Int J Legal Med (2002) 116:212–215
DOI 10.1007/s00414-001-0283-6

ORIGINAL ARTICLE

**Bruce Budowle · Marc W. Allard · Constance L. Fisher
Alice R. Isenberg · Keith L. Monson · John E. B. Stewart
Mark R. Wilson · Kevin W. P. Miller**

HVI and HVII mitochondrial DNA data in Apaches and Navajos

Received: 7 June 2001 / Accepted: 23 November 2001 / Published online: 28 May 2002

Abstract Most mtDNA studies on Native Americans have concentrated on hypervariable region I (HVI) sequence data. Mitochondrial DNA haplotype data from hypervariable regions I and II (HVI and HVII) have been compiled from Apaches (*N*=180) and Navajos (*N*=146). The inclusion of HVII data increases the amount of information that can be obtained from low diversity population groups. Less mtDNA variation was observed in the Apaches and Navajos than in major population groups. The majority of the mtDNA sequences were observed more than once; only 17.8% (32/180) of the Apache sequences and 25.8% of the Navajo sequences were observed once. Most of the haplotypes in Apaches and Navajos fall into the A and B haplogroups. Although a limited number of haplogroups were observed, both sample populations exhibit sufficient variation for forensic mtDNA typing. Genetic diversity was 0.930 in the Apache sample and 0.963 in the Navajo sample. The random match probability was 7.48% in the Apache sample and 4.40% in the Navajo sample. The average number of nucleotide differences between individuals in a database is 9.0 in the Navajo sample and 7.7 in the Apache sample. The data demonstrate that mtDNA sequencing can be informative in forensic cases where Native American population data are used.

Keywords Native Americans · Apache · Navajo · Mitochondrial DNA · HVI · HVII · Random match probability · Genetic diversity · Substitution · Transition · Transversion · Haplotype

Electronic supplementary material to this paper can be obtained by using the Springer Link server located at http://dx.doi.org/10.1007/s00414-001-0283-6.

B. Budowle (✉)
Laboratory Division, FBI,
935 Pennsylvania Avenue, NW, Washington, DC 20535
e-mail: bbudowle@fbi.gov

M.W. Allard
Deptartment of Biological Sciences,
George Washington University, Washington, D.C. 20052

C.L. Fisher · A.R. Isenberg · J.E.B. Stewart · M.R. Wilson
K.W.P. Miller
DNA Analysis Unit II, FBI,
935 Pennsylvania Avenue, NW, Washington, D.C. 20535

K.L. Monson
Forensic Science Research Unit, FBI Academy,
Quantico, VA 22135

Introduction

The genetic sequence(s) of human mitochondrial DNA (mtDNA) of particular interest for forensic analyses are two hypervariable regions, HVI and HVII. These hypervariable regions reside in the control region of the noncoding portion of the mt genome. Because of maternal inheritance and lack of recombination [1, 2, 3, 4, 5], an mtDNA sequence is treated as a single locus or haplotype. When an mtDNA profile from an evidence sample and one from a known reference sample cannot be excluded as originating from the same source, it is desirable to convey some information about the rarity of the mtDNA profile. Typically, the number of matching mtDNA haplotypes in a database(s) can be used to support assessment of the weight to the evidence [6].

The mtDNA haplotypes of Native Americans consist of five major haplogroups, designated as A, B, C, D, and X [7, 8, 9, 10], and Native Americans exhibit lower mtDNA diversity than Africans, Caucasians, Hispanics, and far East Asians [6, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. To date, mtDNA databases on Native Americans developed for forensic applications have not been reported. Generally, mtDNA databases developed for forensic purposes tend to be subjected to more stringent quality assurance and control procedures than those generated for other research endeavors. In this paper, two Native American mtDNA databases (i.e., Apaches and Navajos) have been compiled and analyzed for regions HVI and HVII. The data can be used for forensic analyses and for genetic research studies.

Materials and methods

Blood samples from unrelated Apaches and Navajos were kindly provided by the Arizona Department of Public Safety (Phoenix, Arizona). The Apache samples are from the White Mountain Apache Reservation and were drawn by United Blood Services during a blood drive. The Navajo samples are from reference samples from cases from the Navajo Indian Reservation. The DNA was extracted, amplified, sequenced for regions HVI and HVII, and interpreted as previously described [22, 23].

Sequences are recorded in terms of differences from the profile described by Anderson et al. [24], which is referred to as the Cambridge reference sequence (CRS). Only the nucleotide position(s) (which is a designated number) and the nucleotide(s) differing from the reference standard are recorded. If an unresolved ambiguity is observed at a site, the base number for the site is listed followed by an "N" (e.g., 16228N). Insertions are described by first noting the site immediately 5' to the insertion followed by a point and a "1" (for the first insertion), a "2" (if there is a second insertion), and so on, and then by the nucleotide that is inserted (e.g. 315.1C). Deletions are recorded by listing the missing site followed by a "-" (i.e., 249-).

The population data were analyzed using the program MitoSearch [25], which performs haplotype matching tests and calculates genetic diversity, random match probability (RMP), number of nucleotide differences between and among haplotypes, and the number of transversions, transitions, deletions and insertions. Genetic diversity was calculated according to the method of Tajima [26], and random match probability was calculated according to that of Stoneking et al. [11].

Table 1 Number of different mtDNA haplotypes observed in the Apache database. The most common haplotypes (those occurring at least three times) are specified below the summary table (*Database* Apache, *number of profiles* 180)

Summary of profiles appearing in the database

Number of profiles	Appear	Total
1	32 times	32
1	27 times	27
1	14 times	14
1	12 times	12
1	9 times	9
3	6 times	18
2	5 times	10
1	4 times	4
11	2 times	22
32	1 time	32
54		180

(32) 16111T 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C [A]
(27) 16111T 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 309.1C 315.1C [A]
(14) 16111T 16223T 16290T 16319A 16362C 73G 146C 153G 235G 263G 315.1C [A]
(12) 16111T 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 272G 309.1C 315.1C [A]
(9) 16111T 16147T 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 309.1C 315.1C [A]
(6) 16223T 16298C 16325C 16327T 73G 249D 263G 290D 291D 309.1C 315.1C [C]
(6) 16111T 16183C 16189C 16217C 73G 152C 195C 263G 309.1C 309.2C 309.3C 315.1C [B]
(6) 16111T 16183C 16189C 16217C 73G 152C 195C 263G 309.1C 309.2C 315.1C [B]
(5) 16086C 16111T 16183C 16189C 16217C 73G 263G 309.1C 309.2C 315.1C [B]
(5) 16111T 16183C 16189C 16217C 73G 263G 309.1C 309.2C 315.1C [B]
(4) 16111T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C [A]

Results and discussion

All Apache and Navajo samples have complete sequence data between positions 16024 and 16365 in HVI and between positions 73 and 340 in HVII. Sequences of HVI and HVII regions for all individuals are presented in Table S1.

In African American, Caucasian (i.e., European and/or U.S. origin), Hispanic and Asian population databases of

Table 2 Number of different mtDNA haplotypes observed in the Navajo database. The most common haplotypes (those occurring at least three times) are specified below the summary table (*Database* Navajo, *Number of profiles* 146)

Summary of profiles appearing in the database

Number of profiles	Appear	Total
1	21 times	21
2	10 times	20
1	6 times	6
4	5 times	20
3	4 times	12
5	3 times	15
8	2 times	16
36	1 time	36
60		146

(21) 16111T 16223T 16290T 16319A 16362C 73G 146C 153G 235G 263G 315.1C [A]
(10) 16111T 16183C 16189C 16217C 73G 263G 309.1C 309.2C 315.1C [B]
(10) 16183C 16189C 16217C 73G 150T 263G 309.1C 315.1C [B]
(6) 16183C 16217C 73G 143A 146C 263G 309.1C 309.2C 315.1C [B]
(5) 16111T 16223T 16290T 16319A 16362C 73G 146C 235G 263G 309.1C 315.1C [A]
(5) 16111T 16189C 16192.1T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 315.1C [A]
(5) 16111T 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 272G 309.1C 309.2C [A]
(5) 16093C 16111T 16192T 16223T 16290T 16319A 16362C 73G 146C 153G 235G 263G 315.1C [A]
(4) 16111T 16223T 16290T 16319A 16362C 73G 146C 207A 235G 263G 309.1C 315.1C [A]
(4) 16086C 16223T 16298C 16325C 16327T 73G 249D 263G 290D 291D 309.1C 315.1C [C]
(4) 16092C 16111T 16183C 16189C 16217C 73G 228A 263G 309.1C 315.1C [B]
(3) 16092C 16183D 16186T 16189C 16217C 73G 263G 315.1C [B]
(3) 16093C 16111T 16192T 16223T 16290T 16319A 16362C 73G 146C 153G 235G 263G 309.1C 315.1C [A]
(3) 16111T 16223T 16290T 16319A 16335G 73G 146C 153G 235G 263G 309.1C 315.1C [A]
(3) 16111T 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 309.1C 315.1C [A]
(3) 16111T 16183C 16189C 16217C 73G 263G 315.1C [B]

comparable or greater size than the Native American sample populations in our study, the majority of the mtDNA sequences are observed only once within each population group [6, 15, 16, 17, 18, 20]. In contrast, the Apaches and Navajos have lower diversity, with most mtDNA types being observed more than once. Only 32 out of 180 (or 17.8%) of the haplotypes in Apaches and 36 out of 146 (or 24.7%) of the haplotypes in Navajos are observed only once (Tables 1 and 2). The most commonly observed haplotypes in each population group are listed in Tables 1 and 2. A total of 73 variant sites (42 in HVI and 31 in HVII) were observed in the Apache population sample, and a total of 74 variant sites (41 in HVI and 33 in HVII) were observed in the Navajo population sample.

Native Americans have lower mtDNA diversity because they are composed typically of only five major haplogroups. The data for Apaches and Navajos demonstrate that the A and B haplogroups are the most prevalent with the C haplogroup being less represented (Tables 1 and 2). Of the 180 Apaches, 125 fall into haplogroup A, 35 fall into haplogroup B and 16 fall into haplogroup C. Of the 146 Navajos, 68 fall into haplogroup A, 63 fall into haplogroup B and 11 fall into haplogroup C. At the bottom of Tables 1 and 2, the haplogroup is designated (in brackets) for each of the commonly occurring mtDNA types. In Apaches, the A haplogroup predominates among the five most commonly occurring haplotypes; while in the Navajos, both the A and B haplogroups are well-represented among the top five most commonly occurring haplotypes. The haplogroup distributions for Apache and Navajo data are consistent with the observations of Lorenz and Smith [9]. There are four individuals in each sample population whose mtDNA type could not be categorized into haplogroup A, B, or C. These are Apache samples 37, 75, 130 and 160 and Navajo samples 7, 68, 89, and 114 (Tables 1 and 2). The haplogroup assignment for the mtDNA sequences of the 4 Apache samples are: 37 falls into the K haplogroup, 75 falls into the T haplogroup, 130 falls into the H haplogroup, and 160 falls into the X haplogroup. The haplogroup assignment for the mtDNA sequences of the 4 Navajo samples are: 7 falls into the U2 haplogroup, 68 and 89 fall into the D haplogroup, and 114 falls into the X haplogroup.

Although there are three major haplogroups (i.e., A, B, and C) observed in this study, both population samples are still highly polymorphic (due to private mutations). The Navajos demonstrate a higher degree of diversity compared with Apaches. Genetic diversity is 0.930 in the Apache sample and 0.963 in Navajos (Table 3). The RMP is 7.48% in Apaches and 4.40% in the Navajo sample (Table 3). Although the genetic diversity is lower in Apaches and Navajos compared with other major population groups, the average number of nucleotide differences between individuals in either database is comparable to U.S. Caucasians (9.0 in Navajos and 7.7 in Apaches versus 7.9 in U.S. Caucasians) [6].

The pairwise matching data for the Apaches and Navajos demonstrate that there are similarities between the two groups (Table 4). Pairwise matching frequency is much lower between each of these two Native American groups and other major population groups (i.e., African American and U.S. Caucasian, data not shown).

Table 3 Population groups, sample size, genetic diversity, and random match probability

Database	N	Genetic diversity[a]	Random match probability[b]
Apache	180	0.930	7.48%
Navajo	146	0.963	4.40%

[a]Genetic diversity is calculated according to Tajima [26].
[b]Random match probability is calculated according to Stoneking et al. [11].

Table 4 Pairwise comparisons of mtDNA haplotypes and average number of nucleotide differences between individuals

Populations	Matches/ comparisons	Frequency ($\times 10^{-3}$)	Average # Different nucleotides	(Range)
Navajo	542/10585	0.05120 (1 in 20)	8.969	0–22
Apache	2341/16110	0.14531 (1 in 7)	7.739	0–22
Navajo/ Apache	843/26280	0.03208 (1 in 31)	9.339	0–22

In conclusion, mtDNA population data have been generated for hypervariable regions, HVI and HVII, in Apaches and Navajos. Although the data can be categorized predominantly into three major haplogroups (i.e., A, B, and C), both Native American sample populations exhibit a high degree of mtDNA polymorphism, but not nearly as high as other major population groups. Nevertheless, the data demonstrate that mtDNA sequencing can be informative in forensic cases where Native American population data are used. Moreover, the population data reported here can be considered of high quality because 1) both the light and heavy strands of the amplicons were sequenced, 2) each sequence was read independently by two individuals, and 3) the distribution of the haplogroups is similar to previously reported data [9]. These data will enhance the ability of forensic scientists to estimate the rarity of mtDNA haplotype frequencies.

Acknowledgements This is publication number 01-16 of the Laboratory Division 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. Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 77:6715–6719
2. Jorde L, Barnshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
3. Kivisild T, Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
4. Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931

5. Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029–1032
6. Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL (1999) Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 103:23–35
7. Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani P, Wallace DC (1998) MtDNA haplogroup X: an ancient link between Europe/western Asia and North America? Am J Hum Genet 63: 1852–1861
8. Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59: 935–945
9. Lorenz JG, Smith DG (1996) Distribution of four founding mtDNA haplogroups among native North Americans. Am J Phys Anthropol 101:307–323
10. Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, Cann RL (1999) MtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the North American populations. Am J Hum Genet 65:519–530
11. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48: 370–382
12. Melton T, Stoneking M (1996) Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. J Forensic Sci 41:591–602
13. Melton T, Ginther C, Sensabaugh G, Soodyall H, Stoneking M (1997) Extent of heterogeneity in mitochondrial DNA of sub-Saharan African populations. J Forensic Sci 42:582–592
14. Melton T, Clifford S, Kayser M, Nasidze I, Batzer M, Stoneking M (2001) Diversity and heterogeneity in mitochondrial DNA of North American populations. J Forensic Sci 46:46–52
15. Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steigner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region databases: genetic diversity as a function of sample size and microheterogeneity. Int J Legal Med 112:291–298
16. Pfeiffer H, Forster P, Ortmann C, Brinkmann B (2001) The results of an mtDNA study of 1200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. Int J Legal Med 114:169–172
17. Lutz S, Weisser H-J, Heizmann J, Pollak S (1999) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany (Erratum). Int J Legal Med 112:205–209
18. Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111:124–132
19. Rousselet F, Mangin P (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. Int J Legal Med 111:292–298
20. Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. Int J Legal Med 113:89–97
21. Lee SD, Shin CH, Kim KB, Lee YS, Lee JB (1997) Sequence variation of mitochondrial DNA control region in Koreans. Forensic Sci Int 87:99–116
22. Wilson MR, Polanskey D, Butler J, DiZinno JA, Replogle J, Budowle B (1995) Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. Biotechniques 18:662–669
23. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B (1995) Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 108:68–74
24. Anderson S, Bankier AT, Barrell BG, et al, (1981) Sequence and organization of the human mitochondrial genomes. Nature 290:457–465
25. Monson KL, Miller KWP, Budowle B (2000) MitoSearch: software for managing and searching mitochondrial DNA profiles. Proceedings of the Eleventh International Symposium on Human Identification (in press).
26. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595