis was conducted using the software packages of Winclada and Nona (52,53). For each analysis, 2000 replicates of the parsimony ratchet were conducted to determine the most parsimonious solution. Recommended search strategies for using the parsimony ratchet program for large data matrices were followed (52). These analyses include one or two trees held per replicate and trees built using unambiguous optimizations. Additionally, characters were treated with equal weighting of data, and with N's representing any possible base. Gap regions (i.e., deletions) were treated either as missing data or as a fifth independent character state, so that these coded characters could be used to define haplogroups. Length variants in the hypervariable C stretch regions were omitted from these analyses (54).

An alignment was built and then the most parsimonious phylogenetic tree was determined and characterized. Tree statistics were calculated for tree length, consistency index, and retention index, as well as for the number of times that each character changed on a tree (character length = L). The important defining characters for each haplogroup were based on a tree topology and the lineages and states present. These characters were the variable nucleotide sequence positions and states, listed in comparison with the rCRS. After independently determining the positions and states that define all major clusters on the tree, the results were compared with reports in the mtDNA literature (8–42). A nomenclature for human mtDNA haplogroups is established; therefore, the SWGDAM data were compared with studies that used similar character based methods (11,12,22,23,45–47) rather than alternative methods and/or nomenclatures (32). The nomenclature strategy of Torroni et al. (55) and Kivisild et al. (56) was followed, where haplogroups are defined as monophyletic groups that are inferred from the shared mutations that define all members (accommodating some reversals). The major patterns of character change were described. In addition, it is noted when there are multiple independent single nucleotide polymorphism (SNP) origins for the affected positions.

Sequence Alignment

Human CR sequences were aligned according to a set of rules developed for the consistent placement of gaps (57,58) as was described previously (45–47,59).

Ranking of SNPs

The SNPs were placed into three categories to prioritize which SNPs were most informative in partitioning the Hispanic SWGDAM data set. The first ranking level was based on whether or not the variable characters were found in two or more individuals. All shared derived sites that define groups of 1% ($n = 7$) or more of the individuals in the Hispanic SWGDAM data set were listed as a second arbitrary ranking. These informative sites, defined by nucleotide position and state, were determined by the placement of characters on the tree topology and thus, all were found on branches that gave rise to seven or more terminal samples. Characters that appeared as the most discriminating were chosen after further examination of the amount of the tree that was affected. The final SNP set was selected to identify highly informative sites partly by removing some of the SNP sites that defined the same group.

The Hispanic SWGDAM data set was characterized by comparing the defining and variable positions and states, as well as the observed haplogroups, to other studies (8–42). Individuals were allocated to the most-derived (smallest) published haplogroup that could be identified.

Results and Discussion

Forensically Informative Population Data Statistics

The majority of the mtDNA sequences were observed only once within the Hispanic population data set and its three main subgroups. Single occurring haplotypes comprised 51.2% (351/686) of the Hispanic sequences with a high of 85.4% (169/198) for southwestern Hispanics and a low of 55.0% (82/149) for Connecticut Hispanics (Table 1). The lower percentage of mtDNA haplotypes observed among the total Hispanic group is most likely due to the larger sample size. Genetic diversity, calculated according to Tajima (60), ranged from 0.980 in the Connecticut sample to 0.997 in southwestern Hispanics (Table 2). Random match probability (RMP), calculated according to Stoneking et al. (61), ranged from 2.79% in the Connecticut Hispanic sample to 0.72% in southwestern Hispanics (Table 2), but the total Hispanic group had a RMP of 0.65%. The most commonly observed haplotypes for each group are displayed in Table 1. The haplotypes listed in Table 1 were observed at a minimum of 1.2% or a count of 3 (which ever was larger) in a data set. To assess whether or not the most common haplotypes are shared between the Hispanic subgroups, the southwestern and Connecticut samples were compared, because the provenance is better defined for these two subgroups of the SWGDAM Hispanic data. As expected, the common haplotypes are different between the two groups, although some are similar. This observation is consistent with their ethnohistory.

Variable SNP Sites

From the phylogenetic analysis, including only informative characters for all of the Hispanic samples and gaps treated as fifth states, 849 trees of length 1130 were uncovered (CI = 0.24, RI = 0.86). Variable sites that were found in two or more individuals in the Hispanic SWGDAM data set (i.e., parsimony informative characters) are listed in Table 3. The SNP sites were ranked into three categories according to the size of the cluster and the discriminating quality of the sites. The first ranking (lightest shade) indicated all variable sites observed in two or more individuals ($N = 234$). The second ranking (darker shaded boxes) delineate SNPs ($N = 75$) that defined groups of seven or more individuals, and the darkest boxes (third ranking, $N = 27$) define a subset of the second ranking that best partitioned the Hispanic haplogroups. This third ranking was partly determined by reducing the redundancy among sites for haplogroups that were defined by multiple SNPs, as well as by choosing common and variable SNPs.

Of the 234 SNPs observed, 36 of the variable positions reported in Table 3 were not observed previously in the MITOMAP database (62). These 36 sites all occurred rarely. Accordingly, they

TABLE 1—*Number of different mitochondrial DNA (mtDNA) haplotypes observed in each database.*

Number of Haplotypes Observed	Number of Times Observed	Total
Total (N = 686)[*]		
1	24	24[a]
1	20	20[b]
1	17	17[c]
1	14	14[d]
1	12	12[e]
2	11	22[f]
1	9	9[g]
1	8	8
2	7	14
3	6	18
4	5	20
9	4	36
15	3	45
38	2	76
351	1	351
431		686
Southwest (N = 244)[†]		
1	10	10[a]
1	4	4[b]
7	3	21[c]
20	2	40
169	1	169
198		244
Connecticut (N = 149)[‡]		
1	18	18[a]
2	7	14[b]
1	5	5[c]
3	4	12[d]
2	3	6[e]
6	2	12
82	1	82
97		149
Unknown (N = 288)[§]		
1	11	11[a]
1	8	8[b]
1	7	7[c]
2	6	12[d]
2	5	10[e]
6	4	24[f]
4	3	12
20	2	40
164	1	164
201		288

[*] The most common haplotypes are as follows: [a] 16223T, 16298C, 16325C, 16327T, 73G, 249-, 290-, 291-, 315.1C; [b] 16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.; [c] 16182C, 16183C, 16189C, 16224C, 16311C, 73G, 146C, 152C, 207A, 263G, 315.1C; [d] 16223T, 16298C, 16325C, 16327T, 73G, 249-, 263G, 290-, 291-, 309.1C, 315.1C; [e]16111T, 16129A, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C.; [f] 16223T, 16325C, 16362C, 73G, 263G, 315.1C; 16083T, 16111T, 16223T, 16256T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 214G, 235G, 263G, 315.1C, 523-, 524-; [g] 16223T, 16290T, 16319A, 16362C, 16519C, 64T, 73G, 146C, 153G, 235G, 263G, 315.1C, 523-, 524-.

[†] The most common haplotypes are as follows: [a] 16182C, 16183C, 16189C, 16224C, 16311C, 16519C, 73G, 146C, 152C, 207A, 263G, 315.1C, 527N, 530N; [b] 16223T, 16298C, 16325C, 16327T, 73G, 249-, 263G, 290-, 291-, 309.1C, 315.1C, 489C, 493G, 523-, 524-; [c] 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C, 523-, 524-; 16223T, 16274A, 16298C, 16325C, 16327T, 16519C, 73G, 249-, 263G, 290-, 291-, 315.1C, 489C, 493G, 523-, 524-; 16223T, 16325C, 16362C, 73G, 263G, 315.1C, 489C; 16069T, 16126C, 16145A, 16222T, 16261T, 73G, 263G, 295T, 315.1C 462T, 489C; 16111T, 16223T, 16290T, 16319A, 16356C, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C, 523-, 524-; 16223T, 16270T, 16290T, 16319A, 16362C, 64T, 73G, 146C, 153G, 235G, 263G, 315.1C, 523-, 524-; 16051G, 16188T, 16204A, 16223T, 16325C, 16327T, 16362C, 16519C, 16527T, 73G, 249-, 263G, 290-, 291-, 315.1C, 489C, 523-, 524-.

[‡] The most common haplotypes are as follows: [a] 16223T, 16298C, 16325C, 16327T, 16519C, 73G, 249-, 290-, 291-, 315.1C, 489C, 493G, 523-, 524-; [b]16083T, 16111T, 16223T, 16256T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 214G, 235G, 263G, 315.1C, 523-, 524-; 16111T, 16129A, 16223T, 16290T, 16319A, 16362C, 64T, 73G, 146C, 153G, 235G, 263G, 315.1C, 523-, 524-; [c] 16223T, 16327T, 90A, 97A, 106-, 111-, 150T, 189G, 200G, 263G, 315.1C; [d] 16069T,16126C, 16145A, 16172C, 16222T, 16261T, 73G, 242T, 263G, 295T, 315.1C, 462T, 489C; 16086C, 16183C, 16189C, 16193.1C, 16223T, 16278T, 16298C, 16325C, 16327T, 73G, 249-, 263G, 290-, 291-, 315.1C, 489C, 493G, 523-, 524-; 16093C,16223T, 16278T, 16362C, 16519C, 73G, 263G, 315.1C, 523-, 524-; [e] 16126C, 16187C, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16519C, 73G, 152C, 182T, 185T, 195C, 228A, 247A, 263G, 315.1C, 357G, 523-, 524-; 16189C, 16192T, 16270T, 16320T, 73G, 150T, 263G, 315.1C.

[§] The most common haplotypes are as follows: [a] 16111T, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 309.1C, 315.1C; [b] 16223T, 16298C, 16325C, 16327T, 73G, 249-, 263G, 290-, 291-, 315.1C; [c] 16182C, 16183C, 16189C, 16224C, 16311C, 73G, 146C, 152C, 207A, 263G, 315.1C; [d] 16223T, 16298C, 16325C, 16327T, 73G, 249-, 290-, 291-, 315.1C; 16111T, 16129A, 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C; [e] 16223T, 16325C, 16362C, 73G, 263G, 315.1C; 16083T, 16111T, 16223T, 16256T, 16290T, 16319A, 16362C, 73G, 146C, 152C, 153G, 214G, 235G, 263G, 315.1C; [f] 16223T, 16298C, 16325C, 16327T, 73G, 198T, 249-, 263G, 290-, 291-, 315.1C; 16111T, 16183C, 16189C, 16217C, 73G, 263G, 315.1C; 16182C, 16183C, 16189C, 16217C, 16319A, 73G, 146C, 263G, 315.1C; 16223T, 16290T, 16319A, 16362C, 73G, 146C, 153G, 235G, 263G, 315.1C.

were ranked in the first category of SNPs (except for position 90), and observation of these may likely be due to sampling.

All SNP sites were examined on a most parsimonious tree topology. The number of times each character state changed on the tree (character length, *L*) was listed for each SNP site in Tables 3 and 4. So that sample size was not a variable in these comparisons, only those Hispanic individuals that were sequenced for the complete CR ($N = 449$) were used in the estimates of character length. For these characters, length values range from 1 to 31 (average 3.9), with site 16519 showing the greatest number of changes on the tree. Site 16519 is the most rapidly changing site in other population groups as well (45–47). SNP sites that changed 13 or more times on the tree based on full CR sequence data were 16519 ($L = 31$), 16189 ($L = 23$), 152 ($L = 23$), 16311 ($L = 19$), 146 ($L = 17$), 195 ($L = 17$), 16093 ($L = 15$), 16362 ($L = 14$), 16129 ($L = 13$), 150 ($L = 13$), and 153 ($L = 13$, Table 4). When individuals without the full CR sequence were excluded from the matrix, 30 SNPs previously informative, but now uninformative, were found and are listed with length = 0 or 1 in Table 3. All of these sites were ranked in the first category.

Hispanic Haplogroup Frequencies

The published Native American and Hispanic data (8–42) and the Hispanic haplogroups herein show similar variation. Approximately 71% of the individuals in the Hispanic data set clustered

TABLE 2—*Hispanic population groups, sample size, genetic diversity (GD), and random match probability.*

Database	*N*	GD[*]	RMP (%)[†]
Total	686	0.995	0.65
Southwest	244	0.997	0.72
Connecticut	149	0.980	2.79
Unknown	288	0.995	0.88

[*] GD is calculated according to Tajima (60).

[†] Random match probability (RMP) is calculated according to Stoneking et al. (61).

TABLE 3—*Single nucleotide polymorphisms (SNPs) determined from phylogenetic analysis of the mtDNA CR sequences of the SWGDAM Hispanic data set.*

CRS		Hispanic	*L*	CRS		Hispanic	*L*	CRS		Hispanic	*L*	CRS		Hispanic	*L*
16042	G		2	16256	C	T	8	16483	G	A	4	234	A		2
16051	A	G	4	16257	C		0	16497	A		2	235	A	G	3
16069	C	T	1	16259	C		1	16519	T	C	31	236	T		2
16075	T		2	16260	C		2	16524	A		2	239	T		2
16083	C	T	1	16261	C	T	4	16526	G		2	241	A		2
16086	T		4	16262.1	D		2	16527	C		5	242	C		1
16092	T		10	16263	T		2	41	C		2	244	A		2
16093	T		15	16264	C	T	5	60.1	D		1	247	G	A	2
16097	T		1	16265	A		1	64	C	T	6	249	A	D	2
16098	A		1	16266	C		2	71	G		4	250	T		0
16104	C		1	16270	C	T	5	72	T		3	257	A		1
16111	C	T	9	16271	T		2	73	A	G	8	263	A	G	2
16114	C		5	16274	G		7	89	T		2	264	C		2
16124	T	C	3	16278	C	T	10	90	G	A	1	275	G		1
16126	T	C	9	16284	A		2	92	G		1	290	A	D	1
16129	G	A	13	16290	C	T	2	93	A		4	291	A	D	2
16136	T		2	16291	C		6	94	G		4	292	T		0
16140	T		4	16292	C		4	95	A		3	295	C	T	3
16142	C		2	16293	A	G	7	97	G	A	3	297	A		1
16145	G	A	7	16294	C	T	6	103	G		2	299	C		1
16148	C		3	16295	C		6	105	C		2	308	C		1
16153	G		5	16296	C		1	106	G	D	4	309	C		2
16163	A		3	16298	T	C	5	107	G	D	4	310	T		4
16168	C		1	16299	A		4	108	A	D	4	315	C		1
16172	T		11	16300	A		1	109	G	D	4	316	G		3
16175	A		1	16301	C		4	110	C	D	4	317	C		1
16176	C		1	16304	T		4	111	A	D	2	325	C		1
16179	C		2	16309	A		3	114	C	G	1	334	T		1
16181	A		1	16311	T	C	19	131	T		2	340	C		4
16182	A	C	3	16316	A		2	143	G		8	357	A	G	1
16183	A	C	7	16318	A		1	146	T	C	17	373	A		2
16184	C		7	16319	G	T	5	150	C	T	13	385	A		4
16186	C		2	16320	C		4	151	C		8	417	G		2
16187	C	T	3	16323	T		1	152	T	C	23	418	C		1
16188	C	T	3	16325	T	C	6	153	A	G	13	437	C		2
16189	T	C	23	16327	C	T	2	159	T		2	456	C		3
16192	C	T	9	16328	C		0	178	A		1	462	C	T	2
16193	C		2	16335	A		2	179	T		1	479	A		2
16204	G		1	16342	T		1	182	C	T	6	487	A		1
16209	T		6	16343	A		2	183	A		2	489	T	C	3
16213	G		4	16344	C		1	185	G	A T	10	493	A	G	4
16217	T	C	3	16352	T		1	186	C		1	497	C		1
16218	C		4	16353	C		1	188	A		0	498	C		2
16219	A		1	16354	C		1	189	A	C G	8	499	G	A	3
16221	C		2	16355	C		6	194	C		3	513	G		3
16222	C	T	3	16356	T		4	195	T	C	17	514	C		2
16223	C	T	9	16357	T		7	196	T		1	518	C		2
16224	T	C	5	16360	C		3	198	C	T	5	523	A	D	7
16230	A		1	16362	T	C	14	199	T		4	524	C	D	7
16234	C		8	16368	T		3	200	A	G	5	524.1	D		5
16235	A		3	16390	G	A	7	203	G		1	524.2	D		5
16239	C		9	16391	G		6	204	T		4	524.3	D		3
16240	A		1	16399	A		4	207	G	A	3	524.4	D		3
16241	A		2	16400	C		2	210	A		1	534	C		2
16243	T		2	16456	G		2	214	A	G	4	551	A		2
16247	A		1	16465	C		2	215	A	G	3	573.1	D		6
16248	C		1	16467	C		2	225	G		2	573.2	D		4
16249	T		5	16468	T		2	226	T		3	573.3	D		4
								228	G		4	573.4	D		3

The numbers refer to the revised rCRS nomenclature system for sites. Light bars refer to the presence of an SNP in the data set (variable site found in two or more individuals). Medium gray and black bars are SNPs that defined groups with seven or more individuals. Black bars refer to the most informative SNPs based on phylogenetic analysis and a close examination of the evolution of the character data on a tree. This entails removal of some of the SNPs that defined the same groups. Character states were listed both for the rCRS and for the more common variable sites (medium gray and black bars). Nucleotides that were observed as defining characters were listed. When more than one character was listed, this referred to multiple state changes at a site. Length (*L*) of characters was determined by counting the numbers of character changes occurring on a most parsimonious tree for the SWGDAM Hispanic data set, with greater length scores indicating multiple independent gains and/or reversals.

rCRS, Cambridge reference sequence; CR, control region; mtDNA, mitochondrial DNA; SWGDAM, Scientific Working Group on DNA Analysis Methods.

in the A, B, C, or D haplogroups common to Native American and Hispanic populations. The informative SNPs used to identify the four main haplogroups are listed in Table 5 and are consistent with other studies (32–42). The remaining 29% of individuals in the data set were found to have either recognized European or African haplogroups.

TABLE 4—*Single nucleotide polymorphisms (SNPs) determined from phylogenetic analysis of the SWGDAM Hispanic mtDNA CR sequences.*

L	CRS	*L*	CRS	*L*	CRS	*L*	CRS	*L*	CRS	*L*	CRS
31	16519	6	16295	4	107	3	499	2	111	1	16240
23	16189	6	16325	4	108	3	513	2	131	1	16247
23	152	6	16355	4	109	3	524.3	2	159	1	16248
19	16311	6	16391	4	110	3	524.4	2	183	1	16259
17	146	6	64	4	199	3	573.4	2	225	1	16265
17	195	6	182	4	204	2	16042	2	234	1	16296
15	16093	6	573.1	4	214	2	16075	2	236	1	16300
14	16362	5	16114	4	228	2	16136	2	239	1	16318
13	16129	5	16153	4	310	2	16142	2	241	1	16323
13	150	5	16224	4	340	2	16179	2	244	1	16342
13	153	5	16249	4	385	2	16186	2	247	1	16344
11	16172	5	16264	4	493	2	16193	2	249	1	16352
10	16092	5	16270	4	573.2	2	16221	2	263	1	16353
10	16278	5	16298	4	573.3	2	16241	2	264	1	16354
10	185	5	16319	3	16124	2	16243	2	291	1	60.1
9	16111	5	16527	3	16148	2	16260	2	309	1	90
9	16126	5	198	3	16163	2	16262.1	2	373	1	92
9	16192	5	200	3	16182	2	16263	2	417	1	114
9	16223	5	524.1	3	16187	2	16266	2	437	1	178
9	16239	5	524.2	3	16188	2	16271	2	462	1	179
8	16234	4	16051	3	16217	2	16284	2	479	1	186
8	16256	4	16086	3	16222	2	16290	2	498	1	196
8	73	4	16140	3	16235	2	16316	2	514	1	203
8	143	4	16213	3	16309	2	16327	2	518	1	210
8	151	4	16218	3	16360	2	16335	2	534	1	242
8	189	4	16261	3	16368	2	16343	2	551	1	257
7	16145	4	16292	3	72	2	16400	1	16069	1	275
7	16183	4	16299	3	95	2	16456	1	16083	1	290
7	16184	4	16301	3	97	2	16465	1	16097	1	297
7	16274	4	16304	3	194	2	16467	1	16098	1	299
7	16293	4	16320	3	207	2	16468	1	16104	1	308
7	16357	4	16356	3	215	2	16497	1	16168	1	315
7	16390	4	16399	3	226	2	16524	1	16175	1	317
7	523	4	16483	3	235	2	16526	1	16176	1	325
7	524	4	71	3	295	2	41	1	16181	1	334
6	16209	4	93	3	316	2	89	1	16204	1	357
6	16291	4	94	3	456	2	103	1	16219	1	418
6	16294	4	106	3	489	2	105	1	16230	1	487
										1	497

The numbers refer to the revised rCRS nomenclature system for sites. Length (*L*) of characters was determined by counting the numbers of character changes occurring on a most parsimonious tree for the data set. Characters were ordered from the fastest changing sites on the tree to the slowest, with an average of 3.9 changes.

rCRS, Cambridge reference sequence; CR, control region; mtDNA, mitochondrial DNA; SWGDAM, Scientific Working Group on DNA Analysis Methods.

A Haplogroup

The A haplogroup was seen in 29.3% of the individuals in the Hispanic data set. This haplogroup was defined by character states 16223T, 16290T, 16319A, 16362C, and 235G. All of these sites were reported by Yao et al. (63) along with site and state 152C. Site 152 was variable within the SWGDAM Hispanic data set. Site 16362 showed higher variability in the Yao et al. (63) data set than in the SWGDAM Hispanic data set, with only two small subclusters (and one other individual) showing reversals. Other studies (8,11,12,23,29,31) have reported the same HVI variants. Nearly all of the A haplogroup individuals in the Hispanic data set were further characterized by 146C and 153G. Sites that further subdivided this haplogroup include character states 16083T, 16092C, 16111T, 16129A, 16239T, 16256T, 16311C, 16519C, 64T, and 214G.

The A haplogroup could be divided into several subclusters. The addition of SNPs 16083T, 16111T, 16256T, 152C, and 214G ($N = 18$, 2.6%) define the major subcluster, while another subcluster is defined by the addition of SNPs 16111T and 16129A ($N = 8$, 1.2%). Forster et al. (26) defined haplogroup A2 based on the additional presence of site 16111T, although in the present analysis this site had numerous reversals within the A haplogroup. These two subclusters are similar to those described in Eskimos based on variation in HV1, although these groups were not formally named (36). The above subclusters could be newly named A2a and A2b, respectively.

TABLE 5—*List of some of the most important positions and nucleotides that identify major haplogroups in the SWGDAM Hispanic data set.*

Haplogroup	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7
A	16223T	16290T	16319A	16362C	235G		
B	16189C	16217C	499A				
C	16223T	16298C	16327T	249D	290D	291D	489C
D	16223T	16362C	489C				

SNP, single nucleotide polymorphism; SWGDAM, Scientific Working Group on DNA Analysis Methods.

B Haplogroup

The B haplogroup was distinguished by SNPs 16183C, 16189C, and 16217C. It was present in 15.7% of the SWGDAM data set

($N = 108$). The SNPs 16183C, 16189C, and 16217C were defined by Easton et al. (22) as haplogroup B1. Only two of 47 individuals from the Yao et al. (63) data set did not exhibit SNP 16183C, though this site was not included for defining the B haplogroup. Similarly, this SNP is not included here as an identifier for the haplogroup due to the absence of SNP 16183C in other studies (12,23). In contrast, all but one of the Hispanic individuals in the SWGDAM data set that were sequenced for the complete CR had SNP 499A, thus this site was included as an identifier for the haplogroup (Table 5). Because of the presence of 16217C, several studies (54,59) defined this haplogroup as B4. Other sites that further subdivide haplogroup B include 16182C, 16111T, 16298C, 16483A, 114G, 146C, and 152C. The B4a haplogroup was observed rarely ($N = 2$), and the B4b haplogroup was not observed in the SWGDAM Hispanic data set; these were defined as subgroups of B by the presence of the 16261T and 16136C polymorphisms (63).

Several additional distinct subclusters within the B haplogroup were observed in the Hispanic SWGDAM data set. Their presence likely is due to sampling. One subcluster was defined by the presence of SNP 114G ($N = 13$, 1.9%). This group could be treated as an additional subcluster of the B haplogroup and may be represented as B4g. Several other named subclusters (B5a and B5b haplogroups, defined by site 16140 with either of the variable SNPs 16266 or 16243, respectively) were not observed in the SWGDAM Hispanic data set.

C Haplogroup

The C haplogroup cluster was observed in 20.6% of the Hispanic data set ($N = 141$) and was defined by character states 16223T, 16298C, 16327T, 489C, and deletions at positions 249, 290, and 291. This pattern was fully described by Yao et al. (63). However, these earlier studies did not include variant 489C or the deletions at positions 290 and 291, as observed in the SWGDAM Hispanic data, most likely because they were not analyzed. Similarly, Torroni et al. (10–12), Santos et al. (23), and Easton et al. (22) described the same HV1 variants referred to as the C2 haplogroup. Moraga et al. (37) reported all of the above deletions in South American aboriginal populations, although they did not use the same rules of alignment (57,58) for placing these deletions. Additional SNPs that further partitioned this haplogroup in the Hispanic SWGDAM data set included 16051G, 16153A, 16188T, 16298C, 16311C, 16362C, 16519C, 215G, 263G, 493G, and deletions at positions 523 and 524. Variation at site 16325C has also been reported for some populations (23,37). This site rarely changed in the SWGDAM Hispanic data set and thus could have been used to define the C2 haplogroup (also see (22,24)).

Variation within the C haplogroup has been partitioned into clusters C1 and C2 of Easton et al. (22) based on digestion with HaeIII that corresponds to variation at SNP 16519C. In the SWGDAM data set, SNP 16519C is variable and arises several times independently. Based on SNP nucleotide differences, two additional haplogroups could be identified, one defined by SNP 16051G ($N = 12$, 1.7%) and another subcluster by SNP 493G and deletions at positions 523 and 524 ($N = 65$, 9.5%). These may be regarded as two new C haplogroup subclusters, C1a and C2a, respectively.

D Haplogroup

The D haplogroup cluster was observed in 4.8% of the Hispanic data set ($N = 33$) and was defined by character states 16223T, 16362C, and 489C. Santos et al. (23) include the same HVI sites as defining all Amazonian Native Americans and included 16325C. In the SWGDAM Hispanic data set site 16325C was variable in 10% of the D haplogroup SWGDAM samples ($N = 4$) and was reported to be variable in some North and South American populations (12,14,37), although this variability is not always observed (22,23). The most characteristic site for defining the D haplogroup in southern Chilean and Argentinean Amerindian populations (i.e., 16187T) was only rarely observed in the Hispanic SWGDAM data set (37). Subclusters D1, (26), D2, D4a (16129A, 152C), D4b (16319A), and D5a (16266T) were not observed in the Hispanic data set; and haplogroup G2 was observed only once by defining SNP 16278T (63).

Other Native American Haplogroups

One sample was defined as belonging to the X haplogroup according to the character definition of Kivisild et al. (56).

European Caucasian Haplogroups

Many of the previously defined European haplogroups were observed ($N = 122$, 17.8%) in the SWGDAM Hispanic data set. These include haplogroups H ($N = 45$), I ($N = 3$), J ($N = 20$), K ($N = 21$), T ($N = 10$), U ($N = 18$), and V ($N = 5$). The specific SNPs used to define the European haplogroups are described elsewhere (45,55,64).

African Haplogroups

African haplogroups also were observed ($N = 80$, 11.7%) in the SWGDAM Hispanic data set. These include haplogroups L1 ($N = 22$), L2 ($N = 19$), and L3 ($N = 39$). The specific SNPs used to define the African haplogroups are described elsewhere (46,65).

Admixture Component in Hispanics

Based on autosomal markers, the largest admixture contribution is from European genes (1–7). The mtDNA component differs, due to a uniparental inheritance, and is comprised predominately of Native American haplogroups (6). There are also differences in the percentage of mtDNA haplogroups derived from European and African descent between southeastern and southwestern (SE and SW) Hispanic populations (6). SE Hispanic populations have a larger European and African component and a corresponding decrease in the Native American contributions compared with SW populations (1–7).

The largest component of admixture (70.8%) based on mtDNA analysis for the Hispanic SWGDAM samples derives from the four major haplogroups previously observed in Native American ancestry, including A (29.3%), B (15.7%), C (20.6%), and D (4.8%). European (17.8%) and African (11.8%) haplogroups also were observed within this data set. Merriweather et al. (6) observed 85.1% of Native American mtDNA haplogroups (combined A, B, C, and D) in a southwestern Hispanic data set from Colorado (6). Our Hispanic data set is comprised of SE, SW, and undefined origin samples. Therefore, those from Texas and New Mexico were evaluated separately and compared with Merriweather et al.'s (6) observation as well as the remaining SWGDAM data.

Of the Hispanic samples, 245 were collected in Texas ($N = 222$) and New Mexico ($N = 23$) and these were designated

as southwestern in origin. The remaining Hispanic samples in the SWGDAM data set ($N = 441$) do not have clear SW or SE geolocations of origin. The southwestern sample shows the admixture component as 79.9% of Native American origin inlcuding A (29.8%), B (21.2%), C (22.0%), and D (6.9%). Many recognized European haplogroups were observed ($N = 38$, 15.5%) in the SWGDAM SW Hispanic data set. These include haplogroups H ($N = 17$), J ($N = 5$), K ($N = 12$), T ($N = 2$), and U ($N = 2$). Fewer African American haplogroups were observed in this SW sample with a combined L1, L2, and L3 haplogroup frequency of 4.5% ($N = 11$). The total percentages are comparable with the 85.1% of Native American mtDNA haplogroups, as well as the admixture components of European and African origin observed in Colorado Hispanics (6). The remaining Hispanic samples were examined ($N = 441$) and fewer Native American haplogroups were observed ($N = 288$, 65.3%), more African American haplogroups were observed ($N = 69$, 15.7%) and the European haplogroup frequencies were slightly higher (19%) than those seen in the southwestern populations.

Conclusions

All SNP sites were ranked into three groups based on the level of discrimination they provided in separating Hispanic haplogroups from one another. The variability observed in SNPs was determined from a phylogenetic perspective with site 16519 showing the greatest variability. Many of these positions have been well characterized (45–47,55,56,63–65). Additional character combinations for haplogroup definition were identified beyond those currently published. The new information for determining Hispanic haplogroups will provide a useful baseline for forensic analyses of these populations.

The mitochondrial lineages observed in Hispanics were readily characterized as predominantly recognized Native American, with smaller contributions from European and African haplogroups. Approximately 71% of the individuals in the SWGDAM Hispanic data set carry mtDNA haplogroups recognized to Native Americans (A, B, C, and D). The specific SNPs that defined these major haplogroups correspond well to the published Native American data. The proportion of Native American to European/African mtDNA haplogroups is comparable with other studies. In contrast to nuclear autosomal data where European Caucasian ancestry predominates, the major component for mtDNA is Native American (6,8,9). The overall consistency of the Hispanic SWGDAM data set with other published sequences supports the utility of this data set for forensic applications (44).

Acknowledgments

This research was supported in part by an appointment to the Visiting Scientists Program at the Federal Bureau of Investigation, Counterterrorism Forensic Science Research Unit administered by the Research Participation Program of the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the FBI-CTFSRU. We appreciate the helpful editorial suggestions that were provided to us by Kerri Dugan, John Stewart, and Alice Isenberg. This is publication 04-09 of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

1. Bertoni B, Budowle B, Sans M, Chakraborty R. Admixture in Hispanics: distribution of ancestral population contributions in the continental United States. Hum Biol 2003;75(1):1–11.
2. Cerda-Flores RM, Budowle B, Jin L, Barton SA, Deka R, Chakraborty R. Maximum likelihood estimates of admixture in Northeastern Mexico using 13 short tandem repeat loci. Am J Hum Biol 2002;14:429–39.
3. Chakraborty BM, Fernandez-Esquez ME, Chakraborty R. Is being Hispanic a risk factor for non-insulin dependent diabetes mellitus (NIDDM)? Ethn Dis 1999;9:278–83.
4. Sans M. Admixture studies in Latin America: from the 20th to the 21st century. Hum Biol 2000;72:155–77.
5. Chakraborty R, Ferrell RE, Stern MP, Haffner SM, Hazuda HP, Rosenthal M. Relationship of prevalence of non-insulin dependent diabetes mellitus in Amerindian admixture in Mexican-Americans of San Antonio, Texas. Genet Epidemiol 1986;3:435–54.
6. Merriweather DA, Huston S, Iyengar S, Hamman R, Norris J, Shetterly S, et al. Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. Am J Phys Anthropol 1997;102:153–9.
7. Lisker R, Babinsky V. Admixture estimates in nine Mexican Indian groups and five east coast localities. Rev Invest Clin 1986;38:145–9.
8. Malhi R, Eshleman J, Greenberg J, Weiss D, Shoook B, Kaestle F, et al. The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. Am J Hum Genet 2002;70:905–19.
9. Schurr T, Ballinger S, Gan Y, Hodge J, Merriwether D, Lawrence D, et al. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies suggesting they derived from four primary maternal lineages. Am J Hum Genet 1990;46:613–23.
10. Torroni A, Schurr TG, Yang C, Szatmary E, Williams R, Schanfield M, et al. Native American mitochondrial DNA analysis indicates that the Amerindian and Nadene populations were founded by two independent migrations. Genetics 1992;130:153–62.
11. Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, et al. MtDNA variation of Aboriginal Siberians reveals distinct genetic affinities with Native Ameicans. Am J Hum Genet 1993;53:591–608.
12. Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, et al. Asian Affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 1993;53:563–90.
13. Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K. Peopling of the America founded by four major lineages of mitochondrial DNA. Mol Biol Evol 1993;10:23–47.
14. Ginther C, Corach D, Penacino G, Rey J, Carnese F, Hutz M, et al. Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. Exs 1993;67:211–9.
15. Merriwether D, Rothhammer F, Ferrell R. Genetic variation in the New World: ancient teeth, bone, and tissue as sources of DNA. Experientia 1994;50:592–601.
16. Santos M, Barrantes R. D-loop mtDNA deletion as a unique marker of Chibcha Amerindians. Am J Hum Genet 1994;55:413–4.
17. Santos M, Ward R, Barrantes R. Mitochondrial DNA variation in the Chibcha Amerindian Huetar from Costa Rica. Hum Biol 1994;66: 963–77.
18. Bailliet G, Rothhammer F, Carnese F. Founder mitochondrial haplotypes in Amerindian populations. Am J Hum Genet 1994;54:27–33.
19. Torroni A, Chen Y, Semino O, Santachiara-Beneceretti AS, Scott CR, Lott MT, et al. MtDNA and Y-chromosome polymorphisms in four Native American populations from Southern Mexico. Am J Hum Genet 1994;54:303–18.
20. Bianchi N, Bailliet G, Bravi C. Peopling of the Americas as inferred through the analysis of mtDNA. Braz J Genet 1995;18:661–8.
21. Torroni A, Wallace DC. MtDNA haplogroups in Native Americans. Am J Hum Genet 1995;56:1234–6.
22. Easton RD, Merriwether A, Crews, Ferrell RE. MtDNA variation in the Yanomami: evidence for additional New World founding lineages. Am J Hum Genet 1996;59:213–25.
23. Santos SEB, Ribeiro-Dos-Santos AKC, Meyer D, Zago MA. Multiple founder haplotyopes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. Am J Hum Genet 1996;60:305–19.
24. Merriwether D, Ferrell R. The four founding lineage hypothesis for the New World: a critical reevaluation. Mol Phylogen Evol 1996;5: 241–6.
25. Lorenz J, Smith D. Distribution of four founding mtDNA haplogroups among native North Americans. Am J Hum Genet 1996;101:307–23.

26. Forster P, Harding R, Torroni A, Bandelt H. Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 1996;59:935–45.
27. Bonatto S, Salzano F. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci USA 1997;94:1866–71.
28. Huoponen K, Torroni A, Wickman PR, Selitto D, Gurley DS, Scozzari R, et al. Mitochondrial DNA and Y chromosome-specific polymorphisms in the Seminole tribe of Florida. Eur J Hum Genet 1997;5:25–34.
29. Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC. MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. Am J Hum Genet 1998;63:1473–91.
30. Smith DG, Malhi RS, Eshleman J, Lorenz J, Kaestle FA. Distribution of mtDNA haplogroup X among Native North Americans. Am J Phys Anthropol 1999;110:271–84.
31. Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Berubg Sea region during the Neolithic. Am J Phys Anthropol 1999;108:1–39.
32. Richards O, Marinez-Labarga C, Lum JK, De Stefano GF, Cann RL. MtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. Am J Hum Genet 1999;65:519–30.
33. Derenko MV, Malyarchuk BA, Dambueva IK, Shaikhaev GO, Dorzhu CM, Nimaev DD, et al. Mitochondrial DNA variation in two South Siberian aboriginal populations: implications for the genetic history of North Asia. Hum Biol 2000;72:945–73.
34. Carlyle SW, Parr RL, Hayes MG, O'Rourke DH. Context of maternal lineages in the greater Southwest. Am J Phys Anthropol 2000;113:85–101.
35. Alves-Silva J, de Silva Santos M, Guimaraes PEM, Ferreira ACS, Bandelt H, Pena SDJ, et al. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 2000;67:444–61.
36. Saillard J, Forster P, Lynnerup N, Bandelt H, Norby S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 2000;67:718–26.
37. Moraga M, Rocco P, Miquel J, Nervi F, Llop E, Chakraborty R, et al. Mitochondrial DNA polymorphisms in Chilean Aboriginal populations: implications for the peopling of the Southern Cone of the continent. Am J Phys Anthropol 2000;113:19–29.
38. Rodriguez-Delfin LA, Rubin-de-Celis VE, Zago MA. Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. Hum Hered 2001;51:97–106.
39. Keyeux G, Rodas C, Gelvez N, Carter D. Possible migration routes into South America deduced from mitochondrial DNA studies in Columbian Amerindian populations. Hum Biol 2001;74:211–33.
40. Demarchi D, Panzetta-Dutari G, Motran C, de Basualdo M, Marcellino A. Mitochondrial DNA haplogroups in Amerindian populations from the Gran Chaco. Am J Phys Anthropol 2001;115:199–203.
41. Budowle B, Fisher CL, Isenberg AR, Monson K, Stewart JEB, Wilson MR, et al. HVI and HVII Mitochondrial DNA population data in Apache and Navajos. Int J Legal Med 2002;116:283–6.
42. Salzano F. Molecular variability in Amerindians: widespread but uneven information. An Acad Bras Cienc 2002;74:223–63.
43. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensic mitochondrial DNA: applications, debates and foundations. Annu Rev Genom Hum Genet 2003;4:119–41.
44. Monson K, Miller K, Wilson M, DiZinno J, Budowle B. The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Sci Comm 2002;4:2.
45. Allard MW, Miller K, Wilson MR, Monson KL, Budowle B. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA data set for 1771 human control region sequences. J Forensic Sci 2002;47:1215–23.
46. Allard MW, Miller K, Wilson MR, Monson KL, Budowle B. Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. Forensic Sci Int 2005;148: 169–79.
47. Allard MW, Wilson MR, Monson K, Budowle B. Control region sequences for East Asian individuals in the SWGDAM forensic mtDNA data set. Legal Med 2004;6:11–24.
48. Andrews R, Kubacka I, Chinnery P, Lightowlers R, Turnbull D, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23:147.
49. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, et al. Seqeunce and organization of the human mitochondrial genome. Nature 1981;290:457–65.
50. Budowle B, Polanskey D, Allard MW, Chakraborty R. Addressing the use of phylogenetics for identification of sequences in error in the SWGDAM mitochondrial DNA database. J Forensic Sci 2004;49(6):1256–61.
51. Polanskey D, Budowle B. Summary of the findings of a quality review of the scientific working group on DNA analysis methods mitochondrial DNA database. Forensic Sci Comm 2005;7(1):1–3; http://www.fbi.gov/hq/lab/fsc/current/resaerch/2005research.htm
52. Nixon K. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 1999;15:407–14.
53. Goloboff P. Nona: a tree search program. Program and documentation that is available from ftp.unt.edu.ar/pub/parsimony and www.cladistics.org, 1994
54. Stewart JEB, Fisher CL, Aagaard PJ, Wilson MR, Isenberg AR, Polanskey D, et al. Length variation in HV2 of the human mitochondrial DNA control region. J Forensic Sci 2001;46:862–70.
55. Torroni A, Richards M, Macaulay V, Forster P, Villems R, Norby S, et al. mtDNA haplogroups and frequency patterns in Europe. Am J Hum Genet 2000;66:1173–7.
56. Kivisild T, Tolk H, Parik J, Wang Y, Papiha S, Bandelt H, Villems R. The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 2002;19:1737–51.
57. Wilson MR, Allard MW, Monson KL, Miller K, Budowle B. Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Int 2002;129:35–41.
58. Wilson MR, Allard MW, Monson KL, Miller K, Budowle B. Further discussions of the consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Sci Comm 2002;4(4): 1–10.
59. Foran D, Hixson JE, Brown WM. Comparisons of ape and human: sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res 1988;17:5841–61.
60. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989;123:585–95.
61. 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.
62. Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC. MITOMAP: a human mitochondrial genome database—1998 update. Nucleic Acids Res 1998;26:112–5, www.mitomap.org.
63. Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P. Phylogeographic differentiation of mitochondrial DNA in Han Chineese. Am J Hum Genet 2002;70:635–51.
64. Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, et al. mtDNA and the Islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet 2001;68:723–37.
65. Salas A, Richards M, De la Fe T, Lareu M, Sobrino B, Sanchez-Diz P, et al. The making of the African mtDNA landscape. Am J Hum Genet 2002;71:1082–111.

Additional information and reprint requests:
Marc W. Allard, Ph.D.
Counterterrorism and Forensic Sciences Research Unit
FBI Laboratory
2501 Investigation Pkwy
Quantico, VA 22135
E-mail: mallard@fbiacademy.edu