

mtDNA and Y-Chromosome Diversity in Aymaras and Quechuas From Bolivia: Different Stories and Special Genetic Traits of the Andean Altiplano Populations

Magdalena Gayà-Vidal,^{1,2} Pedro Moral,¹ Nancy Saenz-Ruales,² Pascale Gerbault,² Laure Tonasso,² Mercedes Villena,³ René Vasquez,⁴ Claudio M. Bravi,⁵ and Jean-Michel Dugoujon^{2*}

¹Unitat d'Antropologia, Biologia Animal, Universitat de Barcelona, 08028, Spain

²Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), FRE 2960 CNRS, Université de Toulouse III Paul Sabatier, Toulouse 31000, France

³Instituto Boliviano de Biología de Altura (IBBA), Universidad Mayor de San Andrés, La Paz, Bolivia

⁴Instituto Boliviano de Biología de Altura (IBBA), Universidad Autónoma Tomás Frías, Potosí, Bolivia

⁵Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), La Plata 1900, Argentina

KEY WORDS uniparental markers; human populations; native Americans; Andean region

ABSTRACT Two Bolivian samples belonging to the two main Andean linguistic groups (Aymaras and Quechuas) were studied for mtDNA and Y-chromosome uniparental markers to evaluate sex-specific differences and give new insights into the demographic processes of the Andean region. mtDNA-coding polymorphisms, HVI–HVII control regions, 17 Y-STRs, and three SNPs were typed in two well-defined populations with adequate size samples. The two Bolivian samples showed more genetic differences for the mtDNA than for the Y-chromosome. For the mtDNA, 81% of Aymaras and 61% of Quechuas presented haplogroup B2. Native American Y-chromosomes were found in 97% of Aymaras (89% hg Q1a3a and 11% hg Q1a3*) and 78% of Quechuas (100% hg Q1a3a). Our data revealed high diversity values in the two populations, in

agreement with other Andean studies. The comparisons with the available literature for both sets of markers indicated that the central Andean area is relatively homogeneous. For mtDNA, the Aymaras seemed to have been more isolated throughout time, maintaining their genetic characteristics, while the Quechuas have been more permeable to the incorporation of female foreigners and Peruvian influences. On the other hand, male mobility would have been widespread across the Andean region according to the homogeneity found in the area. Particular genetic characteristics presented by both samples support a past common origin of the Altiplano populations in the ancient Aymara territory, with independent, although related histories, with Peruvian (Quechuas) populations. *Am J Phys Anthropol* 000:000–000, 2011. © 2011 Wiley-Liss, Inc.

The present population of the Andean region in Bolivia is the result of complex processes over thousands of years. It was in the central Andes (Andean Altiplano and current Peru) where the first complex societies and civilizations in South America emerged (Chavin, 900–200 BC, Tiwanaku, 100 BC–1200 AD, Huari (700–1200 AD) as well as the first state; the Inca Empire that was conquered by the Spaniards around 1532 AD (Stanish, 2001). Specifically, in the south central Andes (southern Peru, Bolivian Altiplano, north Chile, and northwest Argentina), the Tiwanaku civilization, originating in the Titicaca basin, extended its influence over the south central Andes (Kolata, 1993). After the Tiwanaku collapse, the state fragmented into a number of Aymara polities or *Señorios* (Bouysse-Cassagne, 1986) that persisted until their conquest by the Inca Empire (1300–1532 AD) when they became grouped within the Kollasuyu Inca region. From Cuzco, the Incas expanded their power toward the north and south using strategies such as language imposition (Quechua) and the *mitma* system (a deliberate movement of whole tribes from region to region around their vast Empire).

Linguistically, two main groups are present in the Andean area, the Quechuas (10 million speakers in Ecuador, Peru, southern Bolivia, and northern Chile) and the Aymaras (around 2.5 million of speakers, mainly in Bolivia). Before the Inca period, it is likely that an ancestral form of Quechua (technically referred to as

proto-Quechua) was spoken in the Huari distribution area (around current Ayacucho, Peru), whereas a proto-Aymara, together with Pukina and Uru, was probably

Additional supporting information may be found in the online version of this article.

Grant sponsors: The work of M. GV has been financed by a FI (2007FIB00108) and a BE grants from the “Departament d'Educació i Universitats de la Generalitat de Catalunya i el Fons Social Europeu”. It benefited from funds of Centre National de la Recherche Scientifique (CNRS) and Université Paul Sabatier Toulouse III, France (J.-M.D.), and the CGL2008-03955 project of the Spanish Ministerio de Educación y Ciencia (P.M.).

*Correspondence to: Jean-Michel Dugoujon, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), FRE 2960 CNRS, Université de Toulouse III Paul Sabatier, 37 Allées Jules Guesde, Faculté de Médecine, 31073 Toulouse cedex 3, Toulouse, France. Tel: 05 61 14 59 87. Fax: +33(0)5 61 14 59 79. E-mail: dugoujon@cict.fr

Present address of Pascale Gerbault: Molecular and Cultural Evolution Laboratory, University College London (UCL), London, UK.

Received 25 August 2010; accepted 9 December 2010

DOI 10.1002/ajpa.21487

Published online in Wiley Online Library (wileyonlinelibrary.com).

spoken in the influence area of the Tiwanaku civilization (Kolata, 1993; Browman, 1994; Stanish, 2001). Afterward, the Incas spread the Quechua tongue and imposed it as the official language of the empire, which was subsequently promoted by the Spaniards as *lingua franca* (Rowe, 1963).

This study explores the genetic variability and the genetic relationships of two Bolivian populations belonging to the two main Andean linguistic groups (Aymaras and Quechuas) through the analysis of uniparental markers; mtDNA and Y-chromosome. A previous study (Gayà-Vidal et al., 2010) examined a total number of 32 polymorphic Alu insertions (PAIs) in these two samples. According to these autosomal and X-chromosome data, the two Bolivian populations showed a similar genetic structure and were significantly close to each other when they were compared to Peruvian Quechua-speakers from Tayacaja and Arequipa. This suggested that the arrival of the Quechua language into Bolivia was more likely the result of a cultural spread rather than a demographic expansion. Nevertheless, have there been different population histories according to gender?

The aim of this study is to evaluate sex-specific differences by analyzing maternal (mtDNA) and paternal (Y-chromosome) uniparental markers in these two populations to gain new insights into the relationships between these two linguistic groups in Bolivia and into the demographic processes that have shaped the current Bolivian populations. The existence of more extensive data for uniparental markers than for biparental PAIs will allow us to achieve more robust interpretations. The majority of the central Andean populations studied so far for the mtDNA control region (CR) are located in Peru (Fuselli et al., 2003; Lewis et al., 2005, 2007), but these studies only considered the HVI region. Also, several samples from northwest Argentina (Alvarez-Iglesias et al., 2007; Tamm et al., 2007) and Bolivia (Corella et al., 2007; Afonso Costa et al., 2010; Barbieri et al., 2011) have been studied. As for Bolivian samples, the samples from Corella et al. (2007) and Barbieri et al. (2011) were only studied for the HVI region. Additionally, those from Corella et al. (2007) corresponded to 10 Aymaras and 19 Quechuas that migrated from the highlands to the lowlands in the Beni department of Bolivia, so their original location is imprecise. On the other hand, Afonso Costa et al. (2010) studied a sample from La Paz with a remarkable sample size (106), but it was an urban sample, and thus, individuals may have different origins. As for previous Y-chromosome variation studies, Andean data come mainly from Peru and the Andean area of Argentina (Bianchi et al., 1998; Tarazona-Santos et al., 2001; Iannacone et al., 2005; Toscanini et al., 2008; Blanco Vereá et al., 2010). Two Bolivian samples available (Lee et al., 2007) were described as Highlanders (from the Andean Altiplano) and Lowlanders (a mix of migrants from the Altiplano and natives from the Beni department), but SNPs were not analyzed to confirm the Native American haplogroups.

In this context, the present study (i) increases the number of Andean samples studied for both types of markers, providing haplogroup and haplotype data, (ii) covers the Bolivian Altiplano region, an area with a particular history in the Andean region, and (iii) provides data from two well-defined population samples with large sample sizes. Thus, it will allow us to obtain a



Fig. 1. Location of the populations included in the analyses. Circles, squares, and triangles indicate populations included in Y-chromosome, mtDNA, and both mtDNA and Y-chromosome analyses, respectively, the two samples of this study.

more accurate understanding of the genetic relationships in the Andean region.

SUBJECTS AND METHODS

Population samples

Blood samples from two Native American Bolivian samples, Aymara-speakers from the Titicaca Lake area and Quechua-speakers from the northern Potosi department, a region that was Aymara-speaking before the Inca expansion (Tschopik, 1963; see Fig. 1), were collected with informed consent by the Instituto Boliviano de Biología de Altura (IBBA), with approval from the Ethical Committee of this institution. From the available genealogical records, a total of 189 (93 Quechuas and 96 Aymaras) and 114 (55 Quechuas and 59 Aymaras) unrelated individuals were analyzed for mtDNA and Y-chromosome, respectively. A more detailed description of these populations can be found in Gayà-Vidal et al. (2010).

mtDNA polymorphisms

A mtDNA segment including the HVS-I and most of the HVS-II mtDNA regions was amplified by polymerase chain reaction (PCR), using the primer pair F15973 and R296, and PCR conditions as described in Coudray et al. (2009). DNA purification was undertaken using QIAquick PCR purification Kit (QIAGEN, Courtaboeuf,

France). Both strands were sequenced with the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and run in an ABI PRISM 3730 sequencer (PE, Applied Biosystems). In cases of samples with C-stretch, typical of haplogroup B2, both strands were sequenced twice. The sequences were checked manually with Sequencing Analysis (ABI Prism v3.7) and Chromas 2.13 software. The sequences were aligned and compared to the revised Cambridge Reference Sequence rCRS (Anderson et al., 1981; Andrews et al., 1999) using the Bioedit software (Tom Hall, Carlsbad, USA). To discard as many sequence artifacts as possible (Bandelt et al., 2001, 2002), each chromatogram was revised at least three times, sequences having unusual mutations were resequenced, and data entry and edition revised several times.

Coding region polymorphisms were typed to classify mtDNA into the four major Amerindian haplogroups (A–D) (Torroni et al., 1992). These four haplogroups were defined by restriction size or length polymorphisms (A: *Hae*III np663, B: 9bp-deletion, C: *Hinc*II np13259, and D: *Alu*I np5176). The primers and the PCR conditions were previously described in Mazieres et al. (2008). Genotype determinations were performed through 3% agarose gel electrophoresis and ethidium bromide staining. In addition, the SNP 6473T was typed to determine the B2 haplogroup, characteristic of Native Americans, and haplogroups A2, C1, and D1 were assigned according to the CR (Bandelt et al., 2003; Tamm et al., 2007).

Y-chromosome polymorphisms

A total of 20 markers were determined: 17 Y-chromosome short tandem repeat (Y-STR) polymorphic loci (DYS456, DYS389i/ii, DYS390, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATAH4, DYS437, DYS438, and DYS448) and three SNPs (M242, M3, and M346).

The Y-STRs were analyzed according to AmpFISTR® Yfiler™ PCR Amplification Kit (Applied Biosystems), using 1–2 ng of template DNA. The determinations were carried out in ABI 3730 with Genescan® and Genotyper® Analysis Softwares. Allele assignments were based on comparisons with the allelic ladders included in the kit using Genemapper software (Applied Biosystems). The quality of the determinations was assessed using the commercial allelic ladder and the DNA control supplied by Applied Biosystems.

To identify individuals carrying a Native American haplogroup, three biallelic markers were typed: (i) the polymorphic *C* → *T* transition (marker M242) that defines the haplogroup Q, present in Asia and America (Jobling and Tyler-Smith, 2003; Seielstad et al., 2003), using the methodology of Cinnioglu et al. (2004); (ii) the *C* → *T* transition (marker M3) in the DYS199 locus (Underhill et al., 1996), which defines the Q1a3a haplogroup, a lineage falling within the haplogroup Q, restricted to the Americas, and reaching a frequency of 100% in some populations; and (iii) for the Y-chromosomes possessing the M242 mutation, but not the M3, we sequenced the M346 *C* → *T* marker, downstream to M242 and upstream to M3 (Karafet et al., 2008). As in Bailliet et al. (2009), these chromosomes were considered to belong to the paragroup Q1a3*. To assign the most probable haplogroup to the non-Q samples and confirm the Q samples, we used Haplogroup Predictor ([http://](http://www.hprg.com/hapest5/)

www.hprg.com/hapest5/) that assigns the most probable haplogroup from the Y-STR profiles.

Data analysis

For mtDNA analyses, we considered the fragment between the 16,017 and 249 positions, according to the rCRS (Anderson et al., 1981; Andrews et al., 1999). Haplogroups were assigned following criteria described in the literature (Torroni et al., 1992; Bandelt et al., 2003). Haplogroup and haplotype frequencies were calculated by direct counting. Various diversity indices were computed. To determine the genetic relationships between haplotypes found in the two samples, Median-Joining (MJ) networks (Bandelt et al., 1999) were constructed for each haplogroup. For haplogroup B, positions 16,182 and 16,183 were not considered, because they are dependent on the presence of C at site 16189 (Pfeiffer et al., 1999). Following the suggestions of Bandelt et al. (2000), higher weights were assigned to the least variable polymorphisms and lower weights to the more hypervariable sites in our data set.

As for the Y-chromosome, haplogroup and haplotype frequencies were calculated by direct counting. Taking into account only the individuals belonging to the Native American Q haplogroup (lineages Q1a3* and Q1a3a), various diversity indices were computed. MJ networks (Bandelt et al., 1999) were built with the MP postprocessing option (Polzin and Daneschmand, 2003) for the Q1a3a hg. STRs were given weights that were inversely proportional to their allele size variances.

Exact tests of population differentiation (Raymond and Rousset, 1995) were performed to detect whether significant differences in mtDNA haplogroup frequencies and in Y-STR allele frequencies existed between the two Bolivian samples. In addition, the two study populations were compared for both sets of markers using F_{st} indices, as a measure of population differentiation.

For comparative purposes, mtDNA data from 51 South American samples (Table 1) were collected from the literature on the basis of available sequences for the HVI region and a minimum sample size of nine individuals. Their geographical location is shown in Figure 1. For the analyses, only Native American haplogroups were considered. The comparisons were based on the HVI region between 16,051 and 16,362 positions. The San Martin de Pangoa sample (Fuselli et al., 2003) was not included, because it is composed of both Quechua and Nematsiguenga speakers. The Cayapa sample included was from Rickards et al. (1999), because Tamm et al. (2007) did not maintain the proportions of the haplogroups, because their focus was phylogeny. Additionally, the HVII CR (from position 73 to 249) was available for 22 of the 51 samples (Table 1), and the CR between positions 16,024 and 249 was available for 6 of the 51 samples.

Therefore, analyses were carried out considering the four sets of data separately: the HVI (53 samples), HVII (24 samples), HVI–HVII (24 samples) CRs, and the HVI, HVII and the intervening region, from now on designated as CR (eight samples). Haplotype (*h*) and nucleotide (π) diversity within groups were calculated using Nei's formulas (1987). Analyses of molecular variance (AMOVA) (Excoffier et al., 1992) and hierarchical AMOVA analyses under geographical criteria (also for haplogroup frequencies) were performed. Genetic distances between samples were estimated using the Tamura–Nei distance method (Tamura and Nei, 1993) with the α

TABLE 1. Populations included in the comparisons for the mtDNA and for the Y-STRs

mtDNA comparisons			Y-Chromosome comparisons		
Populations ^a	N ^b	References ^c	Populations	N ^e	References
<i>Aymara</i>	96	Present study	<i>Aymara</i> **	57	Present study
<i>Quechua</i>	93	Present study	<i>Quechua</i> **	45	Present study
<i>Ignaciano</i>	15/22	Bert et al., 2004, Bert et al., 2001	<i>Colla</i> **	10	Toscanini et al., 2008
<i>Movima</i>	12/22	Bert et al., 2004, Bert et al., 2001	<i>Kaingang_Guarani</i> **	27	Leite et al., 2008
<i>Trinitario</i>	12/35	Bert et al., 2004, Bert et al., 2001	<i>Kichwa</i> **	72	González-Andrade et al., 2007
<i>Yucarare</i>	15/20	Bert et al., 2004, Bert et al., 2001	<i>Peru</i> **	51	Iannacone et al., 2005
<i>Quechua Beni</i>	19/32	Corella et al., 2007/Bert et al., 2001	<i>Toba</i> **	44	Toscanini et al., 2008
<i>Aymara Beni</i>	10/33	Corella et al., 2007/Bert et al., 2001	<i>Trinitario</i> **	34	Tirado et al., 2009
<i>Chimane</i>	10/41	Corella et al., 2007/Bert et al., 2001	<i>Chimane</i> **	10	Tirado et al., 2009
<i>Moseten</i>	10/20	Corella et al., 2007/Bert et al., 2001	<i>Mojeño</i> **	10	Tirado et al., 2009
<i>Ancash Quechua</i>	33/33	Lewis et al., 2005	<i>Kolla</i> **	12	Blanco-Verea et al., 2010
<i>Aymara Puno</i>	14	Lewis et al., 2007	<i>Diaguita</i> **	9	Blanco-Verea et al., 2010
<i>Quechua Puno</i>	30	Lewis et al., 2007	<i>Mapuche</i> **	23	Blanco-Verea et al., 2010
<i>Jaquaru Tupe</i>	16	Lewis et al., 2007	<i>Bari</i> *	16	YHRD:YA003358 ^f
<i>Yungay Quechua</i>	36	Lewis et al., 2007	<i>Yanomami</i> *	11	YHRD:YA002906 ^f
<i>Arequipa Quechua</i>	22	Fuselli et al., 2003	<i>Yukpa</i> *	12	YHRD:YA003360 ^f
<i>Tayacaja Quechua</i>	61	Fuselli et al., 2003	<i>Cayapa</i>	26	Tarazona-Santos et al., 2001
<i>Toba Chaco</i>	43/67 ^d	Cabana et al., 2006	<i>Tayacaja Quechua</i>	44	Tarazona-Santos et al., 2001
<i>Wichi Chaco</i>	32/99 ^d	Cabana et al., 2006	<i>Arequipa Quechua</i>	15	Tarazona-Santos et al., 2001
<i>Pilaga Formosa</i>	38	Cabana et al., 2006	<i>Gaviao-Zoro-Surui</i>	34	Tarazona-Santos et al., 2001
<i>Toba Formosa</i>	24 ^d	Cabana et al., 2006	<i>Karitiana</i>	8	Tarazona-Santos et al., 2001
<i>Wichi Formosa</i>	67 ^d	Cabana et al., 2006	<i>Ticuna</i>	32	Tarazona-Santos et al., 2001
<i>Ayoreo</i>	91	Dornelles et al., 2004	<i>Mbyá-Guarani</i>	33	Altuna et al., 2006
<i>Aché</i>	63	Schmitt et al., 2004	<i>Humahuaca</i>	10	Bianchi et al., 1998
<i>Gaviao</i>	27	Ward et al., 1996	<i>Wichi</i>	12	Bianchi et al., 1998
<i>Zoro</i>	30	Ward et al., 1996	<i>Susque</i>	16	Bianchi et al., 1998
<i>Xavante</i>	25	Ward et al., 1996	<i>Lowlands</i>	97	Lee et al., 2007
<i>Guarani</i>	200	Marrero et al., 2007			
<i>Kaingang</i>	74	Marrero et al., 2007			
<i>Quechua Titicaca</i>	37	Barbieri et al., 2010			
<i>Aymara Titicaca</i>	20	Barbieri et al., 2010			
<i>Coya</i>	60	Alvarez-Iglesias et al., 2007			
<i>Buenos Aires</i>	89	Bobillo et al., 2010			
<i>Corrientes</i>	23	Bobillo et al., 2010			
<i>Formosa</i>	15	Bobillo et al., 2010			
<i>Misiones</i>	23	Bobillo et al., 2010			
<i>RioNegro</i>	30	Bobillo et al., 2010			
<i>Guahibo</i>	59	Vona et al., 2005			
<i>Mapuche</i>	34/11	Moraga et al., 2000			
<i>Pehuenche</i>	24/105	Moraga et al., 2000			
<i>Yaghan</i>	15/21	Moraga et al., 2000			
<i>Cayapa</i>	30	Rickards et al., 1999			
<i>Arsario</i>	47	Tamm et al., 2007			
<i>Kogui</i>	48	Tamm et al., 2007			
<i>Ijka</i>	29	Tamm et al., 2007			
<i>Wayuu</i>	42	Tamm et al., 2007			
<i>Coreguaje</i>	27	Tamm et al., 2007			
<i>Vaupe</i>	22	Tamm et al., 2007			
<i>Secoya-Siona</i>	12	Tamm et al., 2007			
<i>Tucuman</i>	9	Tamm et al., 2007			
<i>Salta</i>	18	Tamm et al., 2007			
<i>Catamarca</i>	25	Tamm et al., 2007			
<i>La Paz</i>	106	Afonso Costa et al., 2010			

^a Underlined samples: HVI and HVII control regions available, Italic samples: Control Region (16,024–249) available.

^b Individuals included for mtDNA sequences/haplogroup frequencies comparisons.

^c References for mtDNA sequences/haplogroup frequency data.

^d The two Wichi and the two Toba samples (Formosa and Chaco) were considered together for the haplogroup frequency comparisons.

^e Considering only individuals belonging to a Native American haplogroup.

^f Accession number from the YHRD database (<http://www.yhrd.com>).

* The minimal haplotype available.

** The 12 STRs available. Note: three different names (Coya, Kolla, and Colla) were used to differentiate the three Coya samples from the bibliography.

parameter set at 0.26 (Meyer et al., 1999). Distance matrices were visualized in a multidimensional scaling plot (MDS).

To compare the Y-STR data, 25 South American populations were selected from the literature (Table 1), tak-

ing into account only the haplotypes belonging to Native American Y-chromosomes. In cases of surveys where the haplogroups were not indicated, we inferred them using the Haplogroup Predictor web page (<http://www.hprg.com/hapest5/>). Because of the uneven number of Y-STRs

TABLE 2. *mtDNA (from position 16,017 to 249) and Y-chromosome (17 STRs) haplogroup frequencies and diversity parameters*

mtDNA	Population	N	Haplogroup frequencies					Diversity parameters ^a						
			A2	B2	C1	D1	<i>H</i>	<i>K</i>	π	<i>h</i>	P_w	Sub	Trans	In/dels
	Aymara	96	0.07	0.81	0.06	0.05	0.331	66	0.011	0.990	8.762	95	7	1
	Quechua	93	0.15	0.61	0.19	0.04	0.569	71	0.014	0.993	11.122	108	14	14 (8) ^b
Y-chromosome		N	Haplogroup frequencies			Diversity parameters ^a								
			Q1a3*	Q1a3a	Other	<i>H</i>	<i>K</i>	<i>D</i>	<i>H</i>	P_w				
	Aymara	59	0.10	0.86	0.03	0.192	42	0.465	0.988	7.91				
	Quechua	55	0.00	0.78	0.22	0.000	33	0.425	0.981	7.22				

N, number of individuals; *H*, gene diversity calculated from haplogroups; *K*, number of haplotypes; π , nucleotide diversity; *h*, haplotype diversity; P_w , mean number of pairwise differences; Sub, number of substitutions; Trans, number of transversions; In/dels, number of insertions and deletions; *D*, average gene diversity over loci.

^a Y-chromosome diversity parameters calculated only considering Native American haplogroups.

^b If we consider the 59-60d and the 206-211d as unique events, the number of in/dels would be reduced to eight.

analyzed in the other studies, three sets of data were analyzed including (i) the minimal haplotype (DYS19-DYS389i-DYS389ii-DYS390-DYS391-DYS392-DYS393-DYS385a/b), for which we collected 14 populations with a minimum of nine individuals; (ii) 12 STRs (the minimal haplotype plus DYS437, DYS438, and DYS439) with data for 11 of the 14 samples; and (iii) six STRs (DYS19-DYS389i-DYS389ii-DYS390-DYS391-DYS393) with data for 25 samples (the previous 14 plus 11 others) with at least eight individuals. The nomenclature of DYS389i and DYS389ii loci from Bianchi et al. (1998), Tarazona-Santos et al. (2001), and Lee et al. (2007) was homogenized with the rest of the studies. The six STRs were chosen to include in the comparisons the Tayacaja and Arequipa samples, the only Peruvian samples used in the autosomal study of Gayà-Vidal et al. (2010).

For each data set, diversity parameters, AMOVA (Excoffier et al., 1992) based on the sum of squared differences (Rst), hierarchical AMOVA analyses according to geographical criteria, and pairwise Rst genetic distances (depicted in a MDS) to evaluate genetic relationships among populations were calculated.

All the analyses were performed using the programs Arlequin 3.1 (Excoffier et al., 2005), Network v.4.5.1.6 (<http://www.fluxus-engineering.com>), Statistica (StatSoft 2001), and R.

RESULTS

Diversity in the two Bolivian populations

mtDNA. Haplogroup and mtDNA sequence variation data are shown in the additional file: Supporting Information Table S1. All the 189 individuals corresponded to one of the four major Native American mtDNA haplogroups. The four (A2, B2, C1, and D1) haplogroups were present in the two populations (Table 2). The haplogroup B2 was the most frequent in both samples, especially in the Aymaras (81%). The other three haplogroups showed frequencies of less than 10% in the Aymaras and 20% in Quechuas, and in both cases haplogroup D1 appeared at the lowest frequencies.

A total of 130 different haplotypes were found in the two Bolivian samples (66 in Aymaras and 71 in Quechuas), and only seven were common to both samples (Supporting Information Table S1). Haplotype 2 belonged to haplogroup A2, but it also presented the 9bp deletion that is characteristic of haplogroup B. Haplotype 51, shared by two Quechua individuals, presented several mutations between positions 59 and 71 in the HVII

region that were considered as a 59–60 deletion plus two insertions (65 and 71 sites), which represent three events. Within population diversity (Table 2) was higher in Quechuas than in Aymaras, for all the parameters tested and for the number of transitions, transversions, and in/dels.

The MJ networks for the four haplogroups (see Fig. 2) indicated a clear starlike pattern for A2 and B2, with a central node corresponding to the haplotype presenting the characteristic mutations of the haplogroup. Haplogroups C1 and D1 lacked this central node. In general, the four MJ networks showed high-haplotype diversity, mainly for haplogroup B2. Most of the nodes were small, indicating single haplotypes or haplotypes shared by a few individuals, in most cases belonging to the same population.

Y-chromosome. Haplogroup and haplotype distributions are presented in additional file 2: Supporting Information Table S2. 96.6% and 78.2% of Aymara and Quechua individuals, respectively, carried a Native American haplogroup (lineages Q1a3* or Q1a3a) according to the SNPs tested. The remaining individuals belonged, according to the Haplogroup Predictor web page, mainly to haplogroup R1b, the most frequent in Western Europe (Jobling and Tyler-Smith, 2003). Considering only the Native American Y-chromosomes, 100% of Quechuas and 89% of Aymaras presented the haplogroup Q1a3a, the remaining 11% of Aymaras carried Q1a3*.

Taking into account the 17 STRs and all the 114 individuals tested, 87 different haplotypes were found, and 69 of them (79%) were unique. Only one haplotype (haplogroup Q1a3a) was shared between the two Andean populations. Considering only the Native American haplogroups (57 Aymaras and 43 Quechuas), 74 different haplotypes were found and 57 of these (77%) were unique. Haplotype diversity indices (Table 2) were carried out for these 100 individuals and indices were slightly higher in the Aymaras.

The MJ networks for the Q1a3* and Q1a3a haplogroups (see Fig. 3) revealed a high diversity of male lineages in the two Bolivian populations and showed that, in general, most of haplotypes at the end branches belonged to the Aymaras. The only haplotype shared by the two populations presented one of the biggest nodes.

Comparison of the two Bolivian populations

mtDNA. The exact tests of population differentiation for the mtDNA showed significant differences for both the

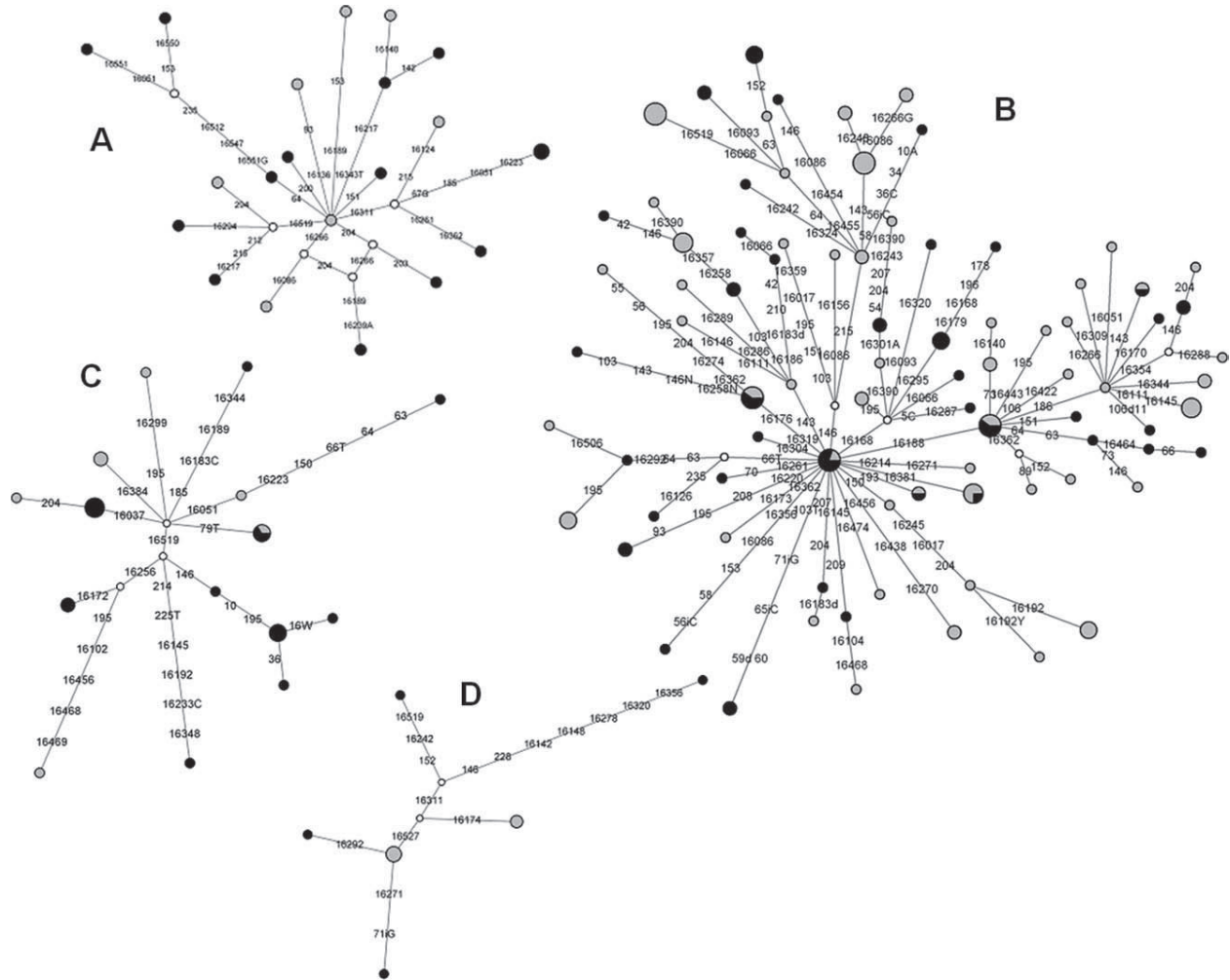


Fig. 2. Haplotype MJ Network for mtDNA haplogroups (A–D). Circle sizes are proportional to the number of individuals carrying the corresponding haplotype. Gray circles correspond to Aymaras and black circles to Quechuas. Small white circles correspond to hypothetical haplotypes. Mutational differences between haplotypes are identified as numbers. Central nodes in haplogroups A2 and B2 correspond to haplotypes with the mutations at nucleotide positions: A2 (16,111, 16,223, 16,290, 16,362, 64, 73, 146, 153, and 235); B2 (16,189, 16,217, 16,519, and 73). All the haplotypes in haplogroups C1 and D1 present the mutations: C1 (16,223, 16,298, 16,325, 16,327, 73, and 249d); D1 (16,223, 16,325, 16,362, 73).

mtDNA haplogroups ($P = 0.012$) and the mtDNA sequences ($P < 0.001$). The same results were obtained through F_{st} values (0.057, $P = 0.003$ for mtDNA haplogroups; 0.009, $P < 0.001$ for mtDNA sequences).

Y-chromosome. The exact tests of population differentiation did not show statistical differences in the allele frequency distributions for any of the 17 loci. Likewise, no significant differentiation of Y-STR haplotypes was found according to the F_{st} value (0.006, $P = 0.239$). On the contrary, slight, but significant differences were obtained for the Y-chromosome haplogroups (0.074, $P = 0.035$).

Comparison with other Native South Americans

mtDNA diversity. Estimates from haplotype and haplogroup data (Table 3) indicated moderate to high levels of diversity in the two Bolivian samples in agreement with other Andean populations. The lowest diversity values were observed in some Colombian samples as well as

in the Aché, and the highest values were observed in the urban Argentinean samples.

The analyses of variance in South America indicated that 27% (HVI), 14.2% (HVII), 18.5% (HVI–HVII), and 9.7% (CR) of the variation could be ascribed to between-population differentiation (Table 4). Hierarchical F_{st} analyses (Table 4) from all sets analyzed, as well as from haplogroup frequencies, did not reveal a genetic structure (diversity among groups > within groups) for Andean versus non-Andean groups. In the same way, according to HVI region and haplogroup frequencies, no genetic structure was found for Andean versus nearby areas (Chaco and Bolivian Lowlands). For all sets analyzed, a significant genetic structure was found in the Andean region when Andean samples were divided into three groups (south Andes vs. central Andes vs. north-west Argentina). It is important to note that the Coya sample was included into the central Andean group and not into the northwestern Argentinean group, in which case, no significant or a weak genetic structure was

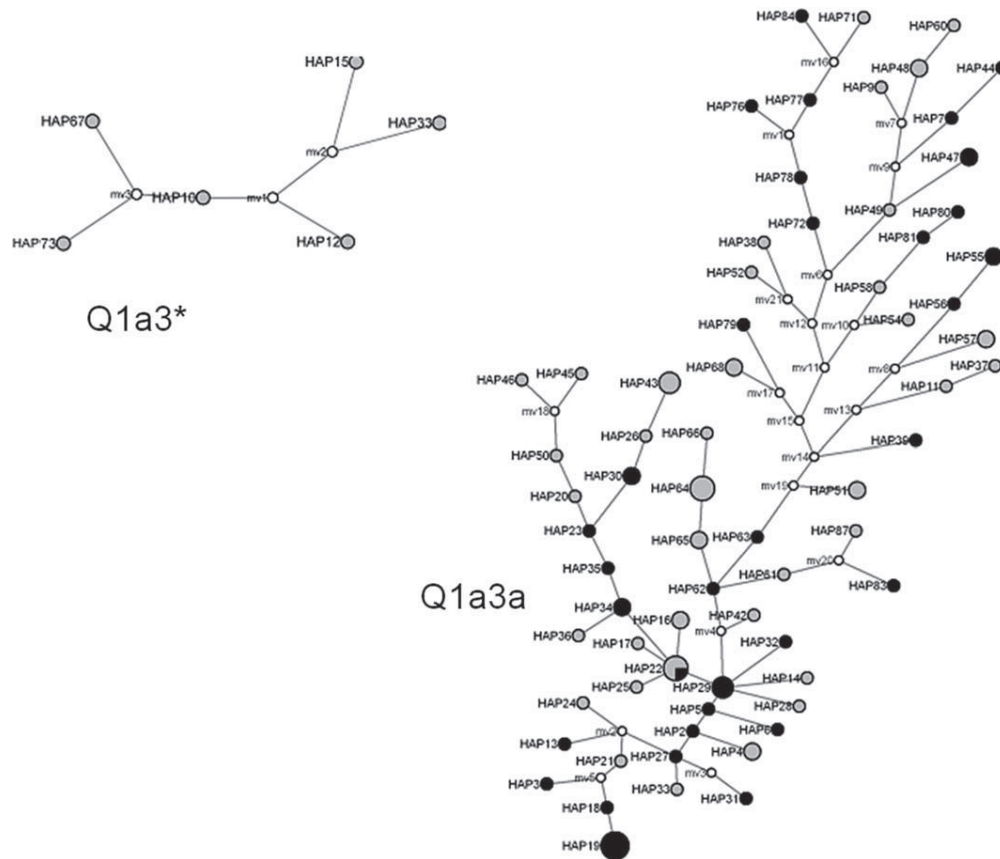


Fig. 3. Network for the Y-chromosome haplogroups Q1a3a and Q1a3*. Circle sizes are proportional to the number of individuals carrying the corresponding haplotype. Gray circles correspond to Aymaras and black circles to Quechuas. Small white circles correspond to hypothetical haplotypes.

found. Considering the central Andes for the HVI region, a slightly significant genetic structure was found for north central Andes versus south central Andes. The CR analysis revealed that most of the genetic differentiation between samples was attributed to between-group differentiation (south central Andes vs. Argentina). Similar results were obtained when the Cayapa sample was included into the central Andean group.

Genetic distances calculated with the Tamura and Nei formula showed that 84% (HVI), 74% (HVII), 82% (HVI-HVII), and 61% (CR) were statistically different from zero (data not shown). The average distance value between all pairs of samples was 0.21 (HVI), 0.16 (HVI-HVII, HVII), and 0.08 (CR). The mean distance between the non-Andean groups was 0.25 (HVI), 0.16 (HVI-HVII, HVII), and 0.01 (CR). The mean distance between the Andean groups was 0.13 (HVI), 0.12 (HVI-HVII), 0.08 (HVII), and 0.02 (CR), values that decreased to 0.09 (HVI), 0.02 (HVI-HVII), and 0.008 (HVII) when the southern Andean and northwest Argentinean samples were removed. It is worth mentioning that all distances between the Aymara sample and the other samples (HVI; HVI-HVII, CR) were statistically significant and showed the lowest distance with the Quechuas (0.036, HVI; 0.029, HVI-HVII; 0.024, CR), closely followed by La Paz (0.038, HVI; 0.03, HVI-HVII) and Aymara Titicaca (0.04, HVI). In contrast, the Quechua sample showed nonstatistically significant distances with Coya for the

four sets, with La Paz (for HVI, HVI-HVII), and with Salta and Quechua Titicaca (for HVI).

MDS plots were built from genetic distances for the four sets of analyses (see Fig. 4). The HVI plot (Fig. 4A) shows most of the Andean samples in the upper left part of the plot (except Catamarca, Tucuman, and Yaghan). The Aymara sample is located at the left extreme, together with Aymara Titicaca and the two Bolivian samples from Corella et al. (2007). The Quechua sample is very close to La Paz, together with most of Andean samples. Wichi Formosa is located in the middle of these Andean samples, and Toba Chaco is very close to Quechua Titicaca. On the other hand, the Peruvian Quechuas from Tayacaja and Yungay, as well as Salta, Mapuche, and Pehuenche, are in the center of the plot, mixed with some non-Andean samples, mainly Chaco and Bolivian lowland samples. The Aché, Ayoreo, Guarani, and Ijka samples appear as the most distant samples. The HVII plot (Fig. 4B) shows the central Andean samples, except Catamarca and Salta, to be relatively grouped in the centre. In the HVI-HVII plot (Fig. 4C), the Quechua, Coya, and La Paz samples are grouped and separated from the Aymaras. The two HVII and HVI-HVII plots reveal the Ijka sample as the most separated population. The CR plot (Fig. 4D) reveals the five Argentinean samples to be practically aligned in the right part of the plot. On the contrary, the Andean samples are rather separated, the Aymara at the left extreme of the plot.

TABLE 3. mtDNA diversity data from contributing populations^a

Populations	HVI-HVII		HVI (16,051–16,362)		HVII (73–249)		Control region (16,024–249)		Haplogroups
	<i>h</i>	π	<i>h</i>	π	<i>H</i>	π	<i>H</i>	π	<i>H</i> ^b
Ijka	0.414	0.005	0.414	0.006	0.197	0.004			0.197
Secoya-Siona	0.818	0.012	0.818	0.016	0.485	0.005			0.53
Kogui	0.668	0.013	0.619	0.014	0.619	0.013			0.503
Arsario	0.790	0.013	0.730	0.014	0.674	0.011			0.472
Yaghan	0.886	0.014	0.886	0.017	0.819	0.008			0.524
Mapuche	0.890	0.014	0.838	0.018	0.793	0.007			0.636
Wayuu	0.820	0.015	0.810	0.017	0.745	0.011			0.675
Aymara	0.988	0.015	0.968	0.015	0.930	0.014	0.990	0.011	0.331
Catamarca	0.913	0.016	0.877	0.018	0.633	0.014			0.68
Pehuenche	0.928	0.016	0.902	0.019	0.859	0.012			0.617
Salta	0.980	0.016	0.967	0.020	0.556	0.010			0.752
Coreguaje	0.872	0.017	0.832	0.018	0.815	0.015			0.527
Cayapa	0.860	0.018	0.837	0.021	0.802	0.014			0.756
Vaupe	0.983	0.018	0.952	0.020	0.878	0.014			0.758
Guahibo	0.895	0.019	0.858	0.016	0.684	0.023			0.541
Quechua	0.988	0.019	0.952	0.020	0.918	0.016	0.993	0.013	0.568
LaPaz	0.993	0.019	0.952	0.020	0.958	0.018			0.579
Tucuman	1.000	0.019	0.972	0.019	0.944	0.019			0.639
Formosa	1.000	0.019	1.000	0.022	0.933	0.014	1.000	0.013	0.762
Misiones	0.984	0.020	0.964	0.022	0.885	0.015	0.984	0.013	0.755
RioNegro	0.991	0.020	0.984	0.022	0.947	0.016	0.991	0.014	0.687
Coya	0.996	0.020	0.980	0.021	0.943	0.017	0.997	0.013	0.556
Buenos Aires	0.993	0.022	0.988	0.023	0.938	0.020	0.994	0.015	0.752
Corrientes	0.984	0.023	0.968	0.022	0.933	0.024	0.988	0.016	0.735
Aché			0.204	0.003					0.175
AymaraBeni			0.667	0.006					0.119
Ayoreo			0.473	0.007					0.281
Guarani			0.764	0.008					0.283
Movima			0.894	0.009					0.571
Xavante			0.677	0.010					0.28
QuechuaBeni			0.673	0.011					0.417
AymaraTiticaca			0.947	0.012					0.195
Zoro			0.775	0.013					0.598
Gaviao			0.866	0.014					0.479
QuechuaYungay			0.954	0.016					0.644
AymaraPuno			0.967	0.016					0.484
Arequipa			0.978	0.016					0.524
JaqaruTupe			0.867	0.017					0.458
QuechuaPuno			0.975	0.017					0.591
Ancash			0.981	0.018					0.669
Kaingang			0.749	0.019					0.545
Chimane			0.800	0.019					0.571
TobaFormosa			0.906	0.019					
Tayacaja			0.968	0.019					0.734
WichiFormosa			0.881	0.020					
TobaChaco			0.888	0.020					0.671 ^c
QuechuaTiticaca			0.954	0.020					0.632
Yucarare			0.952	0.021					0.742
Ignaciano			0.971	0.021					0.697
WichiChaco			0.738	0.022					0.689 ^c
Moseten			0.844	0.022					0.563
Trinitario			0.985	0.022					0.697
PilagaFormosa			0.964	0.023					0.741

h, haplotype diversity; π , nucleotide diversity.

^a Estimates are based on the mtDNA control region from positions 16,051–16,362 (HVI) and from 73 to 249 (HVII), therefore, they may be different from their original published source.

^b Heterozygosities calculated from haplogroup frequencies.

^c Includes also the Wichi or Toba samples from Formosa region.

Y-chromosome diversity. For the Y-chromosome haplotypes, diversity data parameters were calculated for the 27 populations (Table 5). Karitiana, Chimane, and Yukpa presented the lowest diversity values and the Andean samples the highest values. The Andean, Kichwa, Kolla, Peru, Arequipa, and Tayacaja presented the highest levels and Colla, Susque, and Humahuaca the lowest values. The two Bolivian samples of this work presented

intermediate values, the Aymaras presenting slightly higher diversity values than the Quechuas.

The global F_{st} value was similar for the three sets of analysis (6-STRs: 0.24, $P < 0.001$; minimal haplotype and 12-STRs: 0.28, $P < 0.001$). Hierarchical analyses according to geographical criteria (Table 6) revealed a significant genetic structure when central Andes and nearby areas were grouped separately for all sets ana-

TABLE 4. Hierarchical AMOVA analyses with mtDNA sequences and haplogroup frequency data

Population groups ^{a,b}	No. pops	Hierarchical F_{ST} Analyses		
		Values from HVI variation		
		Within groups	Among groups	Total F_{ST}
HVI: Global $F_{st} = 0.270^{***}$				
Andes (21)/non-Andes (32)	53	0.23***	0.10***	0.31***
<i>Andes (21)/non-Andes (30)</i>	51	0.24***	0.09**	0.31***
LowBol (6)/Andes (21)/Chaco (6)	33	0.11***	0.02NS	0.13***
<i>LowBol (6)/Andes (21)/Chaco (4)</i>	31	0.14***	0.01NS	0.14***
LowBol (6)/C Andes (18)/Chaco (6)	30	0.09***	0.03*	0.12***
<i>LowBol (6)/C Andes (18)/Chaco (4)</i>	28	0.08***	0.04*	0.12***
S Andes (3)/C Andes (18)	21	0.09***	0.11**	0.19***
<i>S Andes (3)/C Andes (18)</i>	21	0.08***	0.27***	0.33***
S Andes (3)/C Andes (15)/NWArg (3)	21	0.08***	0.13***	0.19***
<i>S Andes (3)/C Andes (15)/NWArg (3)</i>	21	0.05***	0.26***	0.30***
S Andes (3)/C Andes (15)	18	0.07***	0.13***	0.19***
<i>S Andes (3)/C Andes (15)</i>	18	0.05***	0.30***	0.34***
C Andes (15)/NWArg (3)	18	0.08***	0.13***	0.20***
<i>C Andes (15)/NWArg (3)</i>	18	0.06***	0.19***	0.23***
NC Andes (5)/SC Andes (10)	15	0.05***	0.06**	0.11***
<i>NC Andes (5)/SC Andes (10)</i>	15	0.04***	0.04**	0.08***
HVII: Global $F_{st} = 0.142^{***}$				
Andes (10)/non-Andes (14)	24	0.10***	0.08**	0.17***
S Andes (3)/C Andes (4)/NW Arg (3)	10	0.01**	0.07***	0.08***
S Andes (3)/C Andes (7)	10	0.04***	0.02NS	0.06***
S Andes (3)/C Andes (4)	7	0.01***	0.03*	0.04***
C Andes (4)/NW Arg (3)	7	0.01*	0.10*	0.11***
HVI-HVII: Global $F_{st} = 0.185^{***}$				
Andes (10)/Non-Andes (14)	24	0.13***	0.12***	0.23***
S Andes (3)/C Andes (7)	10	0.07***	0.11NS	0.17***
S Andes (3)/C Andes (4)/NW Arg (3)	10	0.03***	0.15**	0.17***
S Andes (3)/C Andes (4)	7	0.02***	0.15*	0.17***
C Andes (4)/NWArg (3)	7	0.02***	0.16*	0.18***
Control Region: Global $F_{st} = 0.097^{***}$				
SC Andes (3)/Argentina (5)	8	0.02**	0.13*	0.15***

Andes: *South Andes:* Mapuche, Pehuenche, Yaghan; *Central Andes:* (i) *North Central Andes:* Yungay, Tayacaja, Arequipa, Jaqaru, and Ancash; (ii) *South Central Andes:* Aymara, Quechua, Aymara Beni, Quechua Beni, Aymara Puno, Quechua Puno, LaPaz, Coya, Aymara Titicaca, and Quechua Titicaca; *NW Argentina:* Catamarca, Tucuman, and Salta.

Non-Andes: Arsario, Ijka, Coreguaje, Kogui, Wayuu, Vaupe, Secoya-Siona, Guahibo, Zoró, Xavante, Ayoreo, Aché, Gaviao, Cayapa, Misiones, Corrientes, RioNegro, and Buenos Aires; *Bolivian Lowlands:* Ignaciano, Movina, Trinitario, Yucarare, Chimane, and Mosen; *Chaco:* Toba Chaco, Toba Formosa, Wichi Chaco, Wichi Formosa, Pilaga Formosa, and Formosa.

*** $P < 0.001$,

** $P < 0.01$,

* $P < 0.05$.

NS: nonsignificant.

Values in **bold:** significant cases where there was geographic structure.

^{a,b} Number of populations included in each group.

lyzed. Focusing on the central Andes, the 6-STR analysis showed a genetic structure when samples were divided into two groups (north central Andes vs. middle and south central Andes). For the minimal haplotype and 12-STRs, we could not check this differentiation, because only one north sample was available (Kichwa). However, an absence of genetic structure for middle central Andes versus south central Andes was observed. Similar results were found when the Cayapa sample was included into the north central Andean group.

The Rst genetic distance matrices showed 71% (6-STRs and 12-STRs), and 81% (minimal haplotype) of significant distances. The average distance between all pairs of samples was 0.22 (6-STRs, minimal haplotype) and 0.17 (12-STRs). The mean distance between non-Andean samples was 0.27 (6-STRs), 0.32 (minimal haplotype), and 0.17 (12-STRs), seven, five, and almost three times higher than the values between Andean samples (0.04, 6-STRs; 0.07, minimal haplotype, and 12-STRs).

Focusing on the Andean groups, all the analyses highlighted their proximity, with the exception of the Kichwa from Ecuador and the Mapuches. The 6-STR analysis showed that 69.6% of the distances were not statistically significant; the pairs made up of Kichwa were the most significant. Both, the minimal haplotype and the 12-STR analyses revealed nonsignificant distances between all pair of populations composed of Aymara, Colla, Kolla, Diaguita, and Peru. The Quechuas presented nonsignificant distances with Aymara, Kolla, and Diaguita. On the other hand, all distances between the Kichwa and the other Andean samples were significant, except with Colla and Mapuche.

The Rst distances depicted in the MDS plots (see Fig. 5) showed the Andean samples to be relatively grouped and most of the non-Andean populations scattered on the plots. The 6-STR plot (Fig. 5A) revealed the Chimane, Ticuna, Yanomami, Mbyá-Guarani, Bari, Yukpa, and Lowlands as the more distant populations. On the

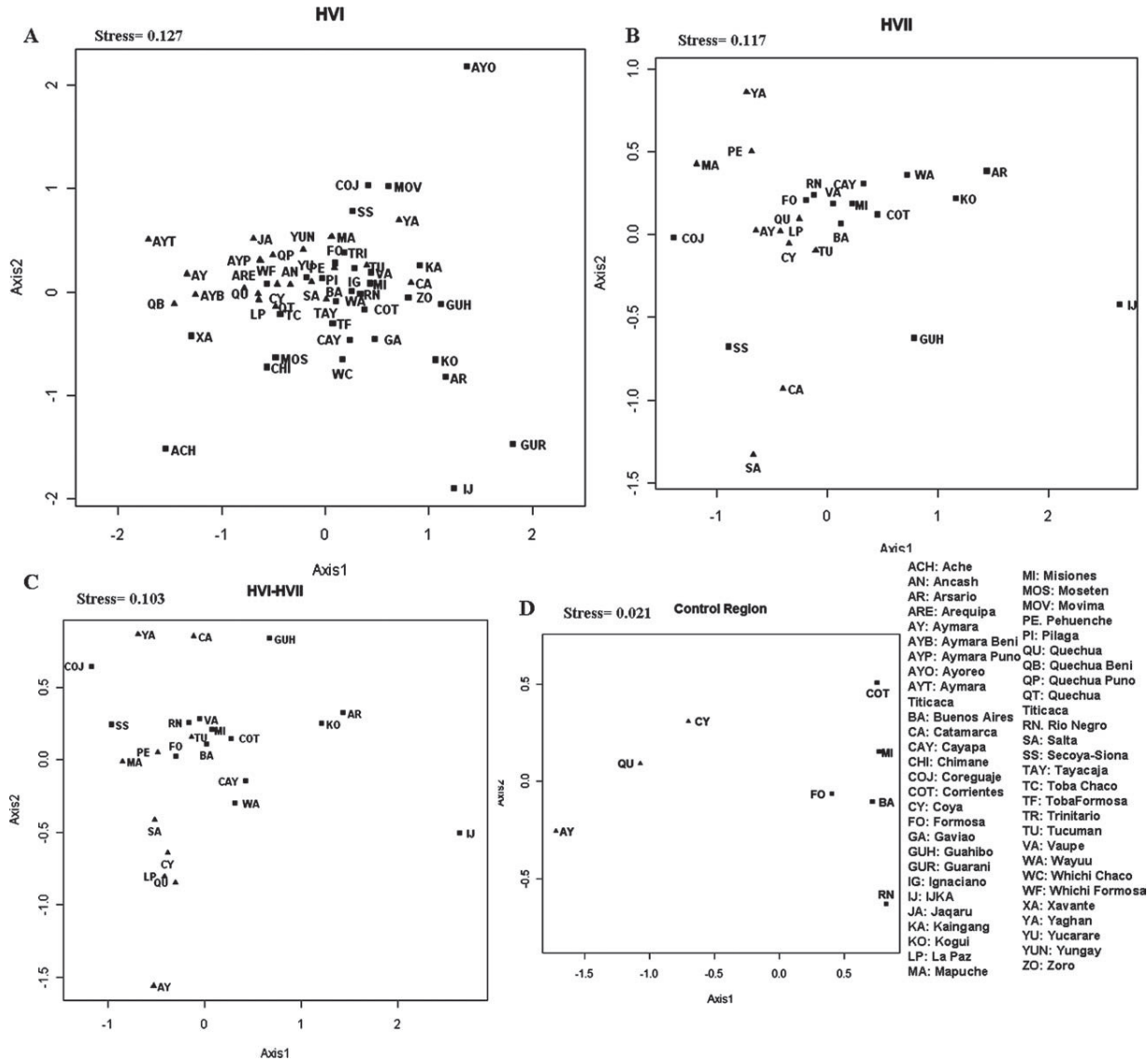


Fig. 4. MDS constructed from mtDNA Tamura and Nei genetic distances. **A:** mtDNA HVI region, **(B)** mtDNA HVII region, **(C)** HVI-HVII regions, **(D)** control region. Triangles and circles represent Andean and non-Andean samples, respectively.

contrary, the Toba, Wichi, Trinitario, Karitiana, and Gaviao-Zoro-Surui samples were the closest to the Andean group, especially to the Susque, Tayacaja, Kichwa, and Cayapa. The two Bolivian, the southernmost Peruvian, and the other northwest Argentinean samples were the most separated from the non-Andean samples. Both, the minimal haplotype and the 12-STR plots (Fig. 5B,C) showed the Kichwas (Ecuador) to be relatively separated from the other six central Andean samples, which formed a group. The Mojeño, Mapuche, Trinitario, Kichwa, and Kaingang-Guarani samples were relatively close to each other.

DISCUSSION

The analysis of both mtDNA and Y-chromosome in this study adds a new perspective to the autosomal data from Gaya-Vidal et al. (2010) for the genetic characterization of the two main linguistic groups in Bolivia, the

Aymarans, and Quechuas and contributes new data on Native American genetic variability.

External contributions to the current gene pool of Bolivian populations

Previous studies on classical markers indicated low external admixture in the two Bolivian samples here examined; around 1% of the specific European haplotype GM5*₃ (Dugoujon JM, personal communication) and 98% of O group from the ABO system (hematological study by the IBBA). In the present study, differences were found by gender. The estimates of the non-Amerindian Y-chromosome and mtDNA haplogroups indicated a total absence of admixture for the mtDNA, but a certain proportion of Y-chromosomes admixture, especially in Quechuas which had 22% of non-Native American Y-chromosomes. This differentiation between the two types of markers is a general trend in all Native American

TABLE 5. Diversity data for Y-STRs haplotypes from populations included in the analyses, considering only Native American haplogroups

Population	Minimal haplotype		12-STRs		6-STRs	
	Haplotype diversity	Expected heterozygosity	Haplotype diversity	Expected heterozygosity	Haplotype diversity	Expected heterozygosity
Chimane	0.378	0.106	0.378	0.096	0.200	0.067
Yukpa	0.303	0.168			0.303	0.152
Bari	0.517	0.259			0.517	0.189
Yanomami	0.946	0.337			0.818	0.236
Toba	0.942	0.429	0.963	0.428	0.834	0.296
Kaingang-Guarani	0.906	0.460	0.914	0.406	0.886	0.451
Mojeño	0.889	0.472	0.889	0.438	0.889	0.421
Diaguíta	0.972	0.500	0.972	0.481	0.944	0.398
Mapuche	0.988	0.502	0.988	0.467	0.964	0.430
Quechua	0.960	0.516	0.976	0.462	0.780	0.446
Trinitario	0.972	0.535	0.972	0.532	0.943	0.506
Colla	0.778	0.541	0.889	0.467	0.778	0.470
Aymara	0.971	0.551	0.982	0.486	0.894	0.484
Peru	0.989	0.570	0.995	0.507	0.962	0.506
Kolla	0.984	0.593	1.000	0.506	0.909	0.503
Kichwa	1.000	0.604	1.000	0.548	0.989	0.551
Karitiana					0.250	0.042
Wichi					0.909	0.316
Ticuna					0.698	0.323
Gaviao-Zoro-Surui					0.882	0.324
Mbya-Guarani					0.854	0.363
Humahuaca					0.978	0.385
Susque					0.967	0.443
Lowlands					0.954	0.461
Cayapa					0.963	0.501
Tayacaja					0.980	0.507
Arequipa					0.952	0.554

Values in bold: the two populations of this study.

populations; a consequence of the colonization by the Europeans. Nevertheless, it is worth noting that the Aymaras showed a remarkably low level of admixture with less than 2% of non-Native American Y-chromosomes compared to other Andean samples (Dipierrri et al., 1998; Tarazona-Santos et al., 2001; González-Andrade et al., 2007) that highlights its isolation from non-Native Americans.

Genetic variation in Aymaras and Quechuas from Bolivia

mtDNA. The results revealed typical Andean characteristics in the two Bolivian samples as well as a certain degree of differentiation between them. Thus, haplogroup B2, the most frequent in the Andean region, was the most frequent in both the Aymaras (81%) and Quechuas (61%). It is interesting to note that around 60% of the diversity of the B2 haplotypes corresponded to groups defined by the variants 16,168, 16,188, 103–143, and 146–215. Particularly, the variant 16,188 was observed in 31% (Aymaras) and 21% (Quechuas) of B2 haplogroups. This variant seems to be characteristic of the Andean Altiplano, because it was also present in Aymara Titicaca (66%), Quechua Titicaca (38%), La Paz (44%), Coya (15%), Aymara Beni (11%), Quechua Beni (81%), and Arequipa (31%). Moreover, in a subgroup of these samples (Aymara, Quechua, Aymara Titicaca, Quechua Titicaca, La Paz, and Coya), the variant 16,188 was always combined with the variant 16183C. Within this subbranch, the 186 variant was present in half of the haplotypes and other minor clusters were defined by the 63–64 variants, the lack of variant 73, and so forth. The 186 variant and the lack of the 73 variant were also

found in Coya and La Paz. The presence of the 16,188 variant in one individual of two populations from the Chaco region (Wichi Formosa and Pilaga) could indicate interactions between these two regions. Traces of contacts between different South American regions are also supported by the presence of two haplotypes typical of the Guaraní (combination 16,239A–16,266, Marrero et al., 2007) and northwest Argentinean lineages (combination 16,242–16,311, Tamm et al., 2007) in our Quechua sample.

On the other hand, it is interesting to discuss several particular characteristics found in the mtDNA of the two Bolivian samples. First, the Quechua haplotype 39 presented a 106–111d, also reported in one individual from La Paz (LPAZ070) sharing the same haplotype (considering the HVI and HVII regions separately). The 106–111d was proposed to be characteristic of Chibchan-speaking populations (Santos and Barrantes, 1994; Kolman et al., 1995). However, in those samples, the deletion occurred within haplogroup A2 and not B2 as in the two Andean cases, indicating a recurrent mutation rather than a trait restricted to a certain group. Note that this deletion is different from the 105 to 110d reported in the Coyas. Second, the haplotypes 18 and 19 (haplogroup A2) presented some mutations (lack of the haplogroup A diagnostic site 235, variants 16,512, 16,547, 16,551iG, and absence of the 64 variant) also found (except for 16551iG) in one Coya individual. Third, haplotype 51, with such a particular mutation combination, highlights the huge variability between the 55 and 71 positions in the HVII. These features highlight the importance of sequencing not only the HVI, nor the HVI and HVII separately, because interesting polymorphisms are located outside the classical segments.

TABLE 6. Hierarchical AMOVA analyses with Y chromosome 6 STRs, minimal haplotype, and 12 STRs data

Population groups ^a	Hierarchical F_{ST} Analyses			
	No. pops	Within groups	Among groups	Total F_{ST}
6 STRs Global $F_{st} = 0.239^{***}$				
Andes (12)/non-Andes (14)	26	0.201***	0.015NS	0.213***
C Andes (11)/Chaco (2)/Bol_Low (3)	16	0.054***	0.062**	0.113***
NC Andes (2)/MC Andes, SC Andes (9)	11	0.008NS	0.056*	0.064***
MC Andes (4)/SC Andes (5)	9	0.000NS	0.008NS	0.008NS
Minimal haplotype Global $F_{st} = 0.275^{***}$				
Andes (8)/non-Andes (8)	16	0.256***	0.048NS	0.292***
C Andes (7)/Bol_Low (3)	10	0.081***	0.131*	0.202***
MC Andes (3)/SC Andes (3)	6	0.009NS	0.006NS	0.015NS
12 STRs Global $F_{st} = 0.279^{***}$				
Andes (8)/non-Andes (5)	13	0.242***	0.095NS	0.314***
C Andes (7)/Bol_Low (3)	10	0.085***	0.126*	0.200***
MC Andes (3)/SC Andes (3)	6	0.116NS	0.000NS	0.016NS

*** $P < 0.001$,

** $P < 0.01$,

* $P < 0.05$;

NS, nonsignificant. Values in **bold**: cases with geographic structure.

^a In parentheses, number of populations included in each group.

Andes: Mapuche, Central Andes: (i) North Central Andes: Kichwa and Tayacaja; (ii) Middle Central Andes: Peru, Arequipa, Aymara, and Quechua; (iii) South Central Andes: Colla, Susque, Humahuaca, Kolla, and Diaguita.

Non-Andes: Bari, Toba, Kaingang_Guarani, Yanomani, Yukpa, Trinitario, Chimane, Mojeños, Gavião-Zoro-Surui, Karitiana, Ticuna, Mbyá-Guaraní, Wichi, and Cayapa.

Considering the CR analysis, an important result was the high-mtDNA diversity observed in the two Bolivian samples, especially in the Quechuas, which was similar to the Coyas. This high-mtDNA CR diversity in the two Bolivian and other Andean samples confirm strongly the findings of Fuselli et al. (2003), suggesting a high-long-term effective population size in the Andean region. Higher values were found in some Argentinean samples, but these are most probably due to their mixed nature, because they correspond to a political subdivision.

Y-chromosome. All Native American Y-chromosomes in Quechuas and 89% of Aymaras belonged to haplogroup Q1a3a, which is the most frequent haplogroup in South America. The remaining Aymaras presented the paralog Q1a3* (11%), a value that is double that reported in other Bolivian samples (Bailliet et al., 2009). In any case, this study supports that the northwest border of South America harbors the highest frequencies of the Q1a3* lineage, as proposed in Bailliet et al. (2009). This high-Q1a3* frequency in Aymaras could be attributed to drift effects, but the high-diversity values (haplotype) observed in Aymaras, as well as in other Andean samples, is not consistent with this interpretation.

Concerning the Y-STR variation, an interesting result was the high frequency of the DYS393*14 allele in the two Bolivian samples; 56 and 58% of Aymaras and Quechuas, respectively. In the context of the central Andes, the average frequency of this allele was 19% (Ecuador), 40% (Peru), 42% (northwest Argentina), and 57% (Bolivia), which may indicate that its origin was in the Andean Altiplano with a subsequent expansion to the surrounding areas. These results support the study of Martínez-Marignac et al. (2001) that found that northwest Argentinean samples were characterized by a high frequency of this allele (38.9%), which was suggested to have a likely Altiplano origin, because most of surnames in the region were of Aymara origin. However, the South American distribution of this allele presents two discontinuous regions of high frequencies: on one hand, the central Andes with frequencies ranging from 12% in the

Cayapas to 80% in Humahuaca, and on the other hand, the Venezuelan samples, Bari and Yukpa, with values around 90%. The total discontinuity between these two areas suggests two different events for the origin of this allele, and in order to verify this, we analyzed the composition of the haplotypes carrying this allele for the minimal haplotype. All the Andean samples shared haplotypes that were unique to their group (except a Toba individual that carried a haplotype also present in the Quechuas), and all the haplotypes in the Bari and Yukpa were unique. Moreover, when we removed the two most variable STRs (DYS385a/b), similarly the Andean samples shared haplotypes only with Andeans, except for one Aymara individual carrying a Bari haplotype, one Peru individual carrying a Yukpa haplotype, and the only Chimane individual carrying a haplotype also present in Kichwa. These results support the hypothesis of two independent origins.

Genetic relationships among Native South Americans

Our results revealed a similar value (around 25%) of between-population genetic diversity among South Americans for both sets of markers and failed to indicate a strong clustering of to the two main geographic areas in South America (west vs. east). However, different patterns of variation were observed in the Andean region compared to the east. Eastern samples presented larger within-group genetic distances and lower intrapopulation diversity parameters than for the Andean samples for both sets of markers. This is consistent with different patterns of drift and gene flow related to larger effective population sizes in the Andean area, as suggested by different kinds of data (mtDNA, Fuselli et al., 2003; Y-chromosome, Tarazona-Santos et al., 2001; classical markers, Luiselli et al., 2000; STRs, Wang et al., 2007, and PAIs, Gaya-Vidal et al., 2010). High diversities have also been reported in the Andean surrounding areas of Chaco and the Bolivian lowlands that may suggest a certain influence from the Andean region (also reflected in the MDS

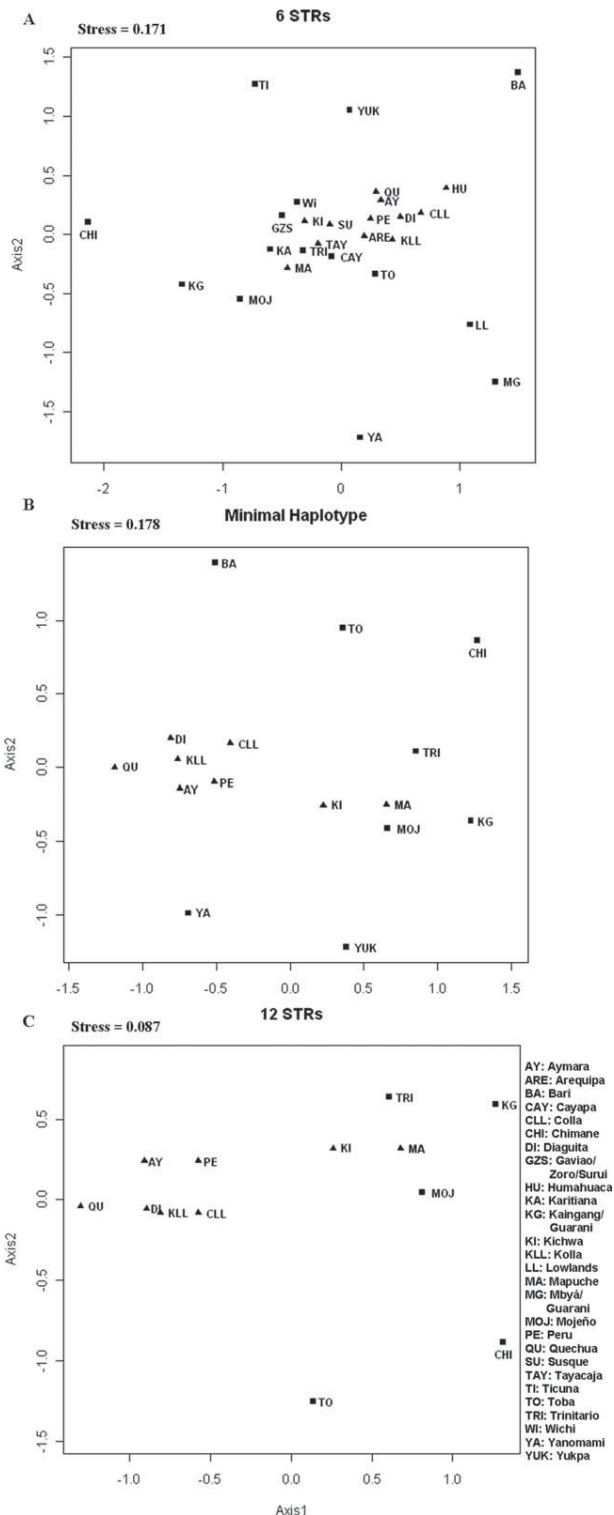


Fig. 5. MDS constructed from Y-chromosome Rst genetic distances. Considering: (A) 6 STRs, (B) the minimal haplotype, and (C) 12 STRs. Triangles and circles represent Andean and non-Andean samples, respectively.

plots), mainly for the mtDNA data, because no genetic structuring was found according to hierarchical AMOVA analyses. On the contrary, this Andean influence has not

been strong enough to avoid significant genetic structuring for the Y-chromosome among these three areas.

Regarding the Andean range, the most important genetic differentiation was observed between south versus central Andean populations for both the mtDNA and Y-chromosome data, in agreement with geographical distance.

Concerning the central Andean region, both markers revealed a general homogeneity of this area according to the hierarchical AMOVA analyses and genetic distances if we exclude the Kichwa (Ecuador) for the Y-chromosome, and the three northwest Argentinean samples (Salta, Tucuman, and Catamarca) for the mtDNA. However, the Coyas, also in northwest Argentina, appeared very close to the Bolivian and Peruvian samples. This fact could be attributed to geographic distance (Tucuman and Catamarca) and political rather than ethnic subdivisions (Salta, Tucuman, and Catamarca), but it is most probably due to the Altiplano influence on the Coyas. In fact, the term Coya was used by the Incas to refer to the Aymara inhabitants. The Incas conquered the Aymara territories forming the southeastern provincial region of the Inca Empire, called “Collasuyu.”

Focusing on our samples, for the HVI region, the mean genetic distance between the Peruvian samples (4) and the Aymaras, Quechuas, and Coyas was 0.17, 0.07, and 0.05, respectively. Moreover, the mean distance between the Peruvian samples and the Bolivian Aymaras (5) and the Bolivian Quechuas (4) was 0.12 and 0.08, respectively; the HVI plot showed the Aymaras slightly separated from the Peruvians and Coya. The HVI-HVII comparisons highlighted the separation of the Aymaras from the Quechuas, Coyas, and LaPaz, which clustered together. These results lead to the conclusion that (i) the Altiplano region, including the Aymaras, Quechuas, and Coyas present a certain degree of similarity, probably due to the ancient Aymara influence area; (ii) the Quechua and Coya samples would have received more Peruvian influences, probably during the Inca Empire that also imposed the Quechua language; and (iii) the Aymaras would have remained more isolated, thus maintaining certain mtDNA characteristics.

As for the Y-chromosome, the analyses of 6-STRs revealed a clear concordance between genetics and geography, with the highest genetic distance between the Kichwa and Humahuaca. All the analyses highlighted the differentiation of the Kichwas from the other central Andean samples, which formed a group. It is interesting to note that for the minimal haplotype most of the Andean samples (Aymara, Peru, Colla, and Diaguita) only shared haplotypes with other Andean samples. The minimal haplotype 13-14-31-23-10-16-14-15-18 was the most frequent among Andean samples, shared by Aymara, Quechua, Kolla, Colla, and Peru samples, and therefore, characteristic of the central Andes. The Quechua, Kolla, and Kichwa shared one haplotype with Toba from the Chaco area (when 12 STRs were considered, only the Quechua shared it), indicating gene flow into Andean populations from this area. These results indicate that male gene flow inside the Andean region, especially within the south central Andes, has been remarkably high.

Genetic relationships between the two Bolivian populations

In a previous study (Gayà-Vidal et al., 2010), the two Bolivian samples presented, according to 32 PAIs, a

genetic similarity and a separation from the two Peruvian Quechua-speaker samples (Tayacaja and Arequipa) from the literature. This suggested a common origin of the two Bolivian populations and an expansion of the Quechua language mainly due to cultural rather than demographic processes. In the present study, the Tayacaja sample also appeared as one of the most differentiated Andean populations. However, a larger number of Andean samples was available for the comparisons, permitting more consistent conclusions, taking into account that the two systems are just two independent loci.

In this study, the comparison of the two Bolivian populations revealed more genetic differences for the mtDNA than for the Y-chromosome; that is, both markers reveal different histories. It is commonly accepted that the social organization of Andean populations was a patrilineal system. Under this assumption, more mtDNA similarities would be expected between the two Bolivian samples, unless a higher proportion of gene flow from external areas affected the Quechuas, as demonstrated by the differences in the frequency of haplogroup B2 and the presence of particular haplotypes from other non-Andean areas. Likewise, it is important to remember the presence in our populations, that is, in the Altiplano, of specific mtDNA features possibly related to high-long-term effective sizes since ancient times. The history of Y-chromosome is different. The distribution of the Y-chromosome variation indicates a clear genetic homogeneity inside the whole central Andean region. This homogeneity could be explained by the higher mobility of males than females across the entire region that might have been favored during the Inca Empire. This apparent controversy could be explained by the different nature of the markers analyzed to date.

CONCLUSION

We can hypothesize a demographic scenario to explain the information supplied by the three kinds of genetic data. According to the very low mutation rate of autosomal Alu markers, these data suggest a past common origin of the Altiplano populations, including the current Aymaras and Quechuas from Bolivia. The arrival of the Inca Empire stimulated the movement of people across the Andean region (probably by the *mitma* system). These movements were especially effective in changing the language (imposition of Quechua), but some regions presented important resistance, including the Titicaca Basin (Aymaras). The demographical consequences of these displacements would have been restricted to the beginning period, according to the very low-genetic distances between these two populations. But, the new Quechua-speaking areas would have been more permeable to the incorporation of foreigners. This is consistent with the closer genetic distances of the Quechuas to the Peruvians and Coyas and the presence of other South American lineages. Finally, in this context, the Y-chromosome homogeneity suggests an important male mobility in the Andean area. Nevertheless, data on additional central Andean samples and more markers are necessary to confirm this scenario.

ACKNOWLEDGMENTS

We are grateful to all the donors for providing DNA samples and to the people who contributed to the collection of samples. We acknowledge support from the CRGS

platform of Genopole Toulouse Midi-Pyrénées, where mtDNA and NRY sequencing/genotyping work was performed. We are grateful to the Service de Coopération et d'Action Culturelle (in particular we thank Marie-Claude Senemaud) of French embassy in La Paz and to Mr. Alain Fouquet (the French ambassador). Genealogical information was collected by IBBA, with the help of Emile Crognier and André Sevin (CNRS). We particularly thank E. Guitard, C. Thèves, S. Mazières, C. Coudray, N. Brucato, and H. Razafindrazaka for their technical help and advice.

LITERATURE CITED

- Afonso Costa H, Carvalho M, Lopes V, Balsa F, Bento AM, Serra A, Andrade L, Anjos MJ, Vide MC, Pantoja S, Vieira DN, Corte-Real F. 2010. Mitochondrial DNA sequence analysis of a native Bolivian population. *J Forensic Leg Med* 17:247–253.
- Altuna ME, Modesti NM, Demarchi DA. 2006. Y-chromosomal evidence for a founder effect in Mbyá-Guaraní Amerindians from northeast Argentina. *Hum Biol* 78:635–639.
- Alvarez-Iglesias V, Jaime JC, Carracedo A, Salas A. 2007. Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forens Sci Int Genet* 1:44–55.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Bailliet G, Ramallo V, Muzzio M, García A, Santos MR, Alfaro EL, Dipierri JE, Salceda S, Carnese FR, Bravi CM, Bianchi NO, Demarchi DA. 2009. Brief communication: restricted geographic distribution for Y-Q* paragroup in South America. *Am J Phys Anthropol* 140:578–582.
- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
- Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kisivild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang YP. 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann Hum Genet* 67:512–524.
- Bandelt HJ, Lahermo P, Richards M, Macaulay V. 2001. Detecting errors in mtDNA data by phylogenetic analysis. *Int J Legal Med* 115:64–69.
- Bandelt HJ, Macaulay V, Richards M. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol Phylogenet Evol* 16:8–28.
- Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet* 71:1150–1160.
- Barbieri C, Heggarty P, Castri L, Luiselli D, Pettener D. 2011. Mitochondrial DNA variability in the Titicaca basin: matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol* 23:89–99.
- Bert F, Corella A, Gené M, Pérez-Pérez A, Turbón D. 2001. Major mitochondrial DNA haplotype heterogeneity in highland and lowland Amerindian populations from Bolivia. *Hum Biol* 73:1–16.
- Bert F, Corella A, Gené M, Pérez-Pérez A, Turbón D. 2004. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. *Ann Hum Biol* 31:9–28.
- Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez-Camelo JS.

1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World Native populations. *Am J Hum Genet* 63:1862–1871.
- Blanco Vereza A, Jaime JC, Brión M, Carracedo A. 2010. Y-chromosome lineages in native South American population. *Forens Sci Int Genet* 4:187–193.
- Bobillo MC, Zimmermann B, Sala A, Huber G, Röck A, Bandelt HJ, Corach D, Parson W. 2010. Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. *Int J Legal Med* 124:263–268.
- Bouysse-Cassagne T. 1986. Urco and Uma: Aymara concepts of space. In: Murra JV, Wachtel N, Revel J, editors. *Anthropological history of Andean polities*. Cambridge: Cambridge University Press. p 201–227.
- Browman DL. 1994. Titicaca Basin archaeolinguistics: Uru, Pukina and Aymara AD 750–1450. *World Archaeol* 26:235–251.
- Cabana GS, Merriwether DA, Hunley K, Demarchi DA. 2006. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. *Am J Phys Anthropol* 131:108–119.
- Cinnioglu C, King R, Kivisild T, Kalfoglou E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114:127–148.
- Corella A, Bert F, Pérez-Pérez A, Gené M, Turbón D. 2007. Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Moseten, Aymara and Quechua. *Ann Hum Biol* 34:34–55.
- Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A, Dugoujon JM. 2009. The complex and diversified mitochondrial gene pool of Berber populations. *Ann Hum Genet* 73:196–214.
- Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70:1001–1010.
- Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM. 2004. Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* 16:479–488.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
- Excoffier L, Smouse P, Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
- Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol* 20:1682–1691.
- 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.
- González-Andrade F, Sánchez D, González-Solórzano J, Gascón S, Martínez-Jarreta B. 2007. Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador. *Hum Biol* 79:51–77.
- Iannacone GC, Tito RY, Lopez PW, Medina ME, Lizarraga B. 2005. Y-chromosomal haplotypes for the PowerPlex Y for twelve STRs in a Peruvian population sample. *J Forens Sci* 50:239–242.
- Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830–838.
- Kolata A. 1993. *The Tiwanaku: portrait of an Andean civilization*. The Peoples of America. Cambridge: Blackwell.
- Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F. 1995. Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. *Genetics* 140:275–283.
- Lee J, Ewis AA, Hurles ME, Kashiwazaki H, Shinka T, Nakahori Y. 2007. Y chromosomal STRs haplotypes in two populations from Bolivia. *Legal Med (Tokyo)* 9:43–47.
- Leite FP, Callegari-Jacques SM, Carvalho BA, Kommers T, Matte CH, Raimann PE, Schwengber SP, Sortica VA, Tsuneto LT, Petzl-Erler ML, Salzano FM, Hutz MH. 2008. Y-STR analysis in Brazilian and South Amerindian populations. *Am J Hum Biol* 20:359–363.
- Lewis CM Jr., Lizarraga B, Tito RY, Lopez PW, Iannacone C, Medina A, Martinez R, Polo SI, De La Cruz AF, Caceres AM, Stone AC. 2007. Mitochondrial DNA and the peopling of South America. *Hum Biol* 79:159–178.
- Lewis CM Jr., Tito RY, Lizarraga B, Stone AC. 2005. Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru. *Am J Phys Anthropol* 127:351–360.
- Luiselli D, Simoni L, Tarazona-Santos E, Pastor S, Pettener D. 2000. Genetic structure of Quechua-speakers of the Central Andes and geographic patterns of gene frequencies in South Amerindian populations. *Am J Phys Anthropol* 113:5–17.
- Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, Bortolini MC. 2007. Demographic and evolutionary trajectories of the Guarani and Kaingang Natives of Brazil. *Am J Phys Anthropol* 132:301–310.
- Martínez-Marignac VL, Bailliet G, Dipierri JE, Alfaro E, López-Camelo JS, Bianchi NO. 2001. Variabilidad y antigüedad de linajes holandricos en poblaciones jujeñas. *Rev Arg Antropol Biol* 3:65–77.
- Mazières S, Guitard E, Crubézy E, Dugoujon JM, Bortolini MC, Bonatto SL, Hutz MH, Bois E, Tiouka F, Larrouy G, Salzano FM. 2008. Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations—implications for the region's colonization. *Ann Hum Genet* 72:145–156.
- Meyer S, Weiss G, von Haeseler A. 1999. Pattern of nucleotide substitution and rate of heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103–1110.
- Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvallo P. 2000. Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. *Am J Phys Anthropol* 113:19–29.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P. 1999. Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112:291–298.
- Polzin T, Daneschmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. *Operations Res Lett* 31:12–20.
- Raymond M, Rousset F. 1995. An exact test for population differentiation. *Evolution* 49:1280–1283.
- Rickards O, Martínez-Labarga C, Lum JK, De Stefano GF, Cann RL. 1999. mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. *Am J Hum Genet* 65:519–530.
- Rowe JH. 1963. Inca culture at the time of the Spanish conquest. In: Steward JH, editor. *Handbook of South American Indians*, Vol. 2. New York: Cooper Square. p 183–330.
- Santos M, Barrantes R. 1994. D-loop mtDNA deletion as a unique marker of Chibchan Amerindians. *Am J Hum Genet* 55:413–414.
- Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, Salzano FM. 2004. Extremely limited mitochondrial DNA variability among the Aché Natives of Paraguay. *Ann Hum Biol* 31:87–94.

- Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet* 73:700–705.
- Stanish C. 2001. The origin of state societies in South America. *Annu Rev Anthropol* 30:41–64.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R, Malhi RS. 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. *Mol Biol Evol* 10:512–526.
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 68:1485–1496.
- Tirado M, López-Parra AM, Baeza C, Bert F, Corella A, Pérez-Pérez A, Turbón D, Arroyo-Pardo E. 2009. Y-chromosome haplotypes defined by 17 STRs included in AmpF1STR Yfiler PCR Amplification Kit in a multi ethnical population from El Beni Department (North Bolivia). *Legal Med (Tokyo)* 11:101–103.
- Torrioni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162.
- Toscanini U, Gusmão L, Berardi G, Amorim A, Carracedo A, Salas A, Raimondi E. 2008. Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data. *Forens Sci Int Genet* 2:274–280.
- Tschopik H. 1963. The Aymara. In: Steward JH, editor. *Handbook of South American Indians*, Vol. 2. New York: Cooper Square. p 501–573.
- Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci USA* 93:196–200.
- Vona G, Falchi A, Moral P, Calò CM, Varesi L. 2005. Mitochondrial sequence variation in the Guahibo Amerindian population from Venezuela. *Am J Phys Anthropol* 127:361–369.
- Wang S, Lewis CM Jr, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsunedo LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in Native Americans. *PLoS Genet* 3:e185.
- Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA JR, Santos RV. 1996. Mitochondrial DNA polymorphism in three Brazilian Indian tribes. *Am J Hum Biol* 8:317–323.

49 1 B2 . . G T CC . C . C C -G
50 1 B2 T CC . C . C T C -G
51 2 B2 CC . C . C C -G C . GG . T
52 1 B2 CC . C . C T C -G A
53 2 B2 CC . C . CG C -G G C . C
54 1 B2 CC . C . C C -G A A
55 5 B2 . . G CC . C . C T -G C G
56 2 B2 CC . C . C T A . . A C -G
57 1 B2 CC . C . C A T -G C G
58 1 B2 CC . C . C A . C -G C
59 1 B2 CC . C . C C -G C G
60 1 1 B2 CC . C . C C -G C
61 1 B2 CC . C . C C -G T
62 2 B2 CC . C . C C -G AC G
63 2 1 B2 T . CC . C . C A C -G C
64 1 B2 . . . C A CC . C . C C -G C
65 1 B2 . . . T A CC . C . C C -G C
66 3 B2 C T C . CT . C T C -G T C
67 1 B2 C C . CY . C T C -G T C
68 1 B2 C C . C . C T A C -G T C
69 1 B2 T C . C . C C A C -G A CA
70 1 B2 T C . C . C A C -G C
71 2 B2 T C . C . C C -G C
72 3 B2 T C . C . C T T C -G C
73 1 B2 . . . C T C . C . C T C -G C
74 2 1 B2 C . C . C C C -G C
75 1 1 B2 C . C . C C -G G
76 1 B2 C . C . C T C . C CT - T -G
77 3 B2 C . C . C T C -G CT - T -G C
78 1 B2 T C . C . C C -G C
79 3 B2 C . C . C C -G AC G
80 2 B2 . . . C C . C . C G C -G AC G
81 2 B2 C . C . C T C -G AC G
82 1 B2 C . C . C C -G A
83 1 B2 T G C . C . C C -G A
84 4 B2 C . C . C G C -G A A
85 1 B2 C . C . C G C . . A -G A A
86 1 B2 C C . C . C C -G A C . T C
87 1 B2 T . C . C . C A A C CG - G C . C
88 2 B2 T . C . C . C A C -G C
89 1 B2 T . C . C . C N A C -G A AN
90 1 B2 C . C . C C -G CA
91 1 B2 C . C . C C C -G C
92 1 B2 C . C . C T C -G CT - T -G G
93 1 B2 C . C . C C -G T -G CA G
94 1 B2 -C . C . C C -G C CA
95 1 B2 . . G -T . C . C . C C -G C A G
96 1 B2 -T . C . C . C C -G C A G
97 1 B2 C . C . C C -G CT - C G
98 1 B2 T . C . C . C T C -G G C
99 1 B2 C . C . C C -G A A
100 3 B2 C . C . C C -G CT - C . C G
101 1 B2 A C . C . C C -G C C
102 2 B2 . . . C C . C . C C -G T -G C G
103 3 B2 C . C . C C -G C
104 1 B2 C . C . C G C -G C A AC
105 1 B2 C . C . C C -G A . AC C . C G
106 1 B2 C . C . C T . G C -G A A
107 1 B2 C . C . C C -G C C G
108 1 B2 C . C . TC C C -G C

109	1	B2 C . C . C C C - - - - -G
110	1	C1 C . C . . T C CT . T - - - - -G A
111	1	C1	. . G C CT - - - - -CT-T -G T
112	1	C1	. . G T C CT - - - - -G
113	4	C1	. . G T C CT - - - - -G
114	1	C1	. . G T C CT - - - - -G C
115	1	C1 T CG CT - - - - -G C
116	1	C1 T C CT C CW - - - - -G C C
117	3	C1 T C CT C C - - - - -G C C
118	1	C1 T C CT C C A - - - - -G C C
119	1	C1 T C CT C - - - - -G C
120	2	C1 C T T C CT C - - - - -G
121	1	C1 C T T C CT A . CC . C - - - - -G C
122	1	C1 A T TC C CT T C - - - - -G G . T
123	2	C1 T C CT A - - - - -G
124	1	2	C1 T C CT - - - - -GT
125	1	D1 T . . T T T C . T . C . . C . C C - - - - -G C A
126	1	D1 T T T C C CT - - - - -G
127	1	D1 T C C C CT - - - - -GG . G
128	1	D1 T . T C C C - - - - -G C
129	2	D1 T T C C C - - - - -G
130	3	D1 T C C C CT - - - - -G

H, Haplotype; A, Aymara population; Q, Quechua population; Hap, Haplogroup; CRS, Cambridge Reference Sequence (Anderson et al., 1981).

Table S2. Y_chromosome STR haplotypes in the Aymaras and Quechuas from Bolivia.

Haplotype	<i>n</i>		Haplogroup ^a	DYS	DYS	DYS							Y_GAT						
	A	Q		389-I	389-II	DYS390	DYS456	DYS19	385a-b	DYS458	DYS437	DYS438	DYS448	H4	DYS391	DYS392	DYS393	DYS439	DYS635
1	1		(R1b)	15	32	24	16	14	11-14	17	14	12	18	11	10	13	13	12	23
2		1	Q1a3a	14	32	23	15	13	15-19	17	14	11	20	12	10	16	14	13	22
3		1	Q1a3a	14	32	23	15	13	15-19	16	14	11	20	14	10	17	14	14	22
4	2		Q1a3a	14	32	23	15	13	15-19	16	14	11	20	12	10	16	14	14	22
5		1	Q1a3a	14	32	23	15	13	15-18	16	14	11	20	12	10	16	14	13	22
6		1	Q1a3a	14	32	21	15	13	14-18	17	14	11	21	12	10	14	13	13	22
7		1	Q1a3a	14	31	25	15	13	14-20	18	14	11	19	12	10	14	13	11	22
8		1	(R1b)	14	31	24	17	14	11-14	17	14	12	18	11	11	13	13	11	23
9	1		Q1a3a	14	31	24	15	13	16-17	18	14	12	20	12	11	14	13	11	22
10	1		Q1a3*	14	31	24	15	13	14-14	18	14	11	20	12	10	14	13	13	23
11	1		Q1a3a	14	31	24	15	13	14-14	17	14	11	20	12	10	14	13	13	23
12	1		Q1a3*	14	31	24	15	13	14-14	17	14	11	20	11	10	14	13	13	23
13		1	Q1a3a	14	31	23	16	13	15-20	16	14	11	19	12	10	16	14	14	22
14	1		Q1a3a	14	31	23	16	13	15-17	17	14	11	20	12	10	16	14	13	22
15	1		Q1a3*	14	31	23	15	13	16-19	17	14	11	20	12	10	16	14	13	22
16	2		Q1a3a	14	31	23	15	13	16-19	16	14	11	20	12	10	16	14	13	22
17	1		Q1a3a	14	31	23	15	13	15-20	16	14	11	20	12	10	16	14	13	22
18		1	Q1a3a	14	31	23	15	13	15-19	17	14	11	20	13	10	18	14	14	22
19		5	Q1a3a	14	31	23	15	13	15-19	17	14	11	20	12	10	18	14	14	22
20	1		Q1a3a	14	31	23	15	13	15-19	16	14	11	21	11	10	16	14	12	22
21	1		Q1a3a	14	31	23	15	13	15-19	16	14	11	20	13	10	16	14	14	22
22	3	1	Q1a3a	14	31	23	15	13	15-19	16	14	11	20	12	10	16	14	13	22
23		1	Q1a3a	14	31	23	15	13	15-19	16	14	11	20	12	10	16	14	12	22
24	1		Q1a3a	14	31	23	15	13	15-19	15	14	11	20	12	10	16	14	14	22
25	1		Q1a3a	14	31	23	15	13	15-19	15	14	11	20	12	10	16	14	13	22
26	1		Q1a3a	14	31	23	15	13	15-18	17	14	11	20	12	10	16	14	12	22
27		1	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	14	22
28	1		Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	13	23
29		3	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	13	22
30		2	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	16	14	12	22
31	1		Q1a3a	14	31	23	15	13	15-18	16	14	11	20	12	10	14	14	14	23

32	1	Q1a3a	14	31	23	15	13	15-18	16	14	11	20	11	10	15	14	13	22
33	2	Q1a3a/Q1a3*	14	31	23	15	13	15-17	16	14	11	20	12	10	16	14	14	22
34	2	Q1a3a	14	31	23	15	13	14-19	16	14	11	20	12	10	16	14	13	22
35	1	Q1a3a	14	31	23	15	13	14-19	16	14	11	20	12	10	16	14	12	22
36	1	Q1a3a	14	31	23	15	13	14-18	16	14	11	20	12	10	17	14	13	22
37	1	Q1a3a	14	31	23	15	13	14-14	19	14	11	21	12	10	15	13	12	23
38	1	Q1a3a	14	30	25	17	13	15-19	17	14	11	19	11	11	15	13	12	26
39	1	Q1a3a	14	30	25	15	13	14-19	17	14	9	20	12	10	14	13	13	22
40	1	(R1b)	14	30	24	15	15	11-14	18	15	12	19	12	10	13	12	12	23
41	1	(R1b)	14	30	23	17	14	11-11	18	15	12	19	12	10	13	14	13	23
42	1	Q1a3a	14	30	23	15	14	15-18	16	14	11	20	12	10	16	14	13	22
43	3	Q1a3a	14	30	21	13	13	14-16	17	14	11	20	11	11	14	14	12	22
44	1	Q1a3a	14	29	25	15	13	14-20	18	14	11	19	12	10	14	13	11	22
45	1	Q1a3a	13	32	24	15	13	17-17	13	14	11	21	11	10	14	14	11	22
46	1	Q1a3a	13	32	23	16	13	15-18	16	14	11	20	11	10	14	13	11	22
47	2	Q1a3a	13	31	24	16	13	16-18	15	14	11	19	11	11	14	13	12	22
48	2	Q1a3a	13	31	24	15	14	14-17	18	14	11	20	12	11	14	13	11	22
49	1	Q1a3a	13	31	24	15	13	14-17	17	14	11	19	12	11	14	13	11	22
50	1	Q1a3a	13	31	23	15	13	15-19	16	14	11	21	11	10	14	13	11	22
51	2	Q1a3a	13	30	25	15	14	16-17	13	14	11	20	12	10	14	13	13	23
52	1	Q1a3a	13	30	25	15	12	14-20	17	14	11	19	11	10	14	13	11	23
53	1	(E1b1b)	13	30	24	16	13	17-17	15	14	10	20	12	10	11	13	13	23
54	1	Q1a3a	13	30	24	15	14	15-20	16	14	10	19	12	10	14	13	12	22
55	2	Q1a3a	13	30	24	15	13	14-15	17	14	12	20	12	11	14	13	13	22
56	1	Q1a3a	13	30	24	15	13	14-14	18	14	12	20	12	10	14	13	13	22
57	2	Q1a3a	13	30	24	15	13	14-14	18	14	11	20	12	10	14	13	14	22
58	1	Q1a3a	13	30	24	14	13	15-17	17	14	11	19	12	10	14	13	12	22
59	1	(T)	13	30	23	15	15	16-17	16	14	9	19	11	11	15	13	11	21
60	1	Q1a3a	13	30	23	15	14	12-14	19	14	11	20	11	10	14	13	11	22
61	1	Q1a3a	13	30	23	15	13	15-19	16	14	11	20	12	10	16	14	13	22
62	1	Q1a3a	13	30	23	15	13	15-18	16	14	11	20	12	10	16	14	13	22
63	1	Q1a3a	13	30	23	15	13	15-17	17	14	11	20	12	10	16	14	13	22
64	4	Q1a3a	13	30	23	15	13	13-19	16	14	11	20	12	10	16	14	14	22
65	2	Q1a3a	13	30	23	15	13	13-18	16	14	11	20	12	10	16	14	13	22
66	1	Q1a3a	13	30	23	11	13	13-19	16	14	11	20	12	10	16	13	14	22

67	1	Q1a3*	13	29	25	16	13	15-19	16	14	11	19	11	9	14	13	12	22
68	2	Q1a3a	13	29	25	15	14	14-17	17	14	10	21	12	10	14	13	12	22
69	1	(R1b)	13	29	26	15	14	11-11	16	15	12	19	12	10	13	13	12	23
70	1	(R1b)	13	29	24	16	14	11-14	16	15	11	19	11	10	13	13	12	23
71	1	Q1a3a	13	29	24	16	13	14-16	16	14	11	21	12	12	14	13	12	22
72	1	Q1a3a	13	29	24	15	13	14-17	17	14	11	19	12	11	14	13	11	22
73	1	Q1a3*	13	29	24	15	13	11-16	18	14	11	20	12	10	14	13	13	23
74	1	(J1)	13	29	23	15	14	14-17	17	14	10	20	11	10	11	12	13	20
75	2	(R1b)	13	29	23	15	14	11-11	17	14	12	18	11	11	13	13	13	23
76	1	Q1a3a	13	28	24	15	13	15-18	17	14	11	20	12	11	14	13	11	22
77	1	Q1a3a	13	28	24	15	13	15-17	17	14	11	20	12	11	14	13	12	22
78	1	Q1a3a	13	28	24	15	13	14-17	17	14	11	20	12	11	14	13	11	22
79	1	Q1a3a	12	30	24	15	13	16-18	17	14	11	21	11	10	14	13	12	22
80	1	Q1a3a	12	30	24	14	15	14-16	17	14	11	19	12	10	14	13	12	24
81	1	Q1a3a	12	30	24	14	14	14-16	17	14	11	19	12	10	14	13	12	24
82	1	(E1b1a)	12	30	21	15	17	16-17	18	14	11	21	12	10	11	13	12	22
83	1	Q1a3a	12	29	23	15	13	16-19	16	14	11	20	11	10	16	14	13	22
84	1	Q1a3a	12	28	24	15	13	14-18	16	14	11	21	12	11	14	13	13	22
85	1	(J)	12	28	23	15	14	13-17	17	14	9	21	11	10	11	12	10	22
86	1	(II)	12	28	23	15	14	14-14	16	16	10	20	12	10	11	13	11	22
87	1	Q1a3a	12	28	21	16	13	14-19	18	14	11	20	12	10	14	14	12	22

^a The haplogroups have been assigned by SNP genotyping, and those in parentheses have been inferred from the web page Haplogroup Predictor (<http://www.hprg.com/hapest5/>).