Forensic Population Genetics – Short Communication

Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina

Sergio Cardoso [a], Leire Palencia-Madrid [a], Laura Valverde [a], Miguel A. Alfonso-Sánchez [b], Luis Gómez-Pérez [b], Emma Alfaro [c], Claudio M. Bravi [d], José E. Dipierri [c], José A. Peña [b], Marian M. de Pancorbo [a,*]

[a] BIOMICs Research Group, Centro de Investigación "Lascaray" Ikergunea, Universidad del País Vasco (UPV/EHU), Vitoria-Gasteiz, Spain
[b] Dpto. Genética, Antropología Física y Fisiología Animal, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), Leioa, Spain
[c] Instituto de Biología de la Altura, Universidad Nacional de Jujuy, Avda. Bolivia 1661, 4600 San Salvador de Jujuy, Argentina
[d] Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), CCT-CONICET-La Plata, Argentina

ARTICLE INFO

ABSTRACT

Mitochondrial control region (16024-576) sequences were generated from 180 individuals of four population nuclei from the province of Jujuy (NW Argentina), located at different altitudes above sea level. The frequency at which a randomly selected mtDNA profile would be expected to occur in the general population (random match probability) was estimated at 0.011, indicating a relatively high diversity. Analysis of the haplogroup distribution revealed that Native American lineages A2 (13.9%), B (56.7%), C1 (17.8%), D1 (8.9%) and D4h3a (1.1%) accounted for more than 98% of the total mtDNA haplogroup diversity in the sample examined. We detected a certain degree of genetic heterogeneity between two subpopulations located at different points along the altitudinal gradient (Valles and Puna), suggesting that altitude above sea level cannot be ruled out as a factor promoting divergences in mtDNA haplogroup frequencies, since altitude is closely associated with human living conditions, and consequently, with low demographic sizes and the occurrence of genetic drift processes in human communities. In all, mitochondrial DNA database obtained for Jujuy province strongly points to the need for creating local mtDNA databases, to avoid bias in forensic estimations caused by genetic substructuring of the populations.

1. Population

The entire control region of the mitochondrial DNA (mtDNA) remains virtually unexplored in many human populations from South America. This obviously hinders the construction of high-quality mtDNA databases, which in turn amplifies the limitations inherent to this molecular marker for forensic applications. The lacking of extensive mtDNA databases might be especially problematic in large countries such as Argentina, characterized by the presence of many different human groups with disparate levels of genetic diversity. A recent study on the mtDNA diversity in Argentina revealed that the country harbors a comprehensive admixture of worldwide mtDNA lineages [1]. Some of these lineages show regional specificity, thus conditioning the construction of a nationwide mtDNA database for forensic and anthropological purposes. In this regard, Catelli et al. [2] demonstrated the existence of a clear-cut genetic substructuring in Argentina and the need for in-depth analyses of the different regions of the country to gradually disentangle its stratification. Here we present the first database of the complete mtDNA control region in a sample from the province of Jujuy (northwestern Argentina). Jujuy is characterized by remark-able differences in altitude along the different geographic regions of the province (see Fig. 1), all of them well differentiated by their typical environmental conditions: Selva or Yungas, located between 500 and 1200 m above sea level (m.a.s.l.), Valles (~1200 m.a.s.l.), which include the capital city of the province (San Salvador de Jujuy), Quebrada de Humahuaca (2000–3500 m.a.s.l.), and Puna (>3500 m.a.s.l.). Because of historico-geographical reasons, Native American individuals are proportionally more numerous in the highest regions of Jujuy, where even at present the Amerindian genetic component is highly predominant [3]. Furthermore, the region of Jujuy has been traditionally characterized by markedly low population densities over time notably in those areas of more difficult access (highlands),
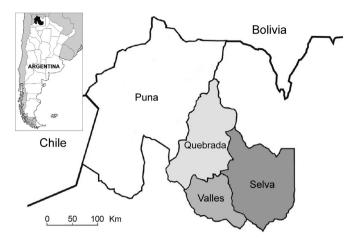
* Corresponding author at: BIOMICs Research Group, Centro de Investigación "Lascaray" Ikergunea, Universidad del País Vasco (UPV/EHU), Avda. Miguel de Unamuno 3, 01006 Vitoria-Gasteiz, Spain. Tel.: +34 945014528; fax: +34 945014458.
E-mail address: marian.mdepancorbo@ehu.es (M.M. de Pancorbo).

Fig. 1. Map showing the location of the province of Jujuy in northwestern Argentina. The subdivision of Jujuy in different regions (Selva, Valles, Quebrada and Puna) is also displayed.

which might have promoted genetic drift events, and therefore, genetic microdifferentiation processes between the diverse population nuclei of the zone.

2. Subjects, materials and methods

Peripheral blood was taken from 180 healthy and maternally unrelated individuals from the province of Jujuy (northwestern Argentina). Autochthony was verified by biographical information traced back at least three generations (mainly birthplaces of the donor's parents and grandparents). The total sample was divided into four different subgroups according to the sampling region: 59 from San Salvador de Jujuy (Valles), 41 from Selva, 41 from Quebrada de Humahuaca, and 39 from Puna. Samples from Valles and Selva were collected in urban and semi-urban areas. In these communities, admixture rates (∼23%) of the Native American genetic background estimated from autosomal markers (polymorphic *Alu* insertions) indicated a major contribution of European ancestry populations, and a much more weak African component (Valle: 6.5%). Sampling locations in Quebrada de Humahuaca were rural communities with admixture rates between 8 and 13%, but in this case with a predominantly African contribution. Puna donors lived in rural, isolated localities with virtually no signals of gene flow in autosomal *Alu* elements [4]. All donors gave their informed consent prior to inclusion in the sample. DNA extraction was performed using the standard phenol:chloroform procedure [5]. DNA was quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) in a DTX880 Multimode Detector (Beckman Coulter, Fullerton, CA). Full mitochondrial control region (CR) was amplified, sequenced and interpreted as reported in Cardoso et al. [6] using rCRS as reference sequence, instead of the recently released RSRS, as proposed by Salas et al. [7] for forensic DNA studies and casework.

3. Analysis of data

Diversity parameters and haplotype frequencies were calculated using the Arlequin Software v3.5 [8]. The probability of two randomly selected individuals from a population having identical mtDNA types was calculated as $p = \sum X_i^2$. The length polymorphisms of the poli-C stretches at HVS-I, HVS-II and HVS-III were disregarded in the analyses.

The mtDNA haplogroup assignment was performed according to the mtDNA tree Build 15 of the Phylotree [9]. Statistical differences in haplogroup distribution among subpopulations

were calculated using PAST software [10]. The conservative Bonferroni correction for multiple testing was used when necessary [11].

All the sequences have been deposited into EMPOP [12] under accession number EMP00512. Likewise, sequences are available online at GenBank under accession numbers JQ736825–JQ737004.

4. Results and discussion

Summary statistics are shown in Table 1, whereas haplotypes are listed in Supplementary Table S1. A total of 188 polymorphic sites were identified, which defined 127 haplotypes. Of them, 96 were unique. This sequence variability rendered a high gene diversity value (0.9945 ± 0.0014) considering the entire control region. This figure is clearly higher than the diversity values estimated in other South American populations where the Native American maternal component is prevalent, such as Kichwa (0.8029 ± 0.0373) from Ecuador [13] and Mapuche (0.9725 ± 0.0107) from Argentina (calculated from de Saint Pierre et al. [14]). Interestingly, the sequence variability observed in Jujuy was comparable to the high values obtained for other South American populations with a long history of admixture, such as Santa Catarina in Brazil (0.9930) [15] or North Central Venezuela (0.9939) [16]. The great diversity of the maternal lineages was also reflected in the probability of a random match (RMP), with a value of 0.011.

Disregarding length variations in the poli-C stretches, the most frequent haplotypes in Jujuy, with 5 occurrences (2.8%) each, were (i) 263G-315.1C-499A-16183C-16188T-16189C-16217C-16519C, (ii) 73G-146C-249DEL-263G-290DEL-291DEL-315.1C-489C-493G-523DEL-524DEL-16223T-16298C-16325C-16327T-16519C, (iii) 63C-64T-66A-73G-185A-263G-315.1C-499A-16183C-16188T-16189C-16217C-16519C, and (iv) 46G-59A-61T-62T-73G-195C-249DEL-263G-290DEL-291DEL-315.1C-489C-493G-523DEL-524DEL-16092C-16223T-16249C-16298C-16325C-16327T-16400T-16463G-16519C. Haplotypes (i), (ii) and (iii) were identified in different Jujuy subregions; however, the distribution of haplotype (iv) was exclusively restricted to Puna, suggesting genetic drift effects in the area most likely promoted by prolonged population isolation, and accordingly, a traditional small community size.

Phylogeographic comparisons of the haplotypes identified in our study with those of other South American groups were hindered by the lack of high quality, complete control region sequences of relevant populations. Notwithstanding these limitations, we used a clade-defining approach to search for (near)-matches in an in-house database of published sequences of Native American origin derived from both indigenous and admixed South American populations.

Overall, most of the (near)-matches available for Jujuy samples pointed to a shared biological (maternal) history of southern Central Andean populations (Supplementary Table S2). The most frequent clade in Jujuy (*N* = 28) was the B2 haplotype carrying 16188T. Even though this clade has been previously described in different South American populations, the more

Table 1
Summary statistics for 180 control region (16024-576) sequences in a sample from Jujuy province (NW Argentina). Insertions at nucleotide positions 16193, 309, and 573 were ignored for all calculations.

Population statistics	Jujuy (*N* = 180)
Random match probability (sum of squares)	0.011
Haplotypes	127 (31 shared)
Polymorphic positions	188
Mean pairwise differences	14.7932 ± 6.6503
Genetic diversity	0.9945 ± 0.0014

Table 2
Mitochondrial DNA haplogroup distribution in four subpopulations from Jujuy province (NW Argentina). Relative frequencies (as percentages) in parentheses.

	N	Haplogroup and frequency						P value[a]		
		A	B	C	D	L	T	Puna	Quebrada	Valles
Puna	39	8 (20.5)	17 (43.6)	8 (20.5)	6 (15.4)		–	–		
Quebrada	41	7 (17.1)	25 (61.0)	7 (17.1)	2 (4.8)	–	–	0.3739	–	
Valles	59	3 (5.1)	39 (66.1)	9 (15.3)	5 (8.5)	2 (3.4)	1 (1.7)	0.0396	0.4096	–
Selva	41	7 (17.1)	21 (51.2)	8 (19.5)	5 (12.2)	–	–	0.9147	0.7907	0.3025

[a] Significance values of likelihood ratio tests (G-tests) to compare haplogroup frequencies between subpopulations. Significance *p* value 0.0125 after Bonferroni correction.

relevant frequencies appear in the southern Central Andes, peaking at 22–29% in North Chile, Bolivia and Peru [14,17,18].

Matches with haplotypes derived from some aDNA studies were also observed. In this regard, five ancient individuals from Nasca-Paracas in coastal southern Peru [19] showed a common pronounced motif with three mtDNA sequences from Puna.

Some lineages, such as those attributable to D1j and B2i2b show affinities with Central-Western Argentina and Patagonia most likely determined by gene flow processes. On the other hand, available mtDNA data clearly indicated the seclusion of two clusters in Jujuy region. One of them was a C1b clade, the second most frequent in our sample and signposted by the co-occurrence of 16092C-16400T, a combination with high incidence in Puna and, to the best of our knowledge, not yet reported. The other is a B2 clade with the motif 16145A-16156A-16157C present in five individuals, four of them from Quebrada.

As for haplogroup composition, the population from Jujuy showed a distribution almost entirely composed of Native American lineages (98.3%). Only three non-American haplogroups were observed: two African paragroup L sequences and a T2c lineage of European ancestry. These three exotic lineages were identified in Valles, the region of Jujuy showing the highest level of genetic admixture according to previous studies based on Y-chromosome markers [20] or on polymorphic *Alu* insertions [4]. These findings suggest that, despite the population admixture processes occurred in Jujuy since the 16th century between the indigenous population, the European settlers and the African slaves, there is a high predominance of the ancestral Native American substrate in the maternal lineages, probably mirroring the strong sexual bias of the immigrants settled in this sub-Andean zone from Argentina during colonial times. Asymmetry in the ancestry composition of maternal and paternal lineages has been reported in previous genetic studies carried out in other Ibero-American human groups from Brazil [21,22] and from other ancient Spanish colonies [23–25]. In Jujuy, as expected, the weak effect of the gene flow was detectable only in samples from the lowest geographic regions, since immigrants concentrated mainly in these zones [26].

With regard to the Native American lineages, we identified haplogroups A2 (13.9%), B4b (56.7%), C1 (17.8%), D1 (8.9%) and D4h3a (1.1%) in the study population. The prevalence of haplogroup B is commonplace among indigenous human groups from Central Andean South America, as corroborated by findings in other Amerindian populations such as Aymara (81%) and Quechua (61%) from Bolivia [27], Atacama (72%) from Chile [28], individuals from the Department of Ancash (52%), Peru [29], or Peruvian Amerindians from the Lake Titicaca basin (69%) [17], among others. Results of a likelihood-ratio test (G-test) indicated no statistically significant differences between the haplogroup distribution of the four Jujuy subpopulations considered (Table 2), at least at the macrohaplogroup level (A, B, C and D). Nevertheless, such results were indicative of a certain degree of genetic heterogeneity between two subpopulations sited at different points of the altitude range (Valles and Puna). As can be noted, the comparison between the haplogroup frequencies of these two subgroups rendered a significance value (P = 0.0396) very close to the threshold value obtained after applying the Bonferroni correction for multiple testing (0.0125), thus suggesting that differences in altitude cannot be ruled out as a potential factor in determining divergences in mtDNA haplogroup frequencies. In all probability, statistically significant differences in haplogroup distribution among Jujuy subpopulations might also be detected at a more refined level of the mtDNA phylogeny (subhaplogroups), or perhaps by increasing the sample size of the communities examined.

The existence of genetic heterogeneity even at a microgeographic scale strongly supports the creation of local databases such as the one presented herein, to avoid underestimation or overestimation of a given haplotype in routine forensic casework. These local databases can also be helpful in biomedical investigations, including case-control studies.

The findings of our study together with the information on mtDNA reported so far for Argentina both substantiate the notion that, in countries with quite isolated populations distributed over large territories, the creation of local mtDNA databases is essential.

5. Quality control

All data generated in this study are in compliance with ISFG and EDNAP guidelines for mitochondrial DNA analysis [30,31]. This publication follows ISFG guidelines for the publication of population genetic data [32].

Conflict of interest

Authors declare no competing interest in the content of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2013.01.007.

References

[1] M.C. Bobillo, B. Zimmermann, A. Sala, G. Huber, A. Röck, H.J. Bandelt, D. Corach, W. Parson, Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database, Int. J. Legal Med. 124 (2010) 263–268.

[2] M.L. Catelli, V. Alvarez-Iglesias, A. Gómez-Carballa, A. Mosquera-Miguel, C. Romanini, A. Borosky, J. Amigo, A. Carracedo, C. Vullo, A. Salas, The impact of modern migrations on present-day multi-ethnic Argentina as recorded on the mitochondrial DNA genome, BMC Genet. 12 (2011) 77.

[3] J.E. Dipierri, E. Alfaro, J.A. Peña, J. Constans, J.M. Dugoujon, GM, KM immunoglobulin allotypes and other serum genetic markers (HP, GC, PI, and TF) among South American populations living at different altitudes (Jujuy Province, Argentina): admixture estimates, Hum. Biol. 72 (2000) 305–319.

[4] L. Gómez-Pérez, M.A. Alfonso-Sánchez, J.E. Dipierri, E. Alfaro, S. García-Obregón, M.M. de Pancorbo, G. Bailliet, J.A. Peña, Microevolutionary processes due to landscape features in the province of Jujuy (Argentina), Am. J. Hum. Biol 23 (2011) 177–184.

[5] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.

[6] S. Cardoso, M.J. Villanueva-Millán, L. Valverde, A. Odriozola, J.M. Aznar, S. Piñeiro-Hermida, M.M. de Pancorbo, Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country, Forensic Sci. Int. Genet. 6 (2012) e106–e108.

[7] A. Salas, M. Coble, S. Desmyter, T. Grzybowski, L. Gusmão, C. Hohoff, M.M. Holland, J.A. Irwin, T. Kupiec, H.Y. Lee, B. Ludes, S. Lutz-Bonengel, T. Melton, T.J. Parsons, H. Pfeiffer, L. Prieto, A. Tagliabracci, W. Parson, A cautionary note on switching mitochondrial DNA reference sequences in forensic genetics, Forensic Sci. Int. Genet. 6 (2012) e182–e184.

[8] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.

[9] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2) (2009) E386–E394.

[10] Ø. Hammer, D.A.T. Harper, P.D. Ryan, PAST: paleontological statistics software package for educational and data analysis, Paleontol. Electron. 4 (2001) 9.

[11] J.M. Bland, D.G. Altman, Statistics notes: multiple significance tests: the Bonferroni method, BMJ 310 (1995) 170.

[12] W. Parson, A. Dür, EMPOP – a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88–92.

[13] M. Baeta, C. Núñez, C. Sosa, M. Bolea, Y. Casalod, F. González-Andrade, L. Roewer, B. Martínez-Jarreta, Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador, Int. J. Legal Med. 126 (2012) 299–302.

[14] M. de Saint Pierre, C.M. Bravi, J.M. Motti, N. Fuku, M. Tanaka, E. Llop, S.L. Bonatto, M. Moraga, An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups, PLoS One 7 (2012) e43486.

[15] L. Palencia, L. Valverde, A. Alvarez, L.M. Cainé, S. Cardoso, M.A. Alfonso-Sánchez, M.F. Pinheiro, M.M. de Pancorbo, Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): predominance of the European input, Int. J. Legal Med. 124 (2010) 331–336.

[16] D. Castro de Guerra, C. Figuera Pérez, C.M. Bravi, J. Saunier, M. Scheible, J. Irwin, M.D. Coble, A. Rodríguez-Larralde, Sequence variation of mitochondrial DNA control region in North Central Venezuela, Forensic Sci. Int. Genet. 6 (2012) e131–e133.

[17] C. Barbieri, P. Heggarty, L. Castrì, D. Luiselli, D. Pettener, Mitochondrial DNA variability in the Titicaca basin: matches and mismatches with linguistics and ethnohistory, Am. J. Hum. Biol. 23 (2011) 89–99.

[18] H.H. Afonso-Costa, M. Carvalho, V. Lopes, F. Balsa, A.M. Bento, A. Serra, L. Andrade, M.J. Anjos, M.C. Vide, S. Pantoja, D.N. Vieira, F. Corte-Real, Mitochondrial DNA sequence analysis of a native Bolivian population, J. Forensic Leg. Med. 17 (2010) 247–253.

[19] L. Fehren-Schmitz, M. Reindel, E.T. Cagigao, S. Hummel, B. Herrmann, Pre-Columbian population dynamics in coastal southern Peru: a diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis, Am. J. Phys. Anthropol. 141 (2010) 208–221.

[20] J.E. Dipierri, E. Alfaro, V.L. Martínez-Marignac, G. Bailliet, C.M. Bravi, S. Cejas, N.O. Bianchi, Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina, Hum. Biol. 70 (1998) 1001–1010.

[21] A.K. Ribeiro-dos-Santos, J.M. Pereira, M.R. Lobato, B.M. Carvalho, J.F. Guerreiro, S.E. Batista Dos Santos, Dissimilarities in the process of formation of Curiaú, a semi-isolated Afro-Brazilian population of the Amazon region, Am. J. Hum. Biol. 14 (2002) 440–447.

[22] M.C. Bortolini, W.A. Da Silva Jr., D.C. De Guerra, G. Remonatto, R. Mirandola, M.H. Hutz, T.A. Weimer, M.C. Silva, M.A. Zago, F.M. Salzano, African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers, Am. J. Hum. Biol. 11 (1999) 551–563.

[23] I. Mendizabal, K. Sandoval, G. Berniell-Lee, F. Calafell, A. Salas, A. Martínez-Fuentes, D. Comas, Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba, BMC Evol. Biol. 8 (2008) 213.

[24] L.G. Carvajal-Carmona, I.D. Soto, N. Pineda, D. Ortíz-Barrientos, C. Duque, J. Ospina-Duque, M. McCarthy, P. Montoya, V.M. Alvarez, G. Bedoya, A.A. Ruiz-Linares, Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia, Am. J. Hum. Genet. 67 (2000) 1287–1295.

[25] G. Bedoya, P. Montoya, J. García, I. Soto, S. Bourgeois, L. Carvajal, D. Labuda, V. Alvarez, J. Ospina, P.W. Hedrick, A. Ruiz-Linares, Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 7234–7239.

[26] C. Martinez-Sarasola, Nuestros paisanos los indios, EMECE, Buenos Aires, 1992.

[27] M. Gayà-Vidal, P. Moral, N. Saenz-Ruales, P. Gerbault, L. Tonasso, M. Villena, R. Vasquez, C.M. Bravi, J.M. Dugoujon, mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: different stories and special genetic traits of the Andean Altiplano populations, Am. J. Phys. Anthropol. 145 (2011) 215–230.

[28] D.A. Merriwether, F. Rothhammer, R.E. Ferrell, Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World, Am. J. Phys. Anthropol. 98 (1995) 411–430.

[29] C.M. Lewis Jr., R.Y. Tito, B. Lizárraga, A.C. Stone, Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru, Am. J. Phys. Anthropol. 127 (2005) 351–360.

[30] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.

[31] G. Tully, W. Bär, B. Brinkmann, A. Carracedo, P. Gill, N. Morling, W. Parson, P. Schneider, Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles, Forensic Sci. Int. 124 (2001) 83–91.

[32] A. Carracedo, J. Butler, L. Gusmão, W. Parson, L. Roewer, P. Schneider, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147.