RESEARCH ARTICLE

WILEY — *American Journal of* PHYSICAL ANTHROPOLOGY — *The Official Journal of the American Association of Physical Anthropologists*

Ancient DNA reveals temporal population structure within the South-Central Andes area

M. Gabriela Russo[1] | **Fanny Mendisco[2]** | **Sergio A. Avena[1,3]** |
Cristian M. Crespo[1] | **Valeria Arencibia[4]** | **Cristina B. Dejean[3,4]** | **Verónica Seldes[5]**

[1]Universidad Maimónides, CONICET, Equipo de Antropología Biológica, Departamento de Cs. Naturales y Antropológicas, CEBBAD, Buenos Aires, C1405BCK, Argentina

[2]Laboratory of Molecular Anthropology and Image Synthesis (AMIS), University Paul Sabatier (Toulouse III), Faculté de Médecine, CNRS, Toulouse, UMR 5288, France

[3]UBA, Sección de Antropología Biológica, ICA, FFyL, Buenos Aires, C1406CQJ, Argentina

[4]Universidad Maimónides, Equipo de Antropología Biológica, Departamento de Cs. Naturales y Antropológicas, CEBBAD, Buenos Aires, C1405BCK, Argentina

[5]UBA, CONICET, Instituto Interdisciplinario Tilcara, Centro Universitario Tilcara, FFyL, Tilcara, Jujuy, Y4624AFI, Argentina

Correspondence
María Gabriela Russo, Equipo de Antropología Biológica, CEBBAD, Hidalgo 775, Ciudad Autónoma de Buenos Aires, CP 1405, Argentina.
Email: russo.mariagabriela@maimonides.edu

Present address
Fanny Mendisco, University of Bordeaux, UMR 5199 PACEA, Equipe Anthropologie des Populations Passées et Présentes, Allée Geoffroy ST Hilaire, 33615 Pessac Cedex, France

Funding information
Agencia Nacional de Promoción Científica y Tecnológica, Grant Number: PICT 2014-3012; Fundación Científica Felipe Fiorellino, Grant Number: Subsidio intramuros 2017; Fundación de Historia Natural Félix de Azara, Grant Number: Subsidio intramuros 2017

Abstract

Objectives: The main aim of this work was to contribute to the knowledge of pre-Hispanic genetic variation and population structure among the South-central Andes Area by studying individuals from Quebrada de Humahuaca, North-western (NW) Argentina.

Materials and methods: We analyzed 15 autosomal STRs in 19 individuals from several archaeological sites in Quebrada de Humahuaca, belonging to the Regional Developments Period (900–1430 AD). Compiling autosomal, mitochondrial, and Y-chromosome data, we evaluated population structure and differentiation among eight South-central Andean groups from the current territories of NW Argentina and Peru.

Results: Autosomal data revealed a structuring of the analyzed populations into two clusters which seemed to represent different temporalities in the Andean pre-Hispanic history: pre-Inca and Inca. All pre-Inca samples fell into the same cluster despite being from the two different territories of NW Argentina and Peru. Also, they were systematically differentiated from the Peruvian Inca group. These results were mostly confirmed by mitochondrial and Y-chromosome analyses. We mainly found a clearly different haplotype composition between clusters.

Discussion: Population structure in South America has been mostly studied on current native groups, mainly showing a west-to-east differentiation between the Andean and lowland regions. Here we demonstrated that genetic population differentiation preceded the European contact and might have been more complex than thought, being found within the South-central Andes Area. Moreover, divergence among temporally different populations might be reflecting socio-political changes occurred in the evermore complex pre-Hispanic Andean societies.

KEYWORDS
Andean groups, autosomal STRs, NW Argentina, pre-Hispanic populations

1 | INTRODUCTION

Genetic studies of Native American populations have become widely used to unraveling their evolutionary history, especially when the main focus is put on the peopling process (e.g., Achilli et al., 2013; Llamas et al., 2016; O'Rourke & Raff, 2010; Reich et al., 2012; Tamm et al.,

2007). Considering that America was the last continent populated by human groups and probably as a result of genetic drift, Native Americans exhibit lower levels of genetic variation compared to populations in other continents (e.g., Wang et al., 2007). Moreover, since the first groups followed a north-to-south migration route, a clinal pattern of genetic diversity appeared. Not only lower heterozygosity levels were

found in South American native populations (Wang et al., 2007), but also the settlement in the Southern Cone was accompanied by the loss of maternal lineages, for example (e.g., Crespo, Russo, Hajduk, Lanata, & Dejean, 2017; de la Fuente et al., 2015).

However, numerous studies have demonstrated that isolation, settlement, and diversification across the different regions of South America allowed the emergence of new genetic variants (e.g., de Saint Pierre et al., 2012; García, Pauro, Nores, Bravi, & Demarchi, 2012; Wang et al., 2007) and high levels of differentiation among populations (Wang et al., 2007). Current South American groups showed strong population structuring when studied from distinct genetic approaches such as mitochondrial DNA (mtDNA) (Rothhammer, Fehren-Schmitz, Puddu, & Capriles, 2017), genome-wide autosomal STRs (Wang et al., 2007), and a combination of several markers (Yang et al., 2010). These studies revealed a diversification across a west-to-east cline, mainly differentiating the Andean populations from the eastern groups, although structure patterns might have been more complicated than previously thought principally in the Andes (Cabana et al., 2014).

Albeit the great usefulness, studies on current South American populations have limitations related to the complex demographic processes occurred since European colonization. First, even the presumably most isolated native groups might exhibit at least some level of admixture with European and African populations (e.g., Crossetti et al., 2008; Reich et al., 2012), which strongly affects the allele frequencies introducing new variants. Also, the European colonization produced severe social and demographic changes leading to a population decrease (e.g., O'Fallon & Fehren-Schmitz, 2011) and the resulting loss of genetic variability. These processes and their consequences over allele frequencies call attention to the precautions that must be taken when inferring past events by studying only current human groups (Pickrell & Reich, 2014) and emphasize the need for increasing ancient DNA (aDNA) studies.

Genetic analyses of pre-Hispanic South American groups represent one of the most useful tools for understanding population dynamics and complementing archaeological studies. Most aDNA studies focused on mtDNA due to its advantage for extraction and PCR amplification in comparison with nuclear DNA, mainly caused by the difference in genome copy number (Pakendorf & Stoneking, 2005). Nevertheless, population analyses gain resolution power when several molecular markers are combined, such as mitochondrial and Y-chromosome lineages along with autosomal markers.

In this study, we present autosomal STR data of pre-Hispanic Andean populations from the Quebrada de Humahuaca valley in North-western (NW) Argentina belonging to the Regional Developments Period (RDP) (1050–520 years BP). During this period, these human groups experienced significant social changes. From the previous small, relatively egalitarian societies, they went through a concentration process into large settlements throughout the central ravine (Nielsen, 2001). Also, they were actively involved in an increasingly long-distance exchange of goods from the Pacific coast to the eastern valleys and plains (Nielsen, 2013). The end of this period is determined by the Incas arrival and the concomitant socio-

political changes, including the relocation of entire local groups in order to exercise control over them (Williams, 2000; Williams, Villegas, Gheggi, & Chaparro, 2005). Later, around 414 yBP the Inca rule was interrupted by the Spaniard conquerors arrival to the current NW Argentina (Bárcena, 2007) and the consequent strong socio-political disruption.

The main aim of this work was to contribute to the knowledge of genetic variation among native South American populations by means of analyzing autosomal genetic markers on pre-Hispanic individuals from Quebrada de Humahuaca (NW Argentina) and comparing them with other South-central Andean groups. Moreover, we aimed to investigate whether socio-political changes among regions and temporalities might have been reflected in population structure.

2 | MATERIALS AND METHODS

2.1 | Archaeological sample

We analyzed 19 individuals from seven archaeological sites in Quebrada de Humahuaca (Jujuy province, NW Argentina) (Figure 1). Based on archaeological context and radiocarbon dating it was established that all individuals belonged to the RDP (Table 1). These samples were included in a previous study in which mitochondrial and Y-chromosome data was obtained (Mendisco et al., 2014).

2.2 | DNA extraction and autosomal STR amplification

From each individual, at least two teeth were analyzed independently. First, teeth were washed with a bleach solution, rinsed with ultra-pure water (Milli-Q®, Millipore) and finally ultraviolet (UV) irradiated for 15 min each side to remove surface exogenous DNA contaminants. Then, the powder was obtained through total milling in liquid nitrogen (6870 SamplePrep Freezer Mill®, Fisher Bioblock, Illkirch, France) and DNA extraction was performed using 200 mg of powder following a previously described protocol (Mendisco et al., 2011). Two independent extractions were done for every tooth. Extracts were quantified through real-time PCR using Quantifiler™ Human Identification kit (Applied Biosystems) on the ABI PRISM® 7000 (Applied Biosystems) detection system, following manufacturer's recommendations.

From each of the four available extracts, PCR amplifications of autosomal STRs were carried out systematically. Following a previously published protocol (Russo, Mendisco, Avena, Dejean, & Seldes, 2016), a total of 15 autosomal STRs (D2S1338, D3S1358, vWA, FGA, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, TH01, CSFPO, and TPOX) were analyzed with AmpFlSTR® Identifiler™ kit (Applied Biosystems). We also used AmpFlSTR® MiniFiler™ kit (Applied Biosystems) which amplified only nine of them but in shorter amplicons. This kit allows a more effective amplification of degraded samples because while with Identifiler™ six *loci* are analyzed with amplicons larger than 250 bp, all loci evaluated with MiniFiler™ are amplified in shorter fragments (only one up to 280 bp). The combined strategy enabled not only a better genotyping of large STRs but also confirming the results. Both kits also allowed amplification of

RUSSO ET AL.

American Journal of
PHYSICAL
ANTHROPOLOGY
The Official Journal of the American Association of Physical Anthropologists

WILEY | 3
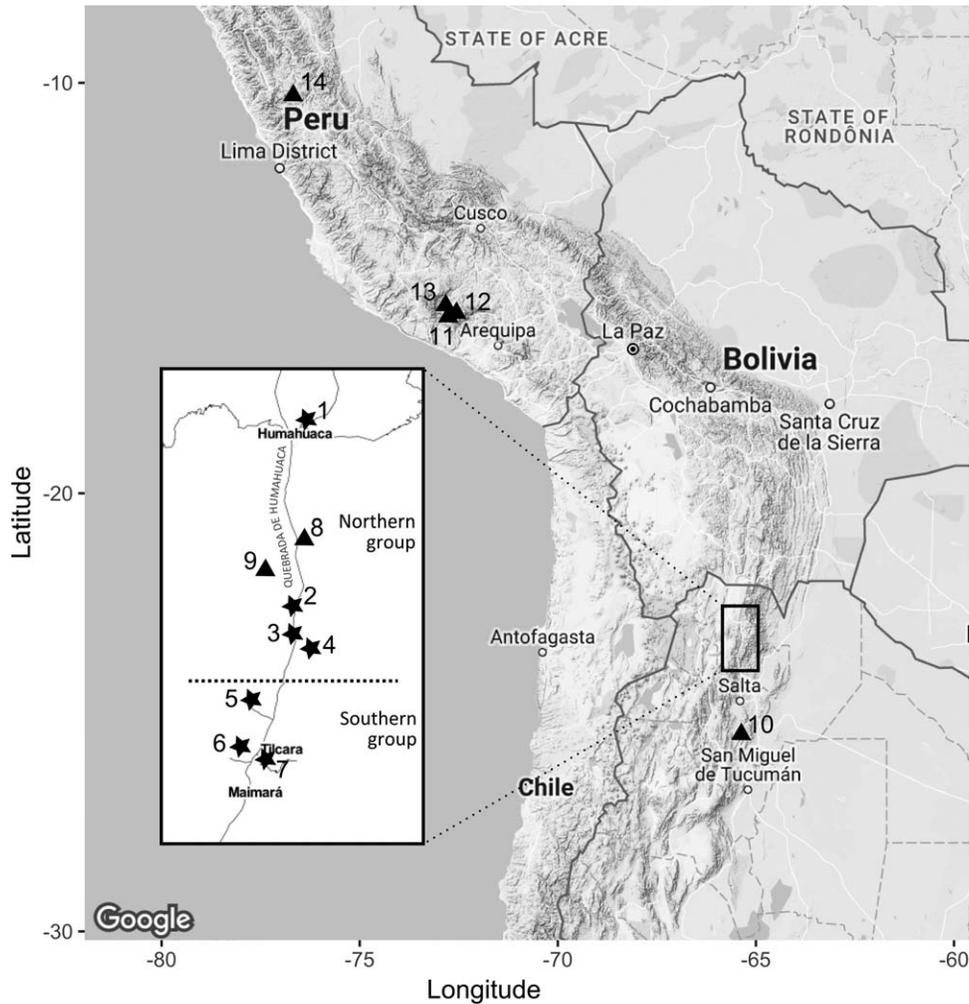
FIGURE 1 Location of the archaeological sites analyzed in this study (stars) and the South-central Andean populations used for comparisons (triangles). The punctuated line indicates the north-south subdivision of Quebrada de Humahuaca.1: Peñas Blancas, 2: San José, 3: Huacalera, 4: Banda de Perchel, 5: Juella, 6: Sarahuaico, 7: Tilcara, 8: Muyuna, 9: Los Amarillos, 10: Las Pirguas, 11: Tompullo 2, 12: Puca, 13: Acchaymarca, 14: Lauricocha. Map constructed from the obtained with the R package *ggmap* (Kahle & Wickham, 2013)

the Amelogenin gene region used for sex determination. A consensus genotype was obtained for each individual from at least four PCR products.

2.3 | Authenticity criteria

All analyses were conducted at laboratories exclusively dedicated to aDNA studies following strict protocols to prevent contamination (Cooper & Poinar, 2000; Gilbert, Bandelt, Hofreiter, & Barnes, 2005). Both pre- and post-PCR working areas were physically isolated. Sample processing was done wearing protective disposable clothes, sterile gloves, and face masks. All materials used during analysis were sterilized by autoclave and a long UV exposure. Extraction and amplification blanks were used as negative controls in each step. Finally, all results were replicated at different times, from multiple extracts and amplifications of the same sample.

For one individual we had the possibility of analyzing both upper jaw and jaw teeth (sample TIL5). Given that in funerary contexts they are often disjointed, this situation was an opportunity to confirm that

the upper jaw and jaw belonged to the same individual and also to improve the authenticity of the results.

2.4 | Database compilation

Analysis of autosomal markers from aDNA extracts represents a challenging task. To date, only a few studies had included autosomal STR amplification of ancient samples across the South-central Andes Area. In the current territory of Peru, three sites from the Inca Period (Baca, Molak, Sobczyk, Węgleński, & Stankovic, 2014), and a rock shelter with human remains from the Pre-Ceramic and Formative Periods (Fehren-Schmitz et al., 2015) have autosomal STR data available.

The remaining autosomal genetic data came from archaeological sites in NW Argentina. On one hand, an early study (Carnese et al., 2010) analyzed some STRs markers in individuals buried during the Formative Period in Las Pirguas site (Salta province). On the other hand, in two RDP sites from Quebrada de Humahuaca STR amplification has been accomplished: Muyuna (Russo et al., 2016) and Los Amarillos (Mendisco et al., 2018).

4 | WILEY · *American Journal of* PHYSICAL ANTHROPOLOGY
The Official Journal of the American Association of Physical Anthropologists

RUSSO ET AL.

TABLE 1 Archaeological sites from Quebrada de Humahuaca (NW Argentina) analyzed in the present study

| Archaeological site | Radiocarbon dates | | n |
	yBP	cal. AD	
Banda de Perchel	850 ± 70	1036–1281[a]	3
Huacalera	n.d.	n.d.	1
San José	889 ± 57	1020–1271[b]	1
Peñas Blancas	n.d.	n.d.	1
Juella	635 ± 140	1066–1613[b]	5
Sarahuaico	690 ± 80; 730 ± 70	1164–1413[b]	2
Tilcara	940 ± 60	989–1222[b]	6
Total			19

All sites dates belong to the Regional Developments Period.
[a]Rivolta, 2007; [b]Nielsen, 2001. *n.d.*: not determined. *n*: sample size.

Altogether, we compiled autosomal STR data of 91 individuals from pre-Hispanic South-central Andean populations (Figure 1 and Table 2).

2.5 | Population analyses for autosomal markers

Before running population analyses, we excluded individuals with first-degree relatives (parent, offspring, and/or full siblings) in the same archaeological site. Pairwise kinship relationships were investigated using Familias3 (Kling, Tillmar, & Egeland, 2014) following a previous study (Russo et al., 2016). This software allows to calculate posterior probabilities for specific family relationships based on Maximum Likelihood estimations. We found several pairs with a relatively high posterior probabilities for parent–offspring and sibling relationships in Tompullo 2 and Puca (Baca et al., 2014), and confirmed previous kinship estimations for Muyuna (Russo et al., 2016) and Los Amarillos (Mendisco et al., 2018) (Table S1 in Supporting Information). Considering the results, eight individuals were excluded for population analyses (Table 2 and Table S1 in Supporting Information).

Each archaeological site with more than five individuals was defined as a distinctive population. Except for the Lauricocha samples (Fehren-Schmitz et al., 2015) that were considered one population despite being only four individuals because of their geographical and chronological differentiation, the sites with less than five individuals came from Quebrada de Humahuaca (including the ones analyzed in this study) and were assigned to either a Northern or Southern group. We took the so-called Angosto de Perchel (23°29'24.5" S, 65°21'47.5" W, approximately) as the limit of the north-south subdivision (Figure 1). Therefore, a total of eight pre-Hispanic South-central Andean populations were considered for analyses: Northern Quebrada, Los Amarillos, Southern Quebrada, Las Pirguas, Tompullo 2, Acchaymarca, Puca and Lauricocha (Figure 1 and Table 2).

We first evaluated population differentiation through AMOVA using Arlequin v. 3.5.2.2 (Excoffier & Lischer, 2010). Pairwise F_{ST} were calculated based on the sum of squared differences (R_{ST}-like), linearized with Slatkin's method (Slatkin, 1995), and graphically displayed using the R (R Core Team, 2017) extension. We chose 10,000 permutations for significance. For these analyses, neither Las Pirguas nor Lauricocha was included due to the former's low number of genotyped STRs and the small sample size of the later. We used a total of 12 STRs since D3S1358, D2S1338 and D19S433 were not available for Tompullo 2, Acchaymarca and Puca populations (Baca et al., 2014). A hierarchical structure was tested by grouping the populations into two regions: the current territories of NW Argentina and Peru.

TABLE 2 Pre-Hispanic South-central Andean populations analyzed in the present study

Current region	Population group	n	Included archaeological sites	Period (radiocarbon dates in yBP)	References
NW Argentina	Northern Quebrada	6 (7)	Banda de Perchel, Huacalera, San José,	RDP (900 ± 40 to 850 ± 70)[a,b]	This study
			Muyuna		Russo et al., 2016
	Los Amarillos	13 (17)	Los Amarillos	RDP (915 ± 85 to 505 ± 50)[a]	Mendisco et al., 2018
	Southern Quebrada	9	Juella, Sarahuaico, Tilcara	RDP (940 ± 60 to 635 ± 140)[a]	This study
	Las Pirguas	18	Las Pirguas	Formative (1310 ± 40)	Carnese et al., 2010
Peru	Tompullo 2	14 (16)	Tompullo 2	Late horizon (474–416)	Baca et al., 2012
	Acchaymarca	11	Acchaymarca	LIP to late horizon (950–416)	Baca et al., 2014
	Puca	8 (9)	Puca	LIP to late horizon (950–416)	Baca et al., 2014
	Lauricocha	4	Lauricocha cave	Pre-ceramic to ceramic (7871 ± 30 to 3337 ± 22)	Fehren-Schmitz et al., 2015
	Total	83 (91)			

n: sample size. Parenthesis indicates total n without excluding close relatives. *RDP*: Regional Developments Period. *LIP*: Late Intermediate Period.
[a]Nielsen, 2001. [b]Rivolta, 2007.
All NW Argentinean and Lauricocha samples dated to pre-Inca moments.

RUSSO ET AL.

American Journal of
PHYSICAL
ANTHROPOLOGY
The Official Journal of the American Association of Physical Anthropologists

WILEY | 5
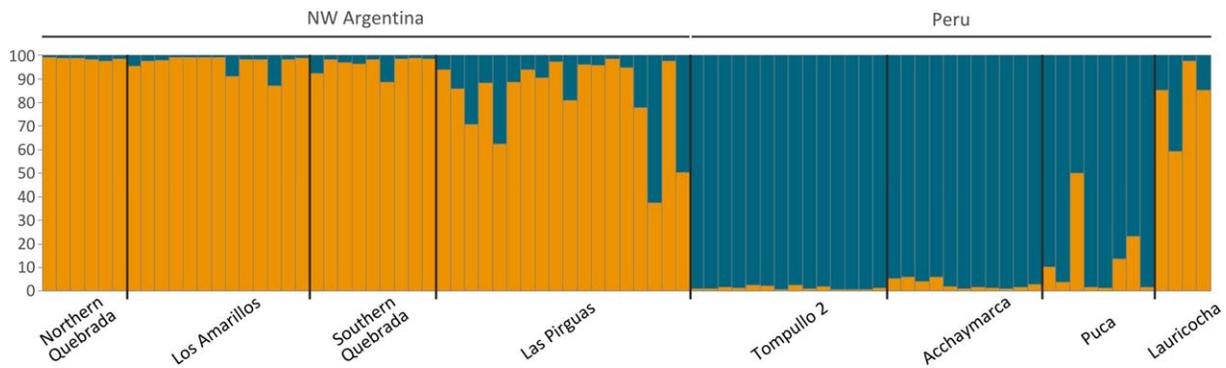
FIGURE 2 Structure analysis of the pre-Hispanic South-central Andean populations analyzed in this study

To further test population structure within the South-central Andes Area, we employed a Bayesian approach implemented in STRUCTURE v. 2.3.4 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). Default *priors* were chosen except for *alpha*, which value was set to 0.5 following the 1/K recommendation of Wang (2017). We ran 1,000,000 MCMC steps with a burn-in period of 100,000. The number of panmictic clusters was estimated from five replicates for each K using the Evanno method (Evanno, Regnaut, & Goudet, 2005) available in Structure Harvester (Earl & von Holdt, 2012). Final individual Q-values were obtained using CLUMPP v. 1.1.2 (Jakobsson & Rosenberg, 2007), available at Mobyle SNAP Workbench portal (Monacell & Carbone, 2014).

Finally, genetic distances between individuals were calculated and used for a Principal Coordinates Analysis (PCoA) in GenAlEx v. 6.5 (Peakall & Smouse, 2012). GenAlEx was also used to test for sex-biased dispersal in the analyzed populations by calculating the corrected mean assignment index (AIc) (Favre, Balloux, Goudet, & Perrin, 1997).

2.6 | Comparison with uniparental markers

Most individuals from the analyzed populations had already been typified for mtDNA (HVR-I sequencing) and Y-chromosome (STRs) in previous studies (Baca, Doan, Sobczyk, Stankovic, & Węgleński, 2012; Baca et al., 2014; Carnese et al., 2010; Fehren-Schmitz et al., 2015; Mendisco et al., 2011, 2014; Russo et al., 2016). Using the available information, we tested whether uniparental lineages could also account for population structuring in the South-central Andes.

The populations that could be analyzed in such fashion varied depending on whether mitochondrial or Y-chromosome data was available. Whenever possible, we tried to sustain the same grouping proposed for autosomal STR data, but the differential availability of uniparental markers made it difficult. Nevertheless, it was possible to group the pre-Hispanic Andean individuals analyzed in this study into nearly the same populations used for autosomal data (Table S2 in Supporting Information).

AMOVA was again employed using Arlequin to test for population differentiation. For mitochondrial sequences, Tamura and Nei (1993) substitution model was chosen using MEGA6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) for Φ_{ST} calculation, and the hierarchical population structuring into two regions (current territories of Peru and Argentina) was tested. For Y-chromosome STRs pairwise F_{ST} were

calculated based on the sum of squared differences (R_{ST}-like). In both cases, a Slatkin's linearization of F_{ST} was employed and we run 10,000 permutations for significance. Because of the small sample sizes, Lauricocha was excluded from these analyses and the individuals from Los Amarillos were included in the Northern Quebrada group based on geographical proximity (Table S2 in Supporting Information).

Additionally, we evaluate haplotype sharing between clusters by constructing median-joining networks (Bandelt, Forster, & Röhl, 1999) for both maternal and paternal lineages using Network v. 5.0.0.1. For mitochondrial data, site weights were assigned following Soares et al. (2009) mutation rates, polycytosine-associated mutations at positions 16182 and 16183 were excluded, and a parsimonious post-processing (Polzin & Daneschmand, 2003) was applied. For Y-chromosome lineages, the median-joining network was constructed after having processed the data with the reduced-median method (Bandelt, Forster, Sykes, & Richards, 1995) and *loci* weights were set proportional to the inverse of their variance.

Because *locus* DYS385a/b is duplicated in the Y-chromosome, it was excluded from analyses due to the impossibility of assigning each allele to the proper duplicate.

3 | RESULTS

3.1 | Nuclear aDNA recovering

Genotyping of autosomal markers and sex determination was accomplished for 12 of the 19 analyzed individuals from Quebrada de Humahuaca (Table S3 in Supporting Information). This 63% rate is a successful result considering the intrinsic difficulty of recovering nuclear DNA from ancient samples. Further, complete profiles of the 15 STR *loci* were obtained for eight individuals. Among the remaining four individuals, three had a 6.67% of missing data (only one *locus*), and one 20% (three *loci*), allowing their inclusion in statistical analyses.

These rates of autosomal genotyping and missing data are similar to the obtained in previous studies on ancient Andean samples. Among the reference populations compiled in this study, for Tompullo 2 16 out of 25 (64%) analyzed individuals were genotyped and a maximum of 8.33% missing data (one *locus*) was found in only three of them (Baca et al., 2012). In Acchaymarca, genotyping of 11 of 22 (50%) individuals was accomplished, with 16.67% of missing data (two *loci*) in

6 | **WILEY** *American Journal of* **PHYSICAL ANTHROPOLOGY** *The Official Journal of the American Association of Physical Anthropologists*

RUSSO ET AL.
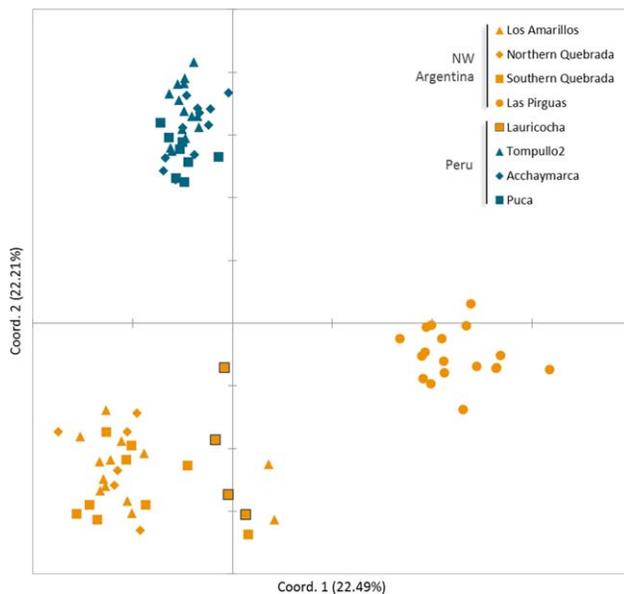
FIGURE 3 Principal coordinates analysis (PCoA) based on individual genetic distances obtained with autosomal STRs data. Percentage of variance explained by each coordinate is shown in parenthesis. Colors were assigned according to the two clusters discovered with structure

one individual and 8.33% (one *locus*) in other (Baca et al., 2014). A lower genotyping rate was obtained for Puca site: 9 of 37 (24%) (Baca et al., 2014). But again, most individuals had no missing data since only one had a 16.67% rate (two *loci*) and three 8.33% (one *locus*). Out of five Lauricocha samples, four reached autosomal genotyping (80%) with no missing data (Fehren-Schmitz et al., 2015). Finally, 17 out of 21 individuals (81%) from Las Pirguas were genotyped, but in this case, a higher missing data rate was found (between 11% and 44%) and almost all individuals had at least one *locus* without data (Carnese et al., 2010).

3.2 | Population structure in pre-Hispanic South-Central Andes

Considering the data obtained in this study, the compiled database, and the excluded individuals based on kinship relationships (Table S1 in

Supporting Information), a total of 83 individuals from eight pre-Hispanic South-central Andean populations (Figure 1 and Table 2) were included in the STRUCTURE analysis. The estimation of *K* gave an optimum number of two clusters (Figure S1 in Supporting Information). This structure reflected the assumed regional division into NW Argentinean or Peruvian with the only exception of the grouping of Lauricocha samples (from current Peru) with NW Argentina populations (Figure 2).

The signal of structuring into two clusters (NW Argentina + Lauricocha and Peruvian Inca) was supported by the PCoA analysis. Genetic distances were minimal among the Peruvian Inca populations showing a strong differentiation from the remaining samples while individuals from Quebrada de Humahuaca and Lauricocha were clearly grouped (Figure 3).

This analysis also revealed a differentiation of Las Pirguas population from the remaining samples including those presumably belonging to the same cluster (Figure 3). Nevertheless, this result could be an effect of the lower available information for this site since only nine autosomal STRs were typified (Carnese et al., 2010), which could also be the reason for the poorer assignment of these individuals into one of the clusters discovered by STRUCTURE (Figure 2). Despite the higher levels of missing data, excluding Las Pirguas population from the analysis did not change the differentiation between the NW Argentina + Lauricocha and the Peruvian Inca groups (Figure S2 in Supporting Information).

The structure tested by the hierarchical AMOVA supported the differentiation between NW Argentina and Peruvian populations. Even though variation between regions was not significant ($p = 0.103$), genetic differentiation was found among populations ($p = 0.047$; Table S4 in Supporting Information) and specifically, strong levels among populations from different regions (Figure 4a and Table S5 in Supporting Information).

This pre-Hispanic population structure in the South-central Andes Area revealed by autosomal STRs was supported by uniparental markers, although with some considerations. First, AMOVA based on Y-chromosome data indicates strong variation among populations ($p < 0.00001$; Table S6 in Supporting Information) and pairwise comparisons showed a clear differentiation between the Peruvian site Tompullo 2 and all NW Argentinean groups (Figure 4b and Table S7 in Supporting Information).
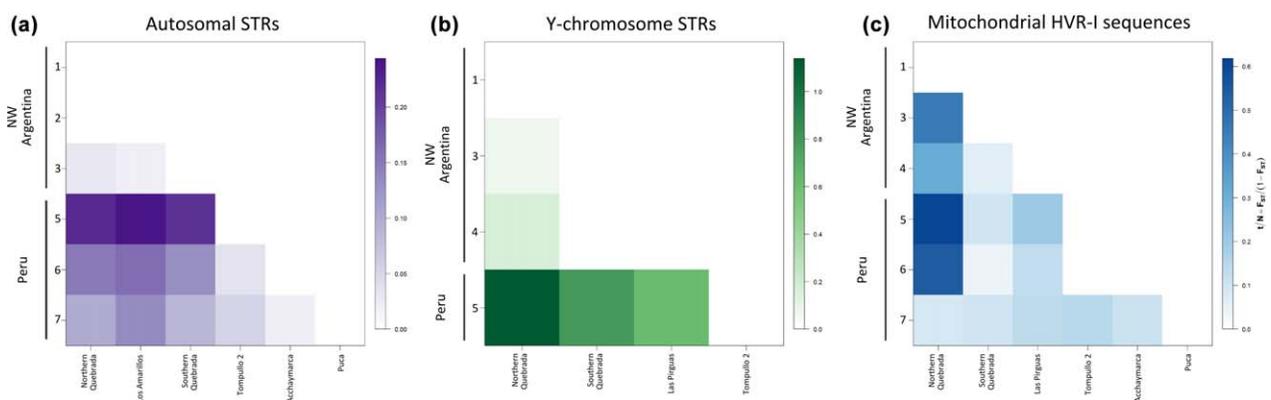
FIGURE 4 Pairwise population differentiation (linearized F_{ST}) for autosomal (a), Y-chromosome (b), and mitochondrial (c) available data. 1: Northern Quebrada, 2: Los Amarillos, 3: Southern Quebrada, 4: Las Pirguas, 5: Tompullo 2, 6: Acchaymarca, 7: Puca

RUSSO ET AL.

American Journal of
PHYSICAL
ANTHROPOLOGY
The Official Journal of the American Association of Physical Anthropologists

WILEY | 7
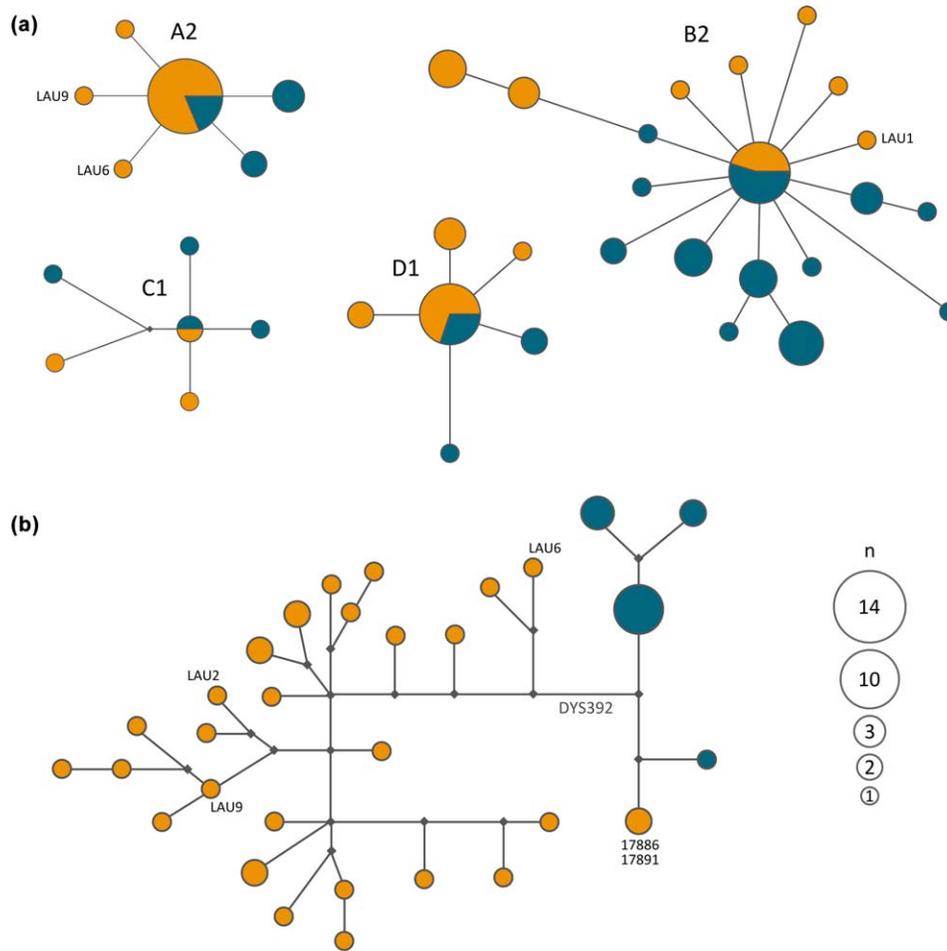
FIGURE 5 Median-joining networks for uniparental lineages of the pre-Hispanic South-central Andean populations analyzed in this study. Colors reflect the clustering resulted from STRUCTURE analysis. (a) Most parsimonious haplotype trees for mitochondrial haplogroups. (b) Y-chromosome haplotypes. *n*: sample size. Mutations are shown below branches

Second, considering mitochondrial HVR I sequences, stronger levels of population differentiation were found through AMOVA (Table S8 in Supporting Information). Although variation between regions was not significant ($p = 0.302$), high levels of differentiation among populations were found ($p = 0.0001$). Nevertheless, in this case, pairwise differences did not show a clear structure pattern since some Φ_{ST} were high even between populations from the same cluster or low between populations from different clusters (Figure 4c and Table S9 in Supporting Information).

Despite this AMOVA result, haplotype composition was clearly different between clusters. For all mitochondrial haplogroups, the two clusters did not share haplotypes besides the nodal lineages, while most derived haplotypes (particularly within A2, B2, and D1) were shared among at least two and up to six individuals of the same cluster (Figure 5a).

Finally, Y-chromosome haplotypes network also showed the divergence between the two clusters having no haplotypes shared and consistent with autosomal-based results, Lauricocha individuals were again linked to the NW Argentina samples (Figure 5b). This divergence could be caused mainly by allelic differences at DYS392 *locus*. All individuals of the Peruvian Inca cluster had allele 16 at this system (Baca et al.,

2014), while NW Argentinean had the allele 14 (Mendisco et al., 2014) along with most Lauricocha samples (Fehren-Schmitz et al., 2015). Only one individual from Lauricocha (LAU6) was an exception, carrying the allele 15 (Fehren-Schmitz et al., 2015). NW Argentinean samples 17886 and 17891 from Las Pirguas site (Carnese et al., 2010) lacked information for this *locus* which could be one of the reasons of their position on the network.

4 | DISCUSSION

The results presented in this study can be reasonably considered authentic given the various precautions followed during the analysis. Also, the developed strategy combining the analysis of various genetic markers was extremely valuable. In addition, (i) all consensus genotypes were determined from many replications (from different extracts and amplifications performed at different times and from different samples); (ii) genetic profiles originating from the researchers who directly handled the samples were never observed during the analyses; (iii) negative controls used confirmed that there was no exogenous or crossed contamination; (iv) results obtained from different markers were all consistent. Moreover, it is unlikely that the pattern of genetic variation

and the phylogeographic coherence found systematically using mitochondrial, Y-chromosome, and autosomal data could be explained by contamination. Finally, missing data seemed not to influence the results mainly because (a) most individuals had a very low percentage of missing information, and (b) excluding the population with the highest rate of missing data did not change the results (Figure S2 in Supporting Information).

Genetic variation among Native Americans is relatively small in comparison with other human populations from the rest of the major continents. Nevertheless, they present strong structuring signals even within South America (e.g., Wang et al., 2007), the last subcontinent populated by humans. Most of the available information for South America came from studies on current native groups, but patterns of genetic variation and structuring might have been strongly affected after European colonization and so might not reflect the actual pre-Hispanic situation. Therefore, a more precise characterization of native groups could be achieved by aDNA analysis.

In the present study, we presented autosomal STRs data of individuals recovered from several archaeological sites in Quebrada de Humahuaca, NW Argentina, contributing to the knowledge of genetic variation of pre-Hispanic South-central Andean populations. Combining the data here obtained with the available in the literature (Table 2) we found a genetic structure across this area, mainly differentiating the pre-Hispanic groups of NW Argentina from those of the current Peru, with the exception of the most ancient Peruvian samples which were linked to the NW Argentinean cluster (Figure 2). This result has several considerable implications.

First, by analyzing aDNA we showed that genetic population differentiation in native South-central Andean groups preceded the European contact. Also, this study is in concordance with previous analysis of population structure in the Andes showing a much more complex pattern than the assumed west-to-east differentiation (Cabana et al., 2014). Particularly, chronological differences between clusters could explain the found structure. All the analyzed NW Argentinean and Lauricocha samples were linked into the same cluster and systematically differentiated from the remaining Peruvian populations (Figure 2–4a). Interestingly, the populations from this Peruvian cluster belonged to the Inca Period (Baca et al. 2014), while all NW Argentinean samples and Lauricocha individuals dated from pre-Inca moments (Carnese et al., 2010; Fehren-Schmitz et al., 2015; Mendisco et al., 2018; Russo et al., 2016).

Taking this into consideration, it is possible to call the NW Argentina and Lauricocha the "ancestral" cluster being the Peruvian Inca populations the "derived." On one hand, this is in concordance with a common origin for pre-Hispanic Andean populations of current territories of Argentina and Peru, with a posterior regional differentiation and demographic increase resulting from the emergence of complex societies (e.g., Nielsen, 2001). Later, genetic differentiation could have prevailed among some populations despite their unification under the Inca State (Williams et al., 2005), favoring the establishment of particular derived variants. Further, in the current Peru, temporal genetic discontinuity among pre-Hispanic populations have been discovered using mtDNA data, although also associated with climate fluctuations (Fehren-Schmitz et al., 2014).

On the other hand, it is possible that the analyzed Peruvian Inca populations have had a different origin compared to the ancient Peruvian and NW Argentinean samples, probably as a result of the resettlement of entire groups among different regions, which was a common Inca practice during the Empire development (Williams, 2000; Williams et al., 2005). In every case, our results showed that pre-Hispanic population structure patterns might be reflecting socio-political changes not only among regions but also among temporalities.

The temporal population structure was mostly supported by uniparental markers. Based on Y-chromosome STRs, we found high levels of differentiation between the NW Argentinean populations and the individuals from Tompullo 2 (Figure 4b). Nevertheless, it must be considered that this was the only Peruvian site that could be included in the analyses. Therefore, differentiation might be the result of the singularity of this small, isolated population (Baca et al., 2014), since considering autosomal STRs it was significantly differentiated even from the populations of the same cluster (Table S5 in Supporting Information). Despite the limitations, the Y-chromosome haplotypes network also supported the population structure and reinforced the link between NW Argentinean samples and Lauricocha individuals (Figure 5b).

Considering maternal lineages, a higher amount of genetic differentiation was found since even populations from the same cluster exhibited significant differences (Figure 4c). The highest Φ_{ST} values were found between the Northern Quebrada de Humahuaca group and most of the remaining populations. This differentiation could be thought as the result of highest levels of endogamy or isolation for the Northern Quebrada populations especially in Los Amarillos site, where most individuals shared mitochondrial haplotypes (Mendisco et al., 2011). However, analyzing autosomal and Y-chromosome STRs, no such differentiation was found (Figure 4a,b). Thus, another explanation could be proposed for the different pattern of population differentiation found with mtDNA.

It is possible that different patterns of sex-biased dispersal were acting on diverse pre-Hispanic Andean populations. To test whether sex-biased dispersal could explain the results; we calculate the corrected mean assignment index (AIc) (Favre, Balloux, Goudet, & Perrin, 1997) for the analyzed groups (data not shown). Although not significant, negative AIc values were found in male individuals from most groups, including Quebrada de Humahuaca, whereas for the Peruvian sites Tompullo 2 and Acchaymarca, males had a positive AIc. This means that males could have been the most dispersive sex in most of the analyzed pre-Hispanic populations explaining the higher levels of differentiation found with mtDNA (Figure 4c). On the other hand, some groups like Tompullo 2 could have had a patrilocal residence pattern being the females the most dispersive sex and confirming previous estimations (Baca et al., 2014). This is also in concordance with the higher amount of differentiation considering Y-chromosome STRs between Tompullo 2 and the rest of the analyzed groups (Figure 4b). Nevertheless, all these results must be taken with caution due to the low availability of autosomal data for Andean archaeological sites. Clearly, the study of pre-Hispanic residence patterns continues to be an intriguing subject for which increasing aDNA information could make significant contributions.

RUSSO ET AL.

American Journal of
PHYSICAL
ANTHROPOLOGY
The Official Journal of the American Association of Physical Anthropologists

WILEY | 9

Finally, the nearly null mitochondrial haplotype sharing between the two clusters also supported the temporal structure found. Moreover, derived haplotypes were only shared among individuals of the same cluster particularly within B2 haplogroup (Figure 5a).

Because we are probably witnessing a temporal effect causing the differentiation revealed in this study, it becomes necessary to increase the autosomal genetic data for ancient Andean populations in order to further investigate pre-Hispanic population structure. Also, we cannot discard a sampling effect due to the small sample sizes of the analyzed populations and the few archaeological sites with available genetic and specifically, autosomal data.

The importance of increasing studies like the present lies not only in improving the knowledge of the pre-Hispanic genetic variability but also in the need of proper reference populations for studies of kinship and residence patterns. Given the genetic complexity exposed by this kind of studies, it is not possible to assume a single residence pattern, an only cause of population structuring, or a general population dynamic that can account for all the diverse pre-Hispanic societies across the South-central Andes.

ACKNOWLEDGMENTS

AUTHOR'S CONTRIBUTIONS

Study design: MGR and VS; Data collection: FM; Data analysis: MGR; Results discussion: MGR, VS, CBD, SAA, CMC and VA; Manuscript preparation: MGR, VS, CBD, SAA, CMC, VA and FM.

ORCID

M. Gabriela Russo http://orcid.org/0000-0002-5727-4956

REFERENCES

Achilli, A., Perego, U. A., Lancioni, H., Olivieri, A., Gandini, F., Hooshiar Kashani, B., ... Torroni, A. (2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proceedings of National Academy of Sciences of the United States of America*, 110, 14308–14313.

Baca, M., Doan, K., Sobczyk, M., Stankovic, A., & Węgleński, P. (2012). Ancient DNA reveals kinship burial patterns of a pre-Columbian Andean community. *BMC Genetics*, 13, 30.

Baca, M., Molak, M., Sobczyk, M., Węgleński, P., & Stankovic, A. (2014). Locals, resettlers, and pilgrims: A genetic portrait of three pre-Columbian Andean populations. *American Journal of Physical Anthropology*, 154, 402–412.

Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37–48.

Bandelt, H.-J., Forster, P., Sykes, B. C., & Richards, M. B. (1995). Mitochondrial portraits of human populations. *Genetics*, 141, 743–753.

Bárcena, J. R. (2007). El Período Inka en el Centro-Oeste y Noroeste argentino: Aspectos cronológicos en el marco de la dominación del Kollasuyu. In V. I. Williams, B. N. Ventura, A. B. M. Callegari & H. D. Yacobaccio (Eds.), *Sociedades precolombinas surandinas: Temporalidad, interacción y dinámica cultural del NOA en el ámbito de los andes centro-sur* (pp. 251–281). Buenos Aires: Taller Internacional de Arqueología del NOA y Andes Centro Sur.

Cabana, G. S., Lewis, C. M., Jr., Tito, R. Y., Covey, R. A., Cáceres, A. M., Cruz, A. F., ... Stone, A. C. (2014). Population genetic structure of traditional populations in the Peruvian Central Andes and implications for South American population history. *Human Biology*, 86, 147–165.

Carnese, F., Mendisco, F., Keyser, C., Dejean, C. B., Dugoujon, J. M., Bravi, C. M., ... Crubézy, E. (2010). Paleogenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina). *American Journal of Physical Anthropology*, 141, 452–462.

Cooper, A., & Poinar, H. (2000). Ancient DNA: Do it right or not at all. *Science (New York, N.Y.)*, 289, 1139.

Crespo, C. M., Russo, M. G., Hajduk, A., Lanata, J. L., & Dejean, C. B. (2017). Variabilidad mitocondrial en muestras pre-colombinas de la Patagonia Argentina. Hacia una visión de su poblamiento desde el ADN antiguo. *Revista Argentina De Antropología Biológica*, 19, 21.

Crossetti, S. G., Demarchi, D. A., Raimann, P. E., Salzano, F. M., Hutz, M. H., & Callegari-Jacques, S. M. (2008). Autosomal STR genetic variability in the Gran Chaco native population: Homogeneity or heterogeneity?. *American Journal of Human Biology: The Official Journal of the Human Biology Council*, 20, 704–711.

de la Fuente, C., Galimany, J., Kemp, B. M., Judd, K., Reyes, O., & Moraga, M. (2015). Ancient marine hunter-gatherers from Patagonia and Tierra Del Fuego: Diversity and differentiation using uniparentally inherited genetic markers. *American Journal of Physical Anthropology*, 158, 719–729.

de Saint Pierre, M., Gandini, F., Perego, U. A., Bodner, M., Gómez-Carballa, A., Corach, D., ... Olivieri, A. (2012). Arrival of Paleo-Indians to the southern cone of South America: New clues from mitogenomes. *PLoS One*, 7, e51311.

Earl, D. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure: Extensions to linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.

Favre, L., Balloux, F., Goudet, J., & Perrin, N. (1997). Female-biased dispersal in the monogamous mammal Crocidura russula: Evidence from field data and microsatellite patterns. *Proceedings of the Royal Society of London, Biological Series B*, 264, 127–132.

Fehren-Schmitz, L., Haak, W., Mächtle, B., Masch, F., Llamas, B., Tomasto Cagigao, E., ... Reindel, M. (2014). Climate change underlies global demographic, genetic, and cultural transitions in pre-Columbian southern Peru. *Proceedings of National Academy of Sciences of the United States of America*, 111, 9443–9448.

Fehren-Schmitz, L., Llamas, B., Lindauer, S., Tomasto-Cagigao, E., Kuzminsky, S., Rohland, N., ... Haak, W. (2015). A re-appraisal of the early andean human remains from Lauricocha in Peru. *PLoS One*, 10, e0127141.

García, A., Pauro, M., Nores, R., Bravi, C. M., & Demarchi, D. A. (2012). Phylogeography of mitochondrial haplogroup D1: An early spread of

subhaplogroup D1j from Central Argentina. *American Journal of Physical Anthropology*, 149, 583–590.

Gilbert, M. T. P., Bandelt, H. J., Hofreiter, M., & Barnes, I. (2005). Assessing ancient DNA studies. *Trends in Ecology and Evolution*, 20, 541–544.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.

Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144–161.

Kling, D., Tillmar, A. O., & Egeland, T. (2014). Familias 3—Extensions and new functionality. *Forensic Science International: Genetics*, 13, 121–127.

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., ... Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, 2, e1501385.

Mendisco, F., Keyser, C., Hollard, C., Seldes, V., Nielsen, A., Crubézy, E., & Ludes, B. (2011). Application of the iPLEXTM Gold SNP genotyping method to the analysis of Amerindian ancient DNA samples: Benefits for ancient population-based studies. *Electrophoresis*, 32, 386–393.

Mendisco, F., Keyser, C., Seldes, V., Nielsen, A. E., Russo, M. G., Crubézy, E., & Ludes, B. An insight into the burial practices of the late pre-Hispanic Los Amarillos community (northwestern Argentina) through the study of ancient DNA. *Journal of Archaeological Science*, 91, 12–19.

Mendisco, F., Keyser, C., Seldes, V., Rivolta, C., Mercolli, P., Cruz, P., ... Ludes, B. (2014). Genetic diversity of a late prehispanic group of the Quebrada de Humahuaca, Northwestern Argentina. *Annals of Human Genetics*, 78, 367–380.

Monacell, J. T., & Carbone, I. (2014). Mobyle SNAP Workbench: A web-based analysis portal for population genetics and evolutionary genomics. *Bioinformatics*, 30, 1488–1490.

Nielsen, A. E. (2001). Evolución social en Quebrada de Humahuaca (AD 700–1536). In E. Berberián & A. E. Nielsen (Eds.), *Historia Argentina prehispánica* (pp. 171–264). Córdoba: Editorial Brujas.

Nielsen, A. E. (2013). Circulating objects and the constitution of South Andean society (500 BC – AD 1550). In K. G. Hirth & J. Pillsbury (Eds.), *Merchants, markets, and exchange in the pre-Columbian world* (pp. 389–418). Washington D.C.: Dumbarton Oaks.

O'Fallon, B. D., & Fehren-Schmitz, L. (2011). Native Americans experienced a strong population bottleneck coincident with European contact. *Proceedings of National Academy of Sciences of the United States of America*, 108, 20444–20448.

O'Rourke, D. H., & Raff, J. A. (2010). The human genetic history of the Americas: The final frontier. *Current Biology*, 20, R202–R207.

Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annual Review of Genomics and Human Genetics*, 6, 165–183.

Pickrell, J., & Reich, D. (2014). Towards a new history and geography of human genes informed by ancient DNA. *Trends in Genetics*, 30, 377–389.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics*, 28, 2537–2539.

Polzin, T., & Daneschmand, S. V. (2003). On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters*, 31, 12–20.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: URL https://www.R-project.org/.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., ... Ruiz Linares, A. (2012). Reconstructing Native American population history. *Nature*, 488, 370–374.

Rivolta, M. C. (2007). Abandono y reutilización de sitios. La problemática de los contextos habitacionales en Quebrada de Humahuaca. *Estudios Atacameños: Arqueología y Antropología Surandinas*, 34, 31–49.

Rothhammer, F., Fehren-Schmitz, L., Puddu, G., & Capriles, J. (2017). Mitochondrial DNA haplogroup variation of contemporary mixed South Americans reveals prehistoric displacements linked to archaeologically derived culture history. *American Journal of Human Biology*, 29(6), e23029.

Russo, M. G., Mendisco, F., Avena, S., Dejean, C. B., & Seldes, V. (2016). Pre-Hispanic mortuary practices in Quebrada de Humahuaca (Northwestern Argentina): Genetic relatedness among individuals buried in the same grave. *Annals of Human Genetics*, 80, 210–220.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139, 457–462.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., ... Richards, M. B. (2009). Correcting for purifying selection: An improved human mitochondrial molecular clock. *American Journal of Human Genetics*, 84, 740–759.

Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., ... Malhi, R. S. (2007). Beringian standstill and spread of native American founders. *PLoS One*, 2, e829.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10, 512–526.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729.

Wang, J. (2017). The computer program structure for assigning individuals to populations: Easy to use but easier to misuse. *Molecular Ecology Resources*, 17, 981. https://doi.org/10.1111/1755-0998.12650.

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., ... Ruiz-Linares, A. (2007). Genetic variation and population structure in native Americans. *PLoS Genetics*, 3, e185.

Williams, V. (2000). El imperio Inka en la provincia de Catamarca. *Intersecciones En Antropología*, 1, 55–78.

Williams, V., Villegas, M. P., Gheggi, M. S., & Chaparro, M. G. (2005). Hospitalidad e intercambio en los valles mesotermales del Noroeste Argentino. *Boletín De Arqueología PUCP*, 9, 335–372.

Yang, N. N., Mazières, S., Bravi, C., Ray, N., Wang, S., Burley, M. W., ... Ruiz-Linares, A. (2010). Contrasting patterns of nuclear and mtDNA diversity in Native American populations. *Annals of Human Genetics*, 74, 525–538.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.