Int J Legal Med (2010) 124:263–268
DOI 10.1007/s00414-009-0366-3

ORIGINAL ARTICLE

Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database

Maria Cecilia Bobillo · Bettina Zimmermann · Andrea Sala · Gabriela Huber · Alexander Röck · Hans-Jürgen Bandelt · Daniel Corach · Walther Parson

Received: 23 December 2008 / Accepted: 22 July 2009 / Published online: 13 August 2009
© Springer-Verlag 2009

Abstract The study presents South American mitochondrial DNA (mtDNA) data from selected north (N=98), central (N=193) and south (N=47) Argentinean populations. Sequence analysis of the complete mtDNA control region (CR, 16024–576) resulted in 288 unique haplotypes ignoring C-insertions around positions 16193, 309, and 573; the additional analysis of coding region single nucleotide polymorphisms enabled a fine classification of the described lineages. The Amerindian haplogroups were most frequent in the north and south representing more than 60% of the sequences. A slightly different situation was observed in central Argentina where the Amerindian haplogroups represented less than 50%, and the European contribution was more relevant. Particular clades of the Amerindian subhaplogroups turned out to be nearly region-specific. A minor contribution of African lineages was observed throughout the country. This comprehensive admixture of worldwide mtDNA lineages and the regional specificity of certain clades in the Argentinean population underscore the necessity of carefully selecting regional samples in order to develop a nationwide mtDNA database for forensic and anthropological purposes. The mtDNA sequencing and analysis were performed under EMPOP guidelines in order to attain high quality for the mtDNA database.

Keywords mtDNA population data · Native America · Control region · Coding region · Forensics

Introduction

The variation observed in the mitochondrial DNA (mtDNA) genome is manifested in a nested array of mtDNA haplogroups (hgs) that are more or less regionally specific and can thus be used to reconstruct the impact of pioneer settlement and later migrations (for a review, see [7]). mtDNA polymorphisms have been applied from the very beginning to study the peopling of the Americas, by discerning four "primary maternal lineages" of Asian origin [13, 22, 25], then named A, B, C and D and later specified as haplogroups A2, B2, C1 and D1 [4] that entered the continent via Beringia. Meanwhile, extensive sequencing efforts have revealed further distinction of subhaplogroups and their role as founders (e.g. C1b, C1c and C1d within C1; [1, 23]). Two further major ethnic components contributed to the extant gene pool in the Americas: successive waves of European immigration and the enforced immigration of African slaves during colonial times. Although visible traits and cultural behaviours denote significant European influence, a strong

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0366-3) contains supplementary material, which is available to authorized users.

M. C. Bobillo · A. Sala · D. Corach
Servicio de Huellas Digitales Genéticas,
Facultad de Farmacia y Bioquímica,
Universidad de Buenos Aires,
Buenos Aires, Argentina

B. Zimmermann · G. Huber · W. Parson (✉)
Institute of Legal Medicine, Innsbruck Medical University,
Innsbruck, Austria
e-mail: walther.parson@i-med.ac.at

A. Röck
Institute of Mathematics, University of Innsbruck,
Innsbruck, Austria

H.-J. Bandelt
Department of Mathematics, University of Hamburg,
Hamburg, Germany

Table 1 Diversity measures for north, central and south Argentinean populations

Population statistics	north Argentina	central Argentina	south Argentina
N	98	193	47
Haplotypes	84	176	45
Unique haplotypes	76	166	44
Mean pairwise differences	15.53±6.99	14.47±6.51	14.50±6.61
Haplotype diversity	0.906	0.937	0.878
Random match probability	0.094	0.063	0.122

Statistics are based on entire control region sequences (16024–16569 and 1–576) with C-insertions at positions 16193, 309 and 573 ignored

evidence of Amerindian ancestry is observed in the Argentinean population, mainly when genetic information is assessed [11, 20]. mtDNA sequence data from Argentina are scarce, fragmentary and mostly restricted to the hypervariable segment I (HVS-I), with the exception of a few earlier studies including also the second hypervariable segment (HVS-II; e.g. [10]). However, the separate analysis of the hypervariable segments is prone to error, as the risk of sample mix-up is increased [5, 6]. For forensic purposes, a comprehensive and high-quality dataset is needed, which also reflects regional differences between populations. This is why we distinguished three major geographical regions of Argentina for the mtDNA samples and pursued a comprehensive approach to type control region (CR) variation that includes redundant amplification of the entire CR in two separate PCRs and independent sequence analysis of both amplicons to generate a reliable consensus sequence ([9], updated in [17]). Entire CR mtDNA sequences were augmented by additional coding region information.

Materials and methods

Samples, DNA extraction and control region sequence analysis

A set of 338 samples of unrelated male and female volunteers from Argentina was selected representing three geographical regions of the country (Fig. S1): north(-east) Argentina (N=98) from Formosa (AFO, N=19), Chaco (ACA, N=5), Misiones (AMI, N=48) and Corrientes (ACO, N=26), central(-west) Argentina (N=193) from Santa Fe (ASF, N=6) and Buenos Aires (ABS, N=187) and south Argentina (N=47) from Río Negro (ARN, N=46) and Chubut (ACH, N=1). The sample size of each region was determined by taking into account the relative contribution of each territory to the entire Argentinean population (National Institute of Statistics and Censuses, INDEC 2001; www.indec.mecon.ar). Samples were collected in the laboratory during the period 2005–2007 from volunteer donors who signed a written consent statement form. Personal information was treated anonymously. Blood samples were obtained by finger puncture and spotted onto FTA paper. DNA extraction was performed following the manufacturer's protocol (FTA, Whatman, Piscataway, NJ).

The entire CR from 16024 to 576 was amplified and sequenced following EMPOP recommendations ([9], updated in [17]). The sequences were aligned to the revised Cambridge Reference Sequence (rCRS; [3]) using Sequencher Vs. 4.8 (GeneCodes, Ann Arbor, MI), following updated nomenclature guidelines for mtDNA [8]. All samples were evaluated twice by two independent analysts and results compared using in-house software and finally reviewed by a third analyst.

Analysis of coding region SNPs

Selected coding region single nucleotide polymorphisms (SNPs) targeting haplogroup-specific mutations in East Asian and Native American lineages were analysed using a modified version (Amory et al., in preparation) of a

Table 2 Most common haplotypes of the three regional populations from north, central and south Argentina

Haplogroup	Type (e.g.)	North	Central	South
H	ABS148	1	7	1
A2	ACA357	5	0	0
D1e1	ARN093	0	0	3
A2m	ABS215	1	3	0
D1f	ABS161	0	3	0
T1a	AMI026	3	1	0
C1b	ACO370	3	2	0
C1b	AMI027	3	2	0
D1	AFO006	2	0	0
H11a	AMI056	2	0	0
A2	ABS137	0	2	0
C1c	ABS178	0	2	0
D1f	ABS251	0	2	0
A2l	ACO365	2	1	0
A2l	AMI025	2	1	0
A2	ABS139	1	2	0
B2	ABS146	1	2	0

Investigated range: 16024–576 (C-insertions at positions 16193, 309 and 573 were ignored)

Table 3 AMOVA design and result (*df* stands for degrees of freedom)

Source of variation	*df*	Sum of squares	Variance components	Percentage of variation
Among populations	2	28.79	0.07 Va	0.97
Within populations	335	2,475.95	7.39 Vb	99.03
Total	337	2,504.74	7.46	

previously published assay [2] to confirm the hg assignment derived from the CR patterns and to detail the hg affiliation in haplotypes where the CR did not provide sufficient information. In lineages of Asian and Native American ancestry, positions 12468 (A2c), 6755 (hg B2b), 1888 (hg C1c), 7697 (hg C1d), 2766 (hg D4c1), 8383, 8419, 9431 (hg D4c2), and 5319 (hg D4c2a) were examined in the respective samples [15, 19]. West Eurasian haplotypes were tested at positions 7028, 2706 (hg H), 9716 (hg K2), 14798 (hg J1c), 4580 (hg V) and 12308 and 12372 (hg U).

Statistical analysis was performed using the software package Arlequin 3.0 [12].

Quasi-median networks were constructed for Amerindian and west Eurasian mtDNA datasets using EMPOP Network (www.empop.org) and EMPOP speedy filter as described in [18].

Results and discussion

Sequence variation analysis and summary statistics

The analysis of the entire Argentinean dataset resulted in 288 (85.2%) distinct CR haplotypes, where 24 sequences occurred more than once, disregarding C-insertions beyond positions 16193, 309 and 573. Within the north (N=98), central (N=193) and south (N=47) Argentinean subsets, we observed 76 (77.6%), 166 (86.0%) and 44 (93.6%) unique haplotypes, respectively (Table 1, Table S1). The most frequent sequence (hg H; 263G 315.1C 16519C) was found nine times in the entire Argentinean population (2.7%) with an uneven distribution between the regional populations, one in 98 (1.0%) in the north, seven in 193 (3.6%) in the central and one in 47 (2.1%) in the south Argentinean population. Table 2 lists the most common haplotypes in each regional population, which were less frequent in the other populations and in the entire Argentinean population. Therefore, when the entire Argentinean database would be used to assess the frequency of an mtDNA haplotype, the frequency of each regional population's most common haplotype would be underestimated. Haplotype diversity was 0.906, 0.937 and 0.878 for the subpopulations from north, central, and south Argentina, respectively (Table 1). Analysis of molecular variance brought significant differences between the regional populations at the 0.05 level (Tables 3 and 4).

Network analysis was performed on the Amerindian subset (N=189) and the west Eurasian subset (N=143) for the regions 16024–16569 (covering HVS-I) and 1–576 (covering HVS-II), respectively. The individual torsos of the networks displayed the expected star-like shapes (Fig. S2a, b, c, d; [7]). Reticulations were observed for mutations that are known to occur in different haplogroups, such as the transition at position 513, which has been described in hgs H5 and J2a, or the transversion at position 57, which went parallel in hgs A2 and B2 (Fig. S2d). We observed numerous transversions (Fig. S1, Table 5), some of which are common [such as those flanked by poly-C tracts (16182C, 16183C, 16265C, and 16293C)]. Some transversions are characteristic for (sub)haplogroups, such as 16114A (in L1b, L2b, G4), 16147G (N1a) or 16188G (L0a), while others have rarely been detected, such as 16058C, 16239A, 16327A and 573A. Besides frequent insertions in the poly-C tracts with respect to rCRS, we observed rare insertions, viz. 46.1T (in C1b6), 71.1G (in D1*) and 291.1A (in A2*). In addition to expected deletions scored at positions 249, 290 and 291 (hg C1), 298 (hg K1c) or within the poly-C tracts and the dimeric repeat region around position 524, we observed one rare deletion event at position 469 in hg H as well as the "Chibcha deletion motif" 106-111del in hgs D1f and D4h3a, which has been described before in other Amerindian populations [16, 21]. Point heteroplasmy was detected in 29 instances at altogether 26 positions in the 338 investigated samples (Table S1), including 16124Y, 16129R, 16142Y, 16157Y, 16163R, 16172Y, 16183M, 16185Y, 16188Y, 16189Y, 16218Y, 16261Y, 16263Y, 16274R, 16298Y, 16318R, 16327Y, 16356Y, 16359Y, 16362Y, 9R, 64Y, 146Y, 153R, 200R, and 204Y.

Within the less variable CR segment 16366–72, a total of 30 nucleotide substitutions were detected with respect to rCRS (defining 46 distinct haplotypes), whereas within the

Table 4 F_{ST} comparison among the regional populations: F_{ST} values are below the diagonal and the p values (1,023 permutations, significance level=0.05) above the diagonal

	North	Central	South
North	–	0.00781	0.02832
Central	0.00971	–	0.05957
South	0.01303	0.00800	–

Table 5 Transversions found in the mtDNA control region of the Argentinean population sample (N=338)

HVS-I	Previously observed	HVS-II/III	Previously observed
16058C	B4b, L3e1	57G	A2, B4b, H1a, H15
16114A	B4a, C4a, G4, HV, L0f, L1b, L1c, L2a1, L2b, L3d, M5, M7b1, R0, U5a, X2	62T	A2, Q3a
16147G	H, N1a	186A	H, L1c, M2b
16166C	C, H, HV, I2, L0k, M47	189C	L1c, L3h
16176A	N1b	431A	K1a7, M28a
16176G	N1b2	512C	B2d, K2a2a
16182C	Common	573A	A2
16183C	Common		
16187A	D1, H		
16188G	H1i, L0a		
16192A	C1, H, L0d3, R22		
16239A	A2, H, L2a1		
16240C	D, H, M2, T2, U2c		
16265C	C, D5b, F, G3a, H5, J, K, L1c, L2, L3e3, M2a, M5a, Q1, U5a, W		
16293C	A, A8, D4, H, I, M1, M4'30, M8, M35a		
16318T	B4a, B5b, C4b2, J1a, K, M18, M28b, U3a, U7		
16327A	H, M, T1, T2		

less variable segment (341–457) between HVS-II and the third hypervariable segment, five additional substitutions were found (defining seven haplotypes). Such information would not be available when sequence analysis would have been restricted to HVS-I (16024–16365) and HVS-II (73–340).

Haplogroup frequencies

Haplogroups were defined on the basis of entire CR sequences, and additional coding region SNPs were typed for confirmation and for refined hg assignment (Table S1). In the north, the paragroups/haplogroups C1b* (17.3%), A2* (13.3%), H (12.2%), B2* (11.2%) and U (10.2%) prevailed, whereas paragroups/haplogroups D1e (27.7%), H (14.9%), A2* (10.6%), C1b6 (6.4%) and B2e (6.4%) were most prominent in the south. In both regions, the Amerindian haplogroups were observed at a frequency exceeding 60%. These results are concordant with the expected distribution of Amerindian people in Argentina. In contrast, the central population represented the Amerindian haplogroups at slightly lower frequency (48.2%) and revealed a predominant west Eurasian pool of haplogroups R0 (24.4%), JT (10.9%), U (10.9%), with only one Amerindian haplogroup above a frequency of 10%, viz. A2 (13.5%). This mtDNA pattern in central Argentina can be explained by the consecutive European migration waves, mostly coming from Spain and Italy, which influenced the population composition in the industrial areas of the country. Interestingly, the matrilineal contribution of African ancestry mainly traceable by haplogroups L0, L1 and L2 and the African-specific subhaplogroups of hg L3 is remarkably low in all three regions (north and south, 2%; central, 1%; Fig. S1) in our sample set.

Coding region analysis

The CR sequence motif of hg B2 lineages cannot be distinguished from the Asian hg B4b in which B2 is nested. Therefore, we screened coding region position 3547 in 33 samples to confirm B2 status [2]. All B2 samples were further tested for the transition at position 6755 (hg B2b status). Hg D1 is characterised by a transition at position 16325 in the CR (in addition to mutations at positions 16223, 16362, 73, and 489 expected for D4 status). In four samples (ARN104, ARN106, ARN110 and ARN127), the transition at position 16325 was missing, although we observed the D1-specific coding region transition at 2092. To determine subhaplogroup status within hg C1, positions 1888 (defining hg C1c) and 7697 (defining hg C1d) were examined in 17 samples. On the basis of these results, samples ABS178, ABS179, ABS293, ABS317, AMI046 and AMI052 were assigned to hg C1c, and samples ABS155, ABS228, ABS229, ABS284, ABS299, ABS326, ACO388, ACO394, ARN112, ARN116 and ABS174 were assigned to hg C1d, although the latter sample did not harbour the transition at position 7697 but instead shared several CR mutations with C1d sequences that were confirmed by the 7697 variant. The analysis of positions

8383, 8419 and 9431 revealed the hg D4c2 status of sample AMI038; moreover, the CR sequence nearly matched (except for the insertion of two cytosines at 309) three Japanese hg D4c2 sequences [24]. The detection of this haplogroup is concordant with the demographic characteristics of the province of Misiones, where this sample was obtained. In this province, the most relevant Asian immigration waves occurred during the early twentieth century [14]. The hg J1c status of sample ABS190 (absence of 228A) was confirmed by inspection of position 14798. The hg V status of samples ABS140, ABS245, ABS264, AMI061, ABS134, ABS165, ABS248, and ARN078 was confirmed by analysing position 4580. Two K samples (ABS244 and AMI049) were assigned to subhaplogroup K2 by analysing position 9716. The analysis of positions 2706 and 7028 verified hg H status of 39 samples: ABS148, ABS152, ABS158, ABS162, ABS166, ABS169, ABS186, ABS187, ABS191, ABS204, ABS209, ABS211, ABS217, ABS223, ABS237, ABS252, ABS265, ABS273, ABS279, ABS295, ABS300, ABS307, ABS309, ABS335, ABS337, ABS345, ABS353, ACA356, AMI041, AMI058, AMI059, AMI069, AMI070, AMI074, ARN084, ARN091, ARN100, ARN108 and ARN125, whereas samples ABS149, ABS219 remained in R0* and ABS340 in HV0*.

Conclusion

The results of the present study corroborate the strong Amerindian influence with a pronounced regional imprint as well as a significant European contribution to the makeup of the Argentinean population. The African input that used to be prevalent during the late eighteenth and early nineteenth century is nowadays limited to a very minor proportion of the gene pool. Haplogroup proportions of the investigated regions reinforce the need of creating local mtDNA population databases as well as of sequencing the entire mitochondrial genome.

The haplotypes reported in the present work will be made available via EMPOP (www.empop.org) upon publication.

This paper follows the recommendations of the ISFG on the use of mtDNA in forensic analysis.

Acknowledgements The authors would like to thank Daniela Niederwieser and Theresa Harm for excellent technical assistance. The study was supported in part by the FWF Austrian Science Fund (TR397) to WP and AR and by the National Research Council of Argentina (CONICET) PIP 6114 Res. 438/05 and UBACyT B-035 grants to DC. DC and AS are members of the Carrera del Investigador Científico y Tecnologico and MCB received a doctoral fellowship from UBACyT.

References

1. Achilli A, Perego UA, Bravi CM et al (2008) The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. PloS ONE 3:e1764
2. Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A (2007) Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. Forensic Sci Int Genet 1: 44–55
3. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147
4. Bandelt H-J, Herrnstadt C, Yao Y-G et al (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet 67:512–524
5. Bandelt H-J, Salas A, Bravi C (2004) Problems in FBI mtDNA database. Science 305:1402–1404
6. Bandelt H-J, Salas A, Lutz-Bonengel S (2004) Artificial recombination in forensic mtDNA population databases. Int J Legal Med 118:267–273
7. Bandelt H-J, Macaulay V, Richards M (eds) (2006) Human mitochondrial DNA and the evolution of *Homo sapiens*. Springer, Berlin
8. Bandelt H-J, Parson W (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med 122:11–21
9. Brandstätter A, Klein R, Duftner N, Wiegand P, Parson W (2006) Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm). Int J Legal Med 120:310–314
10. Corach D, Sala A, Penacino G et al (1997) Additional approaches to DNA typing of skeletal remains: the search for "missing" persons killed during the last dictatorship in Argentina. Electrophoresis 18:1608–1612
11. Corach D, Marino MA, Sala A (2006) Relevant genetic contribution of Amerindian to the extant population of Argentina. Int Congr Ser (Elsevier, Amsterdam) 1288:397–399
12. Excoffier L, Laval LG, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47–50
13. Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol Biol Evol 10:23–47
14. Kikumura-Yano A (ed) (2002) Encyclopedia of Japanese descendants in the Americas: an illustrated history of the Nikkei. Japanese American National Museum, edition number: 336. AltaMira, Walnut Creek
15. Kong QP, Bandelt HJ, Sun C et al (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Gen 15(13):2076–2086
16. Merriwether DA, Ferrell RE, Rothhammer F (1995) mtDNA D-loop 6-bp deletion found in the Chilean Aymara: not a unique marker for Chibcha-speaking Amerindians. Am J Hum Genet 56:812–813
17. Parson W, Bandelt H-J (2007) Extended guidelines for mtDNA typing of population data in forensic science. Forensic Sci Int Genet 1:13–19
18. Parson W, Dür A (2007) EMPOP—a forensic mtDNA database. Forensic Sci Int Genet 1:88–92
19. Perego UA, Achili A, Angerhofer N et al (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr Biol 19:R203–R205

20. Salas A, Jaime JC, Álvarez-Iglesias V, Carracedo Á (2008) Gender bias in the multiethnic genetic composition of central Argentina. J Hum Genet 53:662–674
21. Santos M, Barrantes R (1994) D-loop mtDNA deletion as a unique marker of Chibchan Amerindians. Am J Hum Genet 55:413–414
22. Schurr TG, Ballinger SW, Gan YY et al (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613–623
23. Tamm E, Kivisild T, Reidla M et al (2007) Beringian standstill and spread of Native American founders. PLoS ONE 2:e829
24. Tanaka M, Cabrera VM, González AM et al (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res 14:1832–1850
25. Torroni A, Schurr TG, Yang C-C et al (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162