



Analysis of 21 X-chromosome polymorphisms in urban and rural populations in Salta province (north-western Argentina)

J. F. Ferragut¹ · M. Bassitta¹ · V. Torrens¹ · V. Albeza² · N. Acreche² · J. A. Castro¹ · C. Ramon¹ · A. Picornell¹

Received: 23 August 2018 / Accepted: 12 November 2018 / Published online: 16 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Population genetic data for 21 X-chromosome markers (Alu insertions and STRs) are reported for two populations (rural and urban) in Salta province (north-western Argentina). New variants are described, confirming the complexity and variability of some markers in this set. Results reveal Salta populations harbor a high Native American component, despite their self-recognized European ancestry. Notwithstanding the high genetic similarity of both populations, the rural sample seems to have maintained a larger Amerindian legacy. Data further show these X-linked markers, especially STRs, are highly informative in Salta populations and, therefore, can contribute to the development of a local database for forensic purposes in north-western Argentina.

Keywords X-chromosome · Alu insertions · X-STR · Calchaqui Valleys · Salta · North-western Argentina (NWA) · Investigator Argus X-12 kit

Salta is a province in northwest Argentina. Before the Spanish conquest, numerous native people lived in the region [1]. Salta city (SC) was founded in 1582 in the Lerma Valley, 1152 m above sea level. Including its metropolitan area, it has a population of 535,303 inhabitants [2]. After its foundation, the city grew exponentially, due to agriculture and mineral resources and its strategic position for trade between Lima and Buenos Aires until 1816 when the area was submerged in an economic crisis after the war of independence, a condition that lingered throughout much of the nineteenth century. However, in the early twentieth century, the arrival of Italian, Spanish, and Arab immigrants, together with the arrival of the railroad,

revived trade and agriculture, and the economy of the zone improved [3]. The Calchaqui Valleys (CV), 1700–3000-m altitude in the Andes, cross the provinces of Catamarca, Tucumán, Jujuy, and Salta. The Diaguitas inhabited this area in the pre-Hispanic era. There is little historical information concerning these populations either before or after contact with the Europeans. Therefore, the exact origin and/or degree of admixture of present CV inhabitants is unknown, but a large proportion of the original population probably disappeared due to their particular history: invasion of the Incas, European colonization, and the policy of estrangement of the rebels, from the sixteenth to the late seventeenth centuries [4]. The current population (approximately 36,000 inhabitants) [2] has a low density and is unequally distributed, with Cafayate (approx. 15,000 inhabitants) and Cachi (7000) as the most populated localities (Supplementary Figure 1). Therefore, the present inhabitants of both Salta city and Calchaqui Valleys can be considered “mestizo” populations, the result of a mixture between immigrants of diverse origins, African slaves introduced in the colonial period, and natives of Amerindian tribes through a complex process of conquest and colonization of north-western Argentina (NWA).

The X-chromosome was chosen since it has proved to be an efficient tool in population genetics [5] and forensic practice [6], especially in complex and deficiency cases. Although population data on X-chromosome markers in Argentinian

J. F. Ferragut and M. Bassitta contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-018-1970-x>) contains supplementary material, which is available to authorized users.

✉ J. F. Ferragut
jf.ferragut@uib.es

¹ Institut Universitari d’Investigació en Ciències de la Salut (IUNICS) i Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Carretera de Valldemossa, km 7.5, 07122 Palma de Mallorca, Illes Balears, Spain

² Cátedra de Antropología, Facultad de Humanidades, Universidad Nacional de Salta, Av. Bolivia 5150, 4400 Salta, Argentina

populations can be found in the literature [e.g., 7–10], this is a pioneer study on X-polymorphisms in the NWA region, focusing on the comprehensive analysis of 21 X-chromosomal markers of different types—Alu insertions and STRs—in two Salta province populations, aiming to (i) explore the variation of these polymorphic X-chromosome markers in the population of Salta city, as a reference of an urban population in NWA, and in the Calchaqui Valleys, as a rural population; (ii) correlate this genetic variation with the origin and migration history of these populations; and (iii) explore the usefulness of these markers in a forensic context.

Blood samples were obtained from 178 unrelated individuals (78 males and 100 females) from Salta province after informed consent: 73 living in different villages of the Calchaqui Valleys and 105 in Salta city. Samples were typed for two sets of X-chromosome genetic markers: (i) a set of nine X-chromosome Alu insertions (Ya5DP62, Yb8DP49, Yd3JX437, Yb8NBC634, Ya5DP77, Ya5NBC491, Yb8NBC578, Ya5DP4, and Ya5DP13) described by Callinan et al. [11] and (ii) 12 X-STRs included in the Investigator Argus X-12 kit (Qiagen GmbH, Hilden, Germany). Genotyping and statistical methods were performed as described in Ferragut et al. [12]. Proficiency testing of the Spanish and Portuguese Speakers Working Group of the International Society for Forensic Genetics (GHEP-ISFG, <https://ghep-isfg.org>) was conducted as a quality control. New variant alleles were sequenced by the Sanger method and aligned using the CodonCode Aligner program v.7.1.2 (CodonCode Corporation, Dedham, USA).

Allele frequency data in the studied populations (Salta city and Calchaqui Valleys) are included in Supplementary Tables 1–2. Exact test of population differentiation showed no significant differences between SC and CV. Values of average gene diversity across loci are summarized in Supplementary Tables 3–5. All analyzed loci were in the Hardy-Weinberg equilibrium after the Bonferroni correction for multiple tests ($p > 0.00238$) except for markers DXS10148 and Ya5DP62 in the Calchaqui Valleys population, due to a lack of heterozygotes. Five X-Alu insertions (Ya5DP62, Yb8DP49, Yd3JX437, Ya5DP77, and Ya5DP13) were revealed to be polymorphic in both populations, whereas the others appeared as monomorphic in at least one studied population: Ya5NBC491 and Yb8NBC578 for the insertion in both populations and Yb8NBC634 in CV; and for the absence of insertion, Ya5DP4 in the Calchaqui Valleys. Most polymorphic Alu elements showed moderate to low diversity. Ya5DP77 displayed the highest heterozygosity (0.461) and Ya5DP13 the lowest (0.020). These results agree with those found in other South American populations [8, 10, 13]. Average gene diversity for this set of X-chromosome Alu insertions was 0.157, an intermediate value between African and European populations and similar to Asian and Amerindian populations [11, 13]. All X-STRs were highly

polymorphic in all populations (average gene diversity was 0.783 in VC and 0.798 in SC). Locus-by-locus analyses revealed DXS10135 had the greatest diversity (with 26 alleles and heterozygosity of 0.931), while DXS8378 and DXS10103 were the least diverse markers (with mean heterozygosity lower than 0.676). No identical haplotype-like allelic combinations of the 12 X-STRs markers were found when typing the 78 males from Salta province, although a higher number of samples would better support these results. Linkage groups (LG) 1–4 revealed 67, 60, 39, and 59 haplotypes, respectively (Supplementary Table 6). Of all the haplotypes observed, 76.9% were observed in only one individual, the most common one in 11 males in LG3, displaying a frequency of 0.141. In these populations, LG1 proved to be the most polymorphic group and LG3 the least variable. Notably, the rural population (CV) had slightly lower values in average gene diversity over loci and also in haplotype diversity (0.511 and 0.964, respectively) than the urban SC population (0.525 and 0.990, respectively). Pairwise LD analysis, after the Bonferroni correction for multiple tests, showed significant associations between Yb8NBC634-Yd3JX437 and Yb8NBC634-Ya5DP77 Alu pairs and DXS10103-HPRTB, DXS10103-DXS10101, and DXS10101-HPRTB pairs of STRs (Supplementary Table 7). STR loci associations involved markers located in the same linkage group (LG3); significant LD has also been reported for the same pairs of STRs in other studies [e.g., 12, 14].

During X-STR profiling, new (previously undescribed) alleles were detected in DXS10079, DXS10134, DXS7132, and DXS10148 markers. Traditional Sanger sequencing was performed to identify the structure of these new variants (Supplementary Table 8). Allele 13 in DXS10079 displayed nine straight repeats of the motif AGAA, instead of the 10–21 repeats in the reference structure described by Hering et al. [15]. In DXS10134, alleles 36.1, 37.1, 38.1, and 39.1 were detected, revealing the sequence variations responsible for these intermediate alleles were all due to one additional nucleotide (A) located outside the core repeat in the immediate upstream flanking region when compared to the sequence in Edelman et al. [16]. Allele 39.1 has also been described in a West African sample set [17], but the structure was different: in Guinea-Bissau samples, the additional base (A) was inserted in the last block of GAAA repeats (see Supplementary Table 8). Sequencing methods usually reveal that alleles derived from conventional PCR-CE systems have the same lengths but different sequences, leading to propose a new system of STR allele nomenclature for massively parallel sequencing of forensic STRs [18]. DXS7132 is a locus with a simple tandem of CTAT repeats [19], with intermediate alleles (16.3, 17.3, and 18.3) with a single-base deletion interrupting the perfect repeat motif: (CTAT)_n-CAT-(CTAT)₂. These intermediate alleles have been reported only in individuals with South American origin; this population specificity suggests a

possible Amerindian origin of these alleles. In the present study, along with alleles 16.3 and 17.3, a previously undescribed allele with a fragment size corresponding to allele 13.3 was found. Sequencing data displayed the same deletion, but in a different repeat: $(CTAT)_7$ -**CAT**- $(CTAT)_6$ (Supplementary Table 8). A highly complex compound structure was revealed for DXS10148. Sequencing data of Salta individuals displayed four different structures. The first, in allele 18 and the new variant 28, was according to Hundertmark et al. [20]. In intermediate alleles 18.1, 24.1, and 27.1 and in one individual with allele 26.1, a single-base insertion (A) interrupted the AAGG repeats: $(GGAA)_4$ -**(AAGA)_n**-**A**-**(AAGA)**- $(AAAG)_4$ -N₈- $(AAGG)_2$ -AAAG-(AAGG-AAAG)₂-GGAAA. Sequencing of another individual with a fragment corresponding to allele 26.1 displayed a nucleotide between repeat motifs $(AAGA)_n$ and $(AAAG)_3$, like individuals with alleles 22.1 and 28.1. Finally, the presence of an additional copy of the AAGG-AAAG motif was detected in two samples: one individual with the common allele 31 and one with a new variant, with a fragment size corresponding to allele 25. Regarding Alu insertion in locus Ya5NBC49, fragments with different lengths were detected (Supplementary Figure 2), with 73% of individuals having a band of approximately 415 bp rather than the expected 435 bp band [11]. Sequencing results revealed that length variation was due to a deletion of 20 bases in the poly (A) tail.

Statistical parameters of forensic interest were calculated for each polymorphic marker and population for each set of markers and across the 21-marker set (Supplementary Tables 3–5). As expected, considering the different nature of the polymorphisms, Alu insertion polymorphisms were considerably less informative than STRs. This 21 X-loci set proved to be suitable and highly informative for forensic casework in both populations studied, with values of combined mean exclusion chance (MEC) exceeding 0.999999 for duos and 0.99999999 for trios. In both populations, the combined power of discrimination (PD) yielded values greater than 1 in $4.504 \text{ E} + 15$ in females, and 1 in $1.829 \text{ E} + 9$ in males (Supplementary Table 5). The rural population (CV) had slightly lower values in all forensic efficiency parameters than the urban population (SC), in accordance with genetic diversity values.

To evaluate the genetic relationship between Salta province and other populations, pairwise F_{ST} genetic distances were computed (Supplementary Tables 9 and 10) and represented in an MDS plot (Supplementary Figure 3) separately for the two types of polymorphisms, since no other populations were studied with the full set of 21 X-chromosome markers. Population clustering reveals the ability of both sets of X-markers to clearly discriminate between continents, with European, Amerindian, and African populations distantly positioned. Using available data for these markers, North African groups are positioned much closer to the Europeans

than to sub-Saharan populations. Regarding Amerindian populations, they show up on the left of the *x*-axis in the MDS plot for X-STRs but do not form a tight cluster, not even the ones from the same geographical region, suggesting differentiation by genetic drift in small and isolated populations [e.g., 10, 21]. For the X-Alu set of markers, only two Bolivian Amerindian populations were studied, showing close positioning. Salta and Calchaqui Valleys appear in the center of the plot, between Amerindian and European populations, for both sets of markers (Supplementary Figure 3), in accordance with the admixture from European and Native American ancestries of these populations. Using X-STR markers, a greater number of Amerindian populations were analyzed, allowing better definition. The Calchaqui Valleys were observed to stand closer to Amerindian populations while Salta city is closer to other Latin American populations.

Different admixture proportions from African, European, and Native American ancestries have contributed to the genetic heterogeneity found in Argentinean and Latin American populations in general. The highest Amerindian ancestry is found in areas with historically larger native populations, such as Andean regions and meso-America, where major pre-Columbian civilizations developed, whereas African ancestry is low in most of these populations [22, 23]. In Argentina, rural populations tend to show a greater predominance of the ancestral Native American substrate than urban ones, especially in maternal lineages [24–26], probably mirroring the strong sexual bias of the immigrants who settled during colonial times. Correspondingly, the X-polymorphism results in the present study show the CV population stands closer to the Amerindian cluster than SC, thereby suggesting rural populations from Salta province have maintained more important Native American ancestry in their genetic pool. Considering the positive correlation described between heterozygosity and European ancestry, the lower European component in CV, together with their rural and isolated origin, could explain the lower diversity values in CV versus the Salta urban population, as found in other Latin American populations [23].

The existence of a clear population stratification in Argentina [21, 27] emphasizes the need not to ignore this genetic heterogeneity in routine forensic casework and, hence, to develop local databases. This study provides a useful database of 21 X-chromosome markers, especially powerful for the X-STR set, for populations in Salta province (north-western Argentina). New detected variants are also described, confirming that some tandem repeat structures are more complex and variable than initially observed for this set of markers [28]. Results conclude both Salta populations harbor a high Native American ancestry component, as previously described for Argentina's north-western region [27, 29], despite their self-recognized European ancestry.

Acknowledgments We are grateful to Maria Trinidad Garcia (from the *Serveis Científicotècnics of the Universitat de les Illes Balears*) for her assistance and help with capillary electrophoresis procedures and Meryl Wyn Jones for the English language corrections. We would also like to sincerely thank all the people who volunteered to participate in this study.

Funding information This work was partially supported by grant AAEE196/2005 from the *Direcció General de R + D + I (Comunitat Autònoma de les Illes Balears)* and the European Regional Development Fund (ERDF).

Compliance with ethical standards

All participants provided written informed consent statements, approved by The Argentine Ministry of Health, the review boards at National University of Salta (Argentina) (PICTO-UNSa N° 36656) and the *Direcció General de R + D + I (Government of Balearic Islands, Spain)* (procedure AAEE196/2005). The anonymity of the recruited samples was preserved during the study.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Buliubasich C, González A (2009) Los Pueblos Indígenas de la Provincia de Salta. La posesión y el dominio de sus tierras. Cepiha, Salta
- INDEC (2010) Instituto Nacional de Estadística y Censos de la República Argentina. <http://www.indec.gov.ar/>. Accessed 1 Nov 2017
- Boleda M (1993) La población del Noroeste Argentino. Historia y actualidad. Editorial Legasa, Salta
- Lorandi AM, Boixadós R, Bunster C, Palermo MA (1997) El valle Calchaquí. In: Lorandi AM (ed) *El Tucumán colonial y Charcas I*. Facultad de Filosofía y Letras, Universidad de Buenos Aires, Buenos Aires, pp 205–252
- Schaffner SF (2004) The X chromosome in population genetics. *Nat Rev Genet* 5:43–51
- Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. *Int J Legal Med* 117:67–74
- Bobillo C, Sala A, Gusmão L, Corach D (2011) Genetic analysis of 10 X-STRs in Argentinian population. *Forensic Sci Int Genet* 5: e14–e16
- Resano M, Zanetti D, Esteban EM, Moral P (2016) X-chromosome Alu insertions in Bahía Blanca, Argentina: assessment of population information from varied genetic markers and usefulness of X-chromosome markers to trace sex-biased parental contributions. *Hum Biol* 88:275–286
- Caputo M, Amador MA, Santos S, Corach D (2017) Potential forensic use of a 33 X-InDel panel in the Argentinean population. *Int J Legal Med* 131:107–112
- Di Santo Meztler GP, del Palacio S, Esteban ME, Armoa I, Argüelles CF, Catanesi CI (2018) Genetic differentiation of north-east Argentina populations based on 30 binary X chromosome markers. *Front Genet* 9:1–9
- Callinan PA, Hedges DJ, Salem AH, Xing J, Walker JA, Garber RK, Watkins WS, Bamshad MJ, Jorde LB, Batzer MA (2003) Comprehensive analysis of Alu-associated diversity on the human sex chromosomes. *Gene* 317:103–110
- Ferragut JF, Bentayebi K, Pereira R, Castro JA, Amorim A, Ramon C, Picomell A (2017) Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs. *Forensic Sci Int Genet* 31:e5–e11
- Gayà-Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodríguez A, Villena M, Vasquez R, Moral P (2010) Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. *Am J Hum Biol* 22:154–162
- Nothnagel M, Szibor R, Vollrath O, Augustin C, Edelmann J, Geppert M, Alves C, Gusmão L, Vennemann M, Hou Y, Immel UD, Inturi S, Luo H, Lutz-Bonengel S, Robino C, Roewer L, Rolf B, Sanft J, Shin KJ, Sim JE, Wiegand P, Winkler C, Krawczak M, Hering S (2012) Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome. *Forensic Sci Int Genet* 6:778–784
- Hering S, Augustin C, Edelmann J, Heidel M, Dressler J, Rodig H, Kuhlisch E, Szibor R (2006) DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases. *Int J Legal Med* 120:337–345
- Edelmann J, Hering S, Augustin C, Szibor R (2008) Characterisation of the STR markers DXS10146, DXS10134 and DXS10147 located within a 79.1 kb region at Xq28. *Forensic Sci Int Genet* 2:41–46
- Gomes I, Brehm A, Gusmão L, Schneider PM (2016) New sequence variants detected at DXS10148, DXS10074 and DXS10134 loci. *Forensic Sci Int Genet* 20:112–116
- Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, Gusmão L, Hares DR, Irwin JA, King JL, Knijff P, Morling N, Prinz M, Schneider PM, Neste CV, Willuweit S, Phillips C (2016) Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Sci Int Genet* 22:54–63
- Gomes I, Pereira R, Mayr WR, Amorim A, Carracedo A, Gusmão L (2009) Evaluation of DXS9902, DXS7132, DXS6809, DXS7133, and DXS7423 in humans and chimpanzees: sequence variation, repeat structure, and nomenclature. *Int J Legal Med* 123: 403–412
- Hundertmark T, Hering S, Edelmann J, Augustin C, Plate I, Szibor R (2008) The STR cluster DXS10148–DXS8378–DXS10135 provides a powerful tool for X-chromosomal haplotyping at Xp22. *Int J Legal Med* 122:489–492
- Muñoz A, Albeza MV, Acreche N, Castro JA, Ramon MM, Picornell A (2012) Allele frequencies of 15 STRs in the Calchaqui Valleys population (north-western Argentina). *Forensic Sci Int Genet* 6:e58–e60
- Salzano FM, Bortolini MC (2002) *The evolution and genetics of Latin American populations*. Cambridge University Press, Cambridge
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A (2008) Geographic patterns of genome admixture in Latin American mestizos. *PLoS Genet* 4:e1000037
- Martínez VL, Bertoni B, Parra EJ, Bianchi NO (2004) Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. *Hum Biol* 76:543–557
- Pauro M, García A, Nores R, Demarchi DA (2013) Analysis of uniparental lineages in two villages of Santiago del Estero, Argentina, seat of pueblos de Indios in colonial times. *Hum Biol* 85:699–719

26. Cardoso S, Palencia L, Valverde L, Alfonso MA, Gómez L, Alfaro E, Bravi CM, Dipierri JE, Peña JA, Pancorbo MM (2013) Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina. *Forensic Sci Int Genet* 7:52–55
27. Catelli ML, Álvarez-Iglesias V, Gómez-Carballa A, Mosquera-Miguel A, Romanini C, Borosky A, Amigo J, Carracedo A, Vullo C, Salas A (2011) The impact of modern migrations on present-day multi-ethnic Argentina as recorded on the mitochondrial DNA genome. *BMC Genet* 12:77
28. Tillmar AO, Kling D, Butler JM, Parson W, Prinz M, Schneider PM, Egeland T, Gusmão L (2017) DNA Commission of the International Society for Forensic Genetics (ISFG): guidelines on the use of X-STRs in kinship analysis. *Forensic Sci Int Genet* 29: 269–275
29. Diaz-Lacava A, Walier M, Penacino G, Wienker TF, Baur MP (2011) Spatial assessment of Argentinean genetic admixture with geographical information systems. *Forensic Sci Int Genet* 5:297–302