AMERICAN JOURNAL OF HUMAN BIOLOGY 22:154–162 (2010)

Original Research Article

Autosomal and X Chromosome *Alu* Insertions in Bolivian Aymaras and Quechuas: Two Languages and One Genetic Pool

MAGDALENA GAYÀ-VIDAL,[1,2] JEAN-MICHEL DUGOUJON,[2] ESTHER ESTEBAN,[1] GEORGIOS ATHANASIADIS,[1] ARMANDO RODRÍGUEZ,[3] MERCEDES VILLENA,[3] RENÉ VASQUEZ,[4] AND PEDRO MORAL[1]*
[1]*Unitat d'Antropologia. Biologia Animal, Universitat de Barcelona, Barcelona, Spain*
[2]*Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, FRE 2960 CNRS et Université de Toulouse, Toulouse, France*
[3]*Instituto Boliviano de Biología de Altura (IBBA), La Paz, Bolivia*
[4]*Instituto Boliviano de Biología de Altura (IBBA), Potosí, Bolivia*

ABSTRACT Thirty-two polymorphic *Alu* insertions (18 autosomal and 14 from the X chromosome) were studied in 192 individuals from two Amerindian populations of the Bolivian Altiplano (Aymara and Quechua speakers: the two main Andean linguistic groups), to provide relevant information about their genetic relationships and demographic processes. The main objective was to determine from genetic data whether the expansion of the Quechua language into Bolivia could be associated with demographic (Inca migration of Quechua-speakers from Peru into Bolivia) or cultural (language imposition by the Inca Empire) processes. Allele frequencies were used to assess the genetic relationships between these two linguistic groups. Our results indicated that the two Bolivian samples showed a high genetic similarity for both sets of markers and were clearly differentiated from the two Peruvian Quechua samples available in the literature. Additionally, our data were compared with the available literature to determine the genetic and linguistic structure, and East–West differentiation in South America. The close genetic relationship between the two Bolivian samples and their differentiation from the Quechua-speakers from Peru suggests that the Quechua language expansion in Bolivia took place without any important demographic contribution. Moreover, no clear geographical or linguistic structure was found for the *Alu* variation among South Amerindians. Am. J. Hum. Biol. 22:154–162, 2010.

The Quechuas and the Aymaras are the two main Amerindian linguistic groups inhabiting the Andean Altiplano in Bolivia, an area where genetic studies have mainly focused on mtDNA (Corella et al., 2007; Sandoval et al., 2004). In a wider geographical context, most of the available genetic data on Andean populations derive from studies of uniparental markers, with small-sized samples, from populations geographically restricted to modern Peru (Fuselli et al., 2003; Lewis et al., 2007). The present study is focused on the genetic variability of these two main linguistic groups, through the analysis of significant-sized samples, to provide new autosomal data with a wide set of independent loci (32 *Alu* loci).

Archaeological and historical records suggest that modern Bolivian populations are the result of historic complex interactions among people of different languages and cultures. Most data point to the Central Andes (Bolivian Altiplano and Peru) as the heartland of the first complex societies of South America. It is commonly accepted that important civilizations/states such as Chavin (900–200 BC), Tiwanaku (100 BC–1200 AD), and Huari (700–1200 AD), existed before the establishment of the Inca Empire, which was conquered by the Spaniards around 1532 AD (Stanish, 2001). Specifically, in the South Central Andes, the Tiwanaku civilization, which originated in the Titicaca basin (in the Altiplano at 3,600 m a.s.l), extended its influence from Southern Peru to current Bolivia, Northern and Central Chile and North-Western Argentina, (Kolata, 1993). After the Tiwanaku collapse, the state fragmented into a number of Aymara polities or "*Señorios*" (Qolla, Lupaqa, Pakaq, Caranga, etc; see Bouysse-Cassagne, 1986) that persisted until their conquest by the Inca Empire (1300–1532 AD). From Cuzco, the Incas expanded its power towards 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 in the Andes, two main Amerindian languages of the Andean subfamily (Greenberg, 1987), the Quechua (12 million speakers in Ecuador, Peru, Southern Bolivia, and Northern Chile), and the Aymara (1.5 million speakers mainly in Bolivia), are spoken along with other minor languages, such as Uru-Chipaya which is spoken around the shores of Lake Titicaca and Lake Poopó. It is important to note that this linguistic distribution seems to be relatively recent. 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), whereas a proto-Aymara, Pukina, and Uru were probably spoken in the influence area of the Tiwanaku civilization (Browman, 1994; Kolata, 1993; Stanish, 2001). Afterwards, 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).

To gain new insights into the relationships between the two main Amerindian linguistic groups in Bolivia and the

Contract grant sponsor: Ministerio de Ciencia y Tecnología; Contract grant numbers: CGL2005-3391, CGL2008-03955; Contract grant sponsor: Generalitat de Catalunya; Contract grant number: 2005-SGR00252; Contract grant sponsor: Departament d'Educació i Universitats de la Generalitat de Catalunya i el Fons Social Europeu; Contract grant number: 2007_FI_B00108; Contract grant sponsors: "La Caixa", The French Government

*Correspondence to: Prof. Pedro Moral, Ph.D., Department of Biologia Animal-Antropologia. Fac. Biologia. Universitat de Barcelona, Av. Diagonal, 645. 08028 Barcelona, Spain. E-mail: pmoral@ub.edu

Received 25 March 2009; Revision received 5 May 2009; Accepted 30 May 2009

DOI 10.1002/ajhb.20967

Published online 10 July 2009 in Wiley InterScience (www.interscience.wiley.com).

demographic processes that may have affected these relationships, this study deals with the genetic variability of Aymara-speakers from the Titicaca Lake region and Quechua-speakers from the Northern Potosí department. The Bolivian Quechua population corresponds to the ancient Charaqara region, which was Aymara-speaking before the Inca expansion (Tschopik, 1963). Today, the Tomas Frias province, from where this sample originates, is 98% Quechua-speaker (Fabre, 2005). The aim is to assess the relative importance of the demographic and cultural processes of the Quechua expansion in Bolivia. Different studies have demonstrated that in many cases the use of a language by a population is a sign of genetic identity, but, in other cases it may simply be a cultural trait imposed by a political or economical power without any substantial effect on the genetic structure of the population itself (Belle and Barbujani, 2007; Cavalli-Sforza et al., 1992; Moral et al., 1994). The comparative analysis of our own data with other available data on some Peruvian Quechua-speaking populations, will allow us to determine the demographic implications of the movement of Quechua-speaking groups under the mitma system, under the assumption that Aymara-speakers were the main original inhabitants of the Bolivian Potosi department according to historical sources.

Our hypothesis is that if the Inca mitma system was an effective demographic process for the Andean southward expansion of the Quechua language, it might be recognizable through genetic similarity/difference patterns between current Bolivian Quechua-speakers and other northward Quechua populations from Peru. That is, if the mitma system was effective we can expect greater genetic differences between the two Bolivian groups than between Bolivian and Peruvian Quechua-speakers. The alternative scenario would predict an opposite pattern of genetic similarities. Additionally, our analysis will allow the genetic characterization of Bolivian populations in relation to other Native Americans, and provide new autosomal data to address general issues concerning the South American populations: the genetic East to West (or Amazon vs. Andes) differentiation suggested by some previous surveys (Lewis and Long, 2008; Tarazona-Santos et al., 2001); and the general correspondence between genetics and linguistics in this region (Hunley et al., 2007).

Population genetic analyses were performed using 32 Polymorphic *Alu* Insertions (also known as PAIs), the most abundant short interspersed nuclear elements (SINEs), representing more than 10% of the human genome (Carroll et al., 2001). Typically, an *Alu* insertion is a 300-bp-long sequence ancestrally derived from the 7SL RNA gene inserted into the genome through an intermediate RNA single strand generated by RNA polymerase III transcription. On the basis of the evolution changes into the original genes, these elements are grouped into subfamilies. Some members of the youngest subfamilies are not yet fixed in all human populations and consequently are polymorphic for the presence or absence of the insertion (Roy et al., 1999). These markers present two noteworthy features: (1) the insertion is identical by descent, and (2) the ancestral state is known. These characteristics make the *Alu* insertions a useful group of markers in the study of human population genetics (Cordaux et al., 2007; Resano et al., 2007).

As far as we know, previous data on *Alu* insertions variation in Native Americans range from a few loci (Antunez de Mayolo et al., 2002; Dornelles et al., 2004; Mateus-Pereira et al., 2005; Novick et al., 1998) to 12 loci (Battilana et al., 2006), that have been generally analyzed in population samples of quite small sizes. In the particular case of Andean populations, only two Peruvian Quechua groups have been previously tested (Battilana et al., 2006) but no *Alu* data are available on Aymaran populations. So, in relation to previous studies in the literature, this article represents: (i) the first *Alu* polymorphic survey carried out on Aymara populations, (ii) the first data on 14 X chromosome *Alu* polymorphic elements in Native Americans, and (iii) the first data on 8 out of the 18 autosomal PAI tested in this study in South Americans.

MATERIALS AND METHODS

Population samples

A total of 192 unrelated subjects originating from two linguistically different regions of Bolivia (96 from each population) were analyzed. These subjects were selected from a whole set of 686 individual samples according to available genealogical records. Blood samples were obtained with informed consent under the framework of the High-Altitude Adaptability Project of the IBBA (Instituto Boliviano de Biología de Altura) and with approval from the Ethical Committee of this institution. As an indicator of potential non-Native admixture, an analysis of the GM haplotypes showed around 1% of the specific European haplotype GM5*;3 (Dugoujon JM, personal communication). Also, the two samples analyzed here presented a frequency of 98% of the O group (ABO system).

The geographical location of the two samples studied is shown in Figure 1. The two population samples live in the Central Andes, in the Bolivian Altiplano. The Aymaran sample comes from two agricultural communities or "*Ayllus*" (an endogamous, patrilineal, corporate kin group), the Tuni and Amachuma, which are located 3-km apart between La Paz and the Titicaca Lake and show admixture of 25% (personal communication from pedigree data in Crognier et al., 2002). The Quechua-speaking subjects are inhabitants of rural areas from 13 ayllus near the Tinguipaya city, in the Potosi department.

Genotype determinations

DNA extracted from blood by classical phenol–chloroform method was used for PCR genotype determinations. Eighteen human-specific autosomal *Alu* polymorphic elements (ACE, APOA1, A25, B65, CD4, DM, D1, FXIIIB, PV92, TPA25, HS2.43, HS4.32, HS4.69, Sb19.12, Sb19.3, Yb8NBC120, Yb8NBC125, and Ya5NBC221) were genotyped using the primers and PCR conditions described with minor modifications (Gonzalez-Perez et al., 2003). From these 18 polymorphism, 8 PAIs (ACE, APOA1, FXIIIB, PV92, TPA25, D1, A25, and HS4.32) were selected to provide the best comparative data set regarding the published literature, whereas the remaining ones were included due to their discriminative power among local populations previously shown by other studies (Gonzalez-Perez et al., 2003, 2006; Resano et al., 2007). Additionally, 14 X chromosome *Alu* insertions (Ya5DP62, Ya5DP57, Yb8DP49, Ya5a2DP1, Yb8DP2, Ya5DP3, Ya5NBC37, Yd3JX437, Yb8NBC634, Ya5DP77, Ya5NBC491, Yb8NBC578, Ya5DP4, Ya5DP13) were also determined in each sample by using the primers and PCR

Fig. 1. Geographic location of the populations included into the analyses. 1: Aymara, 2: Quechua "Tin" (from Tinguipaya), 3: Aché, 4: Caingang, 5: Guaraní, 6: Xavante, 7: Cinta Larga, 8: Gavião, 9: Quechua "A" (from Arequipa), 10: Quechua "Tay" (from Tayacaja), 11: Surui, 12: Waiwai, 13: Zoró, 14: Yanomami, 15: Maya.

conditions according to Callinan et al. (2003), with minor modifications (Athanasiadis et al., 2007). Phenotypes were identified by electrophoresis of the PCR products, followed by ethidium bromide staining and observation under UV fluorescence. Positive and negative controls were used in all the PCR runs to assess the quality of the determinations.

Statistical analyses

Allele frequencies were computed by direct counting and Hardy-Weinberg equilibrium was tested by an exact test using the Genepop program (Raymond and Rousset, 1995) to assess data quality. Unbiased estimates of heterozygosity and its average across loci and populations were calculated according to the Nei's formula (Nei, 1978). For the X chromosome PAIs, H-W equilibrium and gene diversities were calculated from female genotype frequencies. The Bonferroni correction was applied in all analyses.

As a first approach to the genetic differentiation between the two Bolivian populations, an exact test based on the allele frequencies of all 32 individual loci was performed using Arlequin statistical package (Schneider et al., 2000).

For comparative purposes, *Alu* frequency data on 13 Native American populations were collected from the literature. These included 12 South Amerindian groups (Aché, Caingang, Guaraní, Cinta Larga, Gavião, Wai Wai, Xavante, Zoró, Quechua "A" from Arequipa, Quechua "Tay" from Tayacaja, Suruí, and Yanomami), and one Central-American population (Maya), whose geographical location is indicated in Figure 1. No North American samples were included in the comparisons due to their low number, high level of admixture, and their geographical irrelevance with our main hypothesis. For all these populations, data were available for 8 out of the 32 loci examined in the present study (Battilana et al., 2006). Using the joint variation in these loci, pairwise population relationships were determined by the analysis of the genetic distances using the Reynolds coefficient (Reynolds et al., 1983) for all the Amerindian populations available. These distance estimates were used (i) to quantify the genetic relationships between the two Bolivian linguistic groups in the framework of the relationships among other Native Americans, (ii) to compare the degree of genetic differentiation between the two Bolivian samples and other similar linguistic groups (Quechua) from Peru to obtain indirect evidence about the demographic impact of the Quechua expansion into Bolivia, and (iii) to approach the between-population variation in different geographical South American population groups (West vs. East). The distance relationships were depicted by a neighbor-joining tree (Saitou and Nei, 1987) and displayed in a Multi-Dimensional Scaling (MDS) graph. The reliability of the tree was tested by bootstrap resampling analysis (1,000 iterations).

The amount of genetic diversity in all Amerindian samples and in different sample groups according to geographical (West and East in South America) and linguistic criteria (Quechua-speakers) was assessed by the analysis of the molecular variance (AMOVA) of the allele frequencies using the Arlequin software (Schneider et al., 2000). Finally, the possible structuring of the genetic diversity in South America according to geography and linguistics was checked by hierarchical AMOVA analyses to test the potential general geography-genetics and linguistics-genetics correlations in South Native Americans.

RESULTS

Allele frequency distributions in Bolivia

Alu insertion frequencies for the 18 autosomal loci in the two Andean populations, the Aymaras and Quechuas from Bolivia, are shown in Table 1. From the 18 loci, 11 were polymorphic in both populations. Four loci were

TABLE 1. Autosomal PAI frequencies and gene diversities in two Bolivian samples and diversity range in other South American Native Populations (Battilana et al., 2006)

Autosomal PAIs	No. of chromosomes		Freq. insertion		Unbiased heterozygosity		Range *H* in Native S. Amer.
Populations	Aymara	Quechua	Aymara	Quechua	Aymara	Quechua	
ACE	170	188	0.853	0.809	0.252	0.311	0.000–0.476
HS4.32	186	186	0.473	0.419	0.501	0.490	0.185–0.470
FXIIIB	190	188	0.942	0.968	0.110	0.062	0.000–0.417
A25	192	182	0.094	0.104	0.171	0.188	0.000–0.365
D1	190	186	0.584	0.505	0.488	0.503	0.403–0.507
TPA 25	190	190	0.679	0.712	0.438	0.413	0.205–0.503
PV92	192	180	0.865	0.917	0.235	0.154	0.075–0.507
Yb8NBC120	192	184	0.630	0.582	0.469	0.489	
Sb19.3	192	184	0.635	0.636	0.466	0.466	
Yb8NBC125	188	184	0.011	0	0.021	0	
B65	192	182	0.219	0.374	0.344	0.471	
DM	192	184	0.031	0.044	0.061	0.084	
Ya5NBC221	192	178	1	1	0	0	
APOA1	192	172	1	1	0	0	0.000–0.131
Sb19.12	188	168	0	0	0	0	
CD4	192	180	1	1	0	0	
HS2.43	190	180	0	0.006	0	0.011	
HS4.69	192	164	0	0.018	0	0.036	
Average	–	–	–	–	0.198	0.204	

TABLE 2. X chromosome PAI frequencies, and gene diversities in the two studied populations and the range of Heterozygosities in other World Populations (Callinan et al., 2003, Athanasiadis et al., 2007)

X chromosome PAIs	No. of chromosomes		Freq. insertion		Unbiased heterozygosity		Range *H* in world populations
Populations	Aymara	Quechua	Aymara	Quechua	Aymara	Quechua	
Ya5NBC37	148	138	0.088	0.029	0.153	0.039	0.160–0.520
Ya5a2DP1	146	142	0.863	0.894	0.196	0.148	0.180–0.470
Yb8DP2	124	134	0.121	0.149	0.209	0.300	0.230–0.430
Yd3JX 437	152	139	0.520	0.554	0.503	0.481	0.080–0.500
Ya5DP62	127	100	0.921	0.860	0.172	0.132	0.080–0.430
Ya5DP77	137	109	0.628	0.624	0.478	0.390	0.000–0.500
Ya5DP57	42	140	1	0.993	0	0.019	0.060–0.410
Yb8DP49	147	144	1	0.986	0	0.037	0.080–0.380
Yb8NBC634	134	139	1	1	0	0	0.000–0.260
Ya5NBC491	131	133	1	1	0	0	0.000–0.500
Yb8NBC578	134	147	1	1	0	0	0.000–0.480
Ya5DP13	142	137	1	1	0	0	0.000–0.080
Ya5DP3	151	147	0	0.007	0	0.018	0.180–0.500
Ya5DP4	140	117	0	0	0	0	0.000–0.280
Average	–	–	–	–	0.122	0.113	0.075–0.377

monomorphic for either the *Alu* presence (Ya5NBC221, APOA1, and CD4) or absence (Sb19.12) of the insertion in the two samples. The absence of the *Alu* element was fixed for HS2.43 and HS4.69 in Aymaras and for Yb8NBC125 in Quechuas. All the polymorphic *Alu* frequency distributions (12 in Aymaras and 13 in Quechuas) fit the Hardy-Weinberg equilibrium. The autosomal *Alu* elements showing the highest gene diversities were HS4.32, D1, TPA25, Yb8NBC120, Sb19.3, and B65 (Table 1). The average heterozygosity for the 18 autosomal loci was ~0.2 (Aymaras: 0.198 and Quechuas: 0.204).

The *Alu* insertion frequencies for the 14 X chromosome loci are displayed in Table 2. Six of them were polymorphic (Ya5NBC37, Ya5a2DP1, Yb8DP2, Yd3JX437, Ya5DP62, Ya5DP77) in both populations. Four were monomorphic for the insertion (Yb8NBC634, Yb8NBC578, Ya5DP13, and Ya5NBC491) and one for the absence (Ya5DP4) in the two samples. Finally, both the insertion for Ya5DP57, Yb8DP49 *Alu* and the absence for Ya5DP3 were fixed in Aymaras. Tests in female samples indicated that most of the observed distributions agree with the H-W equilibrium conditions. Only Ya5DP62 genotype distribution was significant ($P = 0.002$) after Bonferroni correction in the Aymaran population sample.

Two X chromosome *Alu* elements (Yd3JX437, Ya5DP77) showed a high heterozygosity compared with previous data (Athanasiadis et al., 2007; Callinan et al., 2003). The remaining markers examined in the two Andean samples exhibited diversity values that lie close to the lowest values worldwide (Table 2). The average gene diversities (Aymaras: 0.122, Quechuas: 0.113) for the X chromosome PAIs in these Amerindian populations were also low according to the heterozygosity range found in other populations.

The exact test of differentiation between the two populations showed no significant difference for any locus distribution.

Genetic comparisons with other Native Americans

The frequency distribution of eight PAIs (Table 3) was used to estimate the genetic relationships through Rey-

TABLE 3. Population size, allele frequency distribution for the 8 loci, average heterozygosity (considering the 8 loci) and linguistic affiliations of the 15 populations considered

Populations[a]	Ling[b]	*n*	Average *H* (8 loci)	ACE+	APO+	TPA 25+	FXIIIB+	PV92+	A25+	HS4.32+	D1+
1. Aymara (1)	AND	96	0.274	0.853	1	0.679	0.942	0.865	0.094	0.473	0.584
2. Quechua Tin (1)	AND	96	0.265	0.808	1	0.711	0.968	0.917	0.104	0.419	0.505
3. Aché (2)	ET	31–76	0.206	1	1	0.866	0.782	0.855	0.013	0.198	0.581
4. Caingang (2)	GPC	40	0.297	0.543	0.963	0.675	0.872	0.793	0.037	0.250	0.706
5. Guarani (2)	ET	34	0.268	0.829	0.941	0.710	0.935	0.783	0.097	0.130	0.394
6. Xavante (2)	GPC	33	0.305	0.683	1	0.417	1	0.813	0.234	0.242	0.532
7. Cinta larga (3)	ET	25	0.268	0.820	0.960	0.438	0.938	0.538	0.000	0.125	0.283
8. Gavião (3)	ET	28	0.177	0.926	1	0.793	1	0.897	0.000	0.154	0.589
9. Quechua A (3)	AND	21	0.297	0.826	1	0.643	1	0.696	0.136	0.357	0.361
10.Quechua Tay (3)	AND	22	0.312	0.630	0.978	0.714	0.891	0.739	0.043	0.300	0.675
11. Surui (3)	ET	23	0.220	0.870	1	0.409	1	0.938	0.083	0.214	0.286
12. Waiwai (3)	GPC	22	0.205	0.976	1	0.778	0.870	0.870	0.043	0.100	0.543
13. Zoro (3)	ET	28	0.212	0.962	0.983	0.692	1	0.833	0.018	0.217	0.583
14.Yanomami (3)	CP	21	0.216	0.750	1	0.685	1	0.962	0.000	0.241	0.333
15. Maya (3)	May	27	0.312	0.673	0.964	0.643	0.875	0.704	0.000	0.268	0.346

[a]References: (1): Present study, (2): Battilana et al., 2002, (3): Battilana et al., 2006.
[b]Linguistic filiations according to Greenberg (1987): AND: Andean, ET: Equatorial-Tucanoan, GPC: Gê-Pano-Carib, Chib: Chibcan, May: Mayan.

TABLE 4. Pairwise genetic distances between Native Americans

Distances	Aymara	QuechTi	Aché	Cainga	Guaraní	Xavante	Cint.L	Gavião	QuechA	QuechTay	Surui	Waiwai	Zoro	Yanom
Aymara														
QuechTin	0.007													
Aché	0.079	0.079												
Caingang	0.071	0.070	*0.120*											
Guaraní	*0.071*	*0.053*	0.065	0.082										
Xavante	0.072	0.067	0.164	0.066	0.067									
Cint.L	0.153	0.149	0.188	0.150	0.068	0.092								
Gavião	0.069	0.059	*0.039*	0.100	0.046	0.124	*0.173*							
QuechA	0.042	*0.035*	0.108	0.095	*0.032*	0.054	0.064	0.090						
QuechTay	0.046	0.049	*0.091*	0.007	0.064	0.065	0.131	0.078	0.066					
Surui	0.108	0.089	0.180	0.160	0.069	0.061	*0.087*	0.140	0.066	0.148				
Waiwai	*0.084*	*0.075*	0.017	*0.112*	0.035	0.126	0.150	*0.018*	0.090	0.090	0.130			
Zoro	*0.044*	0.044	*0.046*	*0.093*	0.038	0.090	0.127	0.013	0.058	*0.067*	0.102	*0.025*		
Yanomami	0.071	0.040	0.115	0.100	0.036	0.085	*0.123*	0.072	*0.052*	0.088	*0.056*	0.084	0.069	
Maya	0.069	0.057	0.108	*0.061*	0.030	0.065	0.050	0.099	*0.027*	*0.047*	0.079	0.093	0.079	*0.042*

In italics, genetic distance values not significantly different from zero.

nolds's distances (Table 4). Distance errors (Table 4) indicated that around 76% of the distance values were significant. The highest distance was observed between Cinta Larga and the Aché South American groups (0.188). It is worth noting that the distance between the two Bolivian samples of this study was among the lowest values found (0.007). The average distance value between all pairs of South Americans was 0.082; the mean distance between groups of the Eastern region (10 samples) was 0.09, more than twice the value (0.04) of the Western (Andean) region (four samples). It is interesting to note that the distance between the two Bolivian samples examined is 11 times smaller than the average in South America, and 7 times smaller than the distance between any other pairs of Andean populations (range 0.035–0.066).

Population distance relationships were represented through a neighbor-joining tree (see Fig. 2). This tree highlights the similarity between the two Bolivian populations of the present study, grouping them into a tight cluster, clearly differentiated from the rest. The Zoró, Gavião, Aché, and WaiWai South Amerindian populations form another cluster. The Amazon population of Cinta Larga appeared as the most differentiated. Interestingly, the three Quechua samples appeared clearly separated in the tree in spite of sharing the same language and geographical proximity. In an attempt to avoid the dichotomy implied in the tree construction, a MDS analysis (see Fig. 3) was performed that illustrates the close position of the two Bolivian groups. The rest of South American populations appeared scattered in the plot, showing a distribution pattern similar to the tree topology.

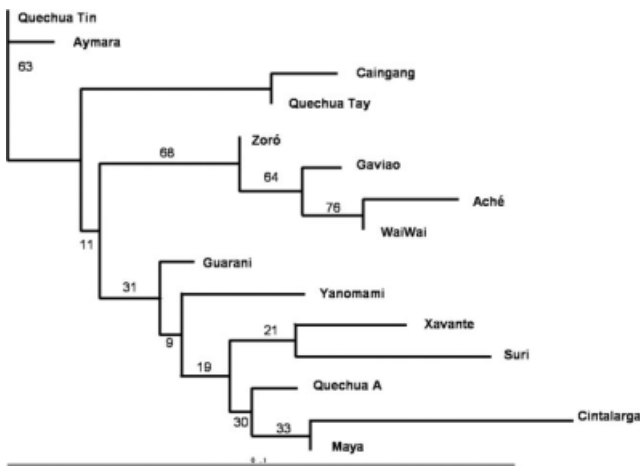

Fig. 2. Neighbor-joining tree obtained from Reynolds's distances. Bootstrap values based on 1,000 replications.

Genetic structuring

A global analysis of the allele frequency variance in Central and South America indicated a significant variation of the PAI markers (Table 5). The global F_{st} for the 15 populations considered showed that ~5% ($P < 0.001$) of the variation could be ascribed to between-population differentiation. A value around 7% was found among South Americans.

Among the Andean populations, the analysis of the genetic variance showed a high similarity between the two Bolivian samples in this study ($F_{st} = -0.003$, $P = 0.91$), whereas the global F_{st} for the three Quechua populations (0.031) was statistically significant ($P = 0.014$).

A hierarchical F_{st} analysis was performed grouping the South American populations according to geographical criteria into two groups: the Western region (Aymara, Quechua Tinguipaya, Quechua Arequipa, and Quechua Tayacaja) and the Eastern region (Aché, Caingang, Guaraní, Cinta Larga, Gavião, WaiWai, Xavante, Zoró, Suruí). In this context, the total between-population diversity ($F_{st} = 0.049$, $P < 0.0001$) can be almost completely explained by the diversity within groups ($F_{sc} = 0.043$, $P < 0.0001$) indicating the absence of geographic structure of the PAI data in South America.

According to linguistic criteria (Greenberg, 1987) the South American population samples were grouped into three of the four linguistic subfamilies: Gê-Pano-Carib languages (Caingang, Xavante, Wai Wai), Equatorial-Tucanoan languages (Aché, Guaraní, Cinta Larga, Gavião, Surui, Zoró), and Andean languages (Aymara, Quechua Tinguipaya, Quechua Arequipa, and Quechua Tayacaja). In this analysis, the Yanomani population was not included because it belongs to a different linguistic subfamily (Chibchan-Paezan). As in the former result, the most important part of the diversity between populations ($F_{st} = 0.051$, $P < 0.0001$) can be attributed to the diversity within groups ($F_{sc} = 0.040$, $P < 0.0001$).

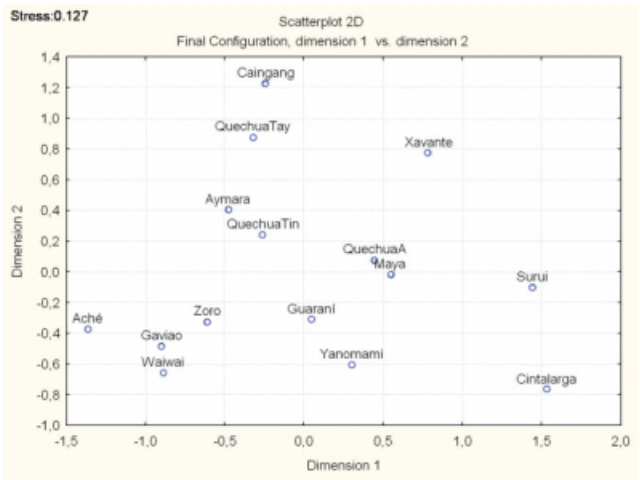

Fig. 3. Multidimensional scaling of Native Americans from Reynolds's distances. Raw stress value was 12.7%. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

DISCUSSION

The analysis of 32 *Alu* polymorphic insertions presented in this study allowed the determination of the genetic characterization of the two main linguistic groups from Bolivia, Aymara, and Quechua, and supplies new data on Native American genetic variability. In fact, so far as we know, the 14 X chromosome *Alu*s have been used for the first time in Amerindians, as well as 8 out of the 18 autosomal PAIs in South Amerindians. Moreover, an Aymara population has never previously been characterized for these markers.

Alu genetic features of the current Bolivian populations

Concerning the distinctiveness/characterization of the Native-American populations based on autosomal *Alu* frequency distributions, the two Bolivian populations show allele frequency patterns similar to other South Amerindian populations for the 10 markers for which data are available, (Antunez de Mayolo et al., 2002; Battilana et al., 2006; Dornelles et al., 2004; Mateus-Pereira et al., 2005; Novick et al., 1998; Tishkoff et al., 1996, 1998), except for the HS4.32 locus that displays the highest insertion frequencies in our study (Aymaras: 0.473, Quechuas: 0.419). For the eight autosomal loci examined for the first time in Native South Americans, it is interesting to note the extreme frequency values found in five loci near the fixation, for both absence (Yb8NBC125, HS2.43, HS4.69, Sb19.12) and presence (Ya5NBC221), as compared with other continents. The remaining three loci (Yb8NBC120, Sb19.3, and B65) present intermediate frequencies in relation to other populations. In general, our results for the eight PAIs previously studied in Amerindians are consistent with the pattern proposed by some authors indicating higher insertion frequencies in Native Americans and Asians than Africans (Mateus-Pereira et al., 2005; Stoneking et al., 1997); however, this trend is not clear for the remaining PAIs analyzed in this study.

Gene diversity variation for the eight *Alu* loci tested so far in South American populations appears to be remarkably high (Battilana et al., 2006; Novick et al., 1998; and present study). Some loci present a large heterozygosity

TABLE 5. Alu *frequency variance analyses in Native Americans*

Nonhierarchical analyses	n	F_{st}	Hierarchical F_{st} analyses			
			Population groups	Within groups	Among groups	Total F_{st}
Native Americans	15	0.045***	Geography: West (4)/East (10)	0.043***	0.006*	0.049***
South Americans	14	0.066***				
Western populations	4	0.015*	Linguistics: Gê-Pano-Carib (3)/	0.040***	0.014*	0.051***
Quechua populations	3	0.031*	Equatorial-Tucanoan (6)/Andean (4)			
Eastern populations	10	0.054***				

*$P < 0.05$; ***$P < 0.001$.

range, for example ACE (H = 0.0–0.5), PV92 (H = 0.0–0.5), and FXIIIB (H = 0.0–0.49). Other loci have a moderate range, for example Hs4.32 (H = 0.19–0.5), A25 (H = 0.0–0.36), and TPA25 (H = 0.21–0.5). For these loci the Bolivian populations are generally presented in high diversity values. In summary, 6 out of 11 of the polymorphic autosomal *Alu* markers in the two Bolivian samples exhibited important gene diversity higher than 0.4 (HS4.32, TPA25, Sb19.3, Yb8NBC120, D1, and B65).

The average gene diversity found in the two populations from the Bolivian Altiplano (H = 0.20) is similar to that described in other South Amerindian populations (H = 0.25; Battilana et al., 2006). Our data are consistent with a general worldwide trend that South Amerindians are the populations with the lowest heterozygosities, followed by Europeans (H = 0.29). This fact could most likely be explained by the genetic drift and bottleneck processes that occurred during the peopling of America, and especially of South America (Watkins et al., 2003).

Of interest is the frequency distribution of the PAIs on the X chromosome. For three loci (Ya5DP57, Yb8DP49, and Ya5a2DP1) Amerindians presented the highest insertion allele frequencies, whereas the Ya5NBC37 and Ya5DP3 Amerindian frequencies are the lowest so far reported in any human population, (Athanasiadis et al., 2007; Callinan et al., 2003). The highest allele frequency differences between the two Bolivian populations corresponded to the Ya5NBC37 locus.

The heterozygosities for most of the X chromosome *Alu* elements followed the general pattern previously described (Athanasiadis et al., 2007; Callinan et al., 2003) with an overall trend towards values lower than those for autosomal PAIs (average values of 0.12 vs. 0.20 in Bolivian populations) according to their chromosomal location. It is interesting to note that two PAIs (Ya5DP77 and Yd3JX437) showed the highest diversity values in the two Bolivian populations like in other African and Asian populations, in contrast with Europeans.

Linguistics vs. genetics in current Bolivian populations

One of the most evident results of this study is the high genetic similarity between the Aymara and Quechua linguistic groups from Bolivia. The two population samples showed very similar allele frequency distributions for the 32 loci analyzed. This close genetic similarity between the two Bolivian groups was also confirmed by the genetic distance and AMOVA analyses. In contrast, the comparison of the two Bolivian samples with other Andean groups underlines the genetic differentiation between Bolivian and Peruvian Quechua-speakers, showing genetic distances seven times higher than those between Aymaras and Quechuas from Bolivia. The high genetic similarity between the two Bolivian samples along with their clear differentiation from other Quechua-speaker peoples from Peru suggests a common genetic origin for the two main linguistic groups in the Bolivian Altiplano. This interpretation implies that the Quechua language expansion under the Inca power into the Bolivian Altiplano was due to cultural diffusion. However, an alternative explanation is also possible. A Quechua language expansion may have also been associated with an early movement of genetically different Quechua speaking people, and that the genetic signature of this movement was erased by subsequent gene flow from original local populations. This explanation is consistent with historical records describing frequent population movements in the Central Andes region during the Inca Empire and afterwards (Platt et al., 2006). However, it seems improbable that gene flow completely erased all genetic signatures in a relatively limited time period (around 500 years) unless an extremely high rate of gene flow was assumed. According to our demographic hypothesis, lower distance values would be expected on comparison of Bolivian vs. Peruvian Quechua-speakers (especially with Peruvian Quechua-speakers from Arequipa who share the same Quechua dialect; Cerron-Palomino, 2003), than to Aymara vs. Peruvian Quechua distances. Nevertheless, the result that the genetic distances of Bolivian Quechuas to Peruvian Quechua-speakers equal those of Aymaras supports that an erasure had to have been complete. On the other hand, the genetic distance between the two Peruvian Quechua-speaking groups is nine times higher than between the two Bolivian samples of this study, suggesting that gene flow in the Central Andes has not been high enough to erase all genetic differences between population groups. In general, the persistence of a certain degree of population divergence in the whole Andean region is shown by the variance of the *Alu* frequencies and is consistent with historically demonstrated (moderate) gene flow that has not completely eliminated the genetic particularities despite the important cultural integration undertaken by the Inca Empire and subsequently by the Spaniards.

Alu-based relationships among Native South Americans

Autosomal *Alu* variation is consistent with significant between-population diversity among South Americans. The N-J tree, MSD graph, and the AMOVA analysis fail to indicate strong clustering according to either geographical or linguistic criteria. However, the average genetic distances seem to indicate a different pattern of variation between the East and West regions of South America. The eastern populations show larger genetic distances and frequency variance than the western ones. Also the high *Alu* heterozygosities found in the Andean region seem to agree with higher within-population diversity as compared with the Eastern region. This could be consistent with different patterns of drift and gene flow, suggested elsewhere from mtDNA (Fuselli et al., 2003; Merriwether et al., 1995), Y chromosome (Tarazona-Santos et al., 2001), classical markers (Luiselli et al., 2000), and STR data (Wang et al., 2007). The Alu-based heterogeneity found in the Eastern South American populations is in agreement with other studies (Lewis and Long, 2008), indicating that they do not appear as a cohesive genetic group. This between-population higher diversity in the East is consistent with the suggested demographical scenario of lower effective population sizes in the East as compared with the West (Fuselli et al., 2003; Tarazona-Santos et al., 2001). Nevertheless, few and uneven population groups (10 from East vs. 4 from West); most of them exhibiting very low sample sizes do not allow a robust test of this hypothesis.

Although a detailed analysis of the correlation between linguistics and genetics in South Native Americans falls out of the scope of this study, it is worth noting that the simple approach of using genetic distance and frequency variance analyses indicates a clear absence of such a correlation. This result is consistent with some previous reports which revealed a positive correlation at a lan-

guage level (Fagundes et al., 2002; Mateus-Pereira et al., 2005) and at a stock level (Mateus-Pereira et al., 2005), but none at a phyla level using the Loukotka language classification. This level corresponds to the linguistic subfamily level considered in the present work according to the Greenberg classification. The controversial results in the literature highlight the complexity of this subject, as discussed recently (Hunley et al., 2007). According to these authors, the observed absence of correlation can be expected considering deep linguistic branches of the Greenberg's classification. In this context, our results indicate that the autosomal *Alu* variation analyzed confirms the absence of genetic-linguistic congruence regarding these linguistic subfamilies in South Native Americans.

CONCLUSIONS

This genetic analysis confirmed the importance of using autosomal genetic markers, such as *Alu* insertions, to unravel the history of human populations. This work underlined the importance of new studies on additional populations to complete the genetic picture of the Andean and South American populations. Finally, this study has revealed the genetic similarity between Bolivian populations belonging to the two main linguistic groups of the region (Aymara and Quechua), reaffirming that languages may not be congruent with the genetic features of the populations. In this sense, the Quechua language, though the main language in the Andean region, is not a safe indicator of the genetic identity of this region.

ACKNOWLEDGMENTS

The authors are grateful to the Service de Coopération et d'Action Culturelle (in particular they thank Marie-Claude Senemaud) of French embassy in La Paz, and to Mr. Alain Fouquet (the French ambassador). This field work received support from Universidad Mayor de San Andres, La Paz. Genealogical information was collected by IBBA, with the help of Emile Crognier and André Sevin (CNRS).

LITERATURE CITED

Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ. 2002. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. Electrophoresis 23:3346–3356.

Athanasiadis G, Esteban E, Via M, Dugoujon JM, Moschonas N, Chaabani H, Moral P. 2007. The X chromosome Alu insertions as a tool for human population genetics: data from European and African human groups. Eur J Hum Genet 15:578–583.

Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Tsuneto LT, Petzl-Erler ML, Salzano FM. 2002. Alu insertions versus blood group plus protein genetic variability in four Amerindian populations. Ann Hum Biol 29:334–347.

Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvshin N, Krylov M, Benevolenskaia L, Arnett FC, Batzer MA, Deininger PL, Salzano FM, Bonatto SL. 2006. Alu insertion polymorphisms in Native Americans and related Asian populations. Ann Hum Biol 33:142–160.

Belle EM, Barbujani G. 2007. Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity. Am J Phys Anthropol 133:1137–1146.

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.

Callinan PA, Hedges DJ, Salem AH, Xing J, Walker JA, Garber RK, Watkins WS, Bamshad MJ, Jorde LB, Batzer MA. 2003. Comprehensive analysis of Alu-associated diversity on the human sex chromosomes. Gene 317:103–110.

Carroll ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, Vincent B, Myers J, Ahmad Z, Nguyen L, Sammarco M, Watkins WS, Henke J, Makalowski W, Jorde LB, Deininger PL, Batzer MA. 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J Mol Biol 311:17–40.

Cavalli-Sforza LL, Minch E, Mountain JL. 1992. Coevolution of genes and languages revisited. Proc Natl Acad Sci USA 89:5620.

Cerron-Palomino R. 2003. Lingüística Quechua. Centro Bartolomé de las Casas: Cuzco, Peru.

Cordaux R, Srikanta D, Lee J, Stoneking M, Batzer MA. 2007. In search of polymorphic Alu insertions with restricted geographic distributions. Genomics 90:154–158.

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.

Crognier E, Villena M, Vargas E. 2002. Helping patterns and reproductive success in Aymara communities. Am J Hum Biol 14:372–379.

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.

Fabre A. 2005. Diccionario etnolingüístico y guía bibliográfica de los pueblos indígenas sudamericanos.

Fagundes NJR, Bonatto SL, Callegari-Jacques SM, Salzano FM. 2002. Genetic, geographic, and linguistic variation among South American Indians: possible sex influence. Am J Phys Anthropol 117:68–78.

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.

Gonzalez-Perez E, Via M, Esteban E, López-Alomar A, Mazieres S, Harich N, Kandil M, Dugoujon JM, Moral P. 2003. Alu insertions in the Iberian Peninsula and North West Africa—genetic boundaries or melting pot? Coll Antropol 27:491–500.

Gonzalez-Perez E, Esteban E, Via M, García-Moro C, Hernández M, Moral P. 2006. Genetic change in the polynesian population of Easter Island: evidence from Alu insertion polymorphisms. Ann Hum Genet 70:829–840.

Greenberg JH. 1987. Language in the Americas. Stanford: Stanford University Press.

Hunley KL, Cabana GS, Merriwether DA, Long JC. 2007. A formal test of linguistic and genetic coevolution in Native Central and South America. Am J Phys Anthropol 132:622–631.

Kolata A. 1993. The Tiwanaku: portrait of an Andean civilization. The people of America. Cambridge: Blackwell Publishers.

Lewis CM, Long JC. 2008. Native South American genetic structure and prehistory inferred from hierarchical modelling of mtDNA. Mol Biol Evol 25:478–486.

Lewis CM, 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.

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.

Mateus-Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM, Herrera RJ. 2005. Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. Am J Phys Anthropol 128:171–184.

Merriwether DA, Rothhammer F, Ferrell RE. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. Am J Phys Anthropol 98:411–430.

Moral P, Marogna G, Salis M, Succa V, Vona G. 1994. Genetic data on Alghero population (Sardinia): contrast between biological and cultural evidence. Am J Phys Anthropol 93:441–453.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez-de-Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. 1998. Polymprphic Alu insertions and the Asian origin of Native American populations. Hum Biol 70:23–39.

Platt T, Bouysse-Cassagne T, Harris O. 2006. Qaraqara-Charka, Mallku, Inka y Rey en la provincia de Charcas (siglos XV-XVII). Historia antropológica de una confederación aymara. Institut français d'études

andines—IFEA; Plural editores; University of St. Andrews; University of London; Interamerican Foundation; Fundación Cultural del Banco Central de Bolivia, La Paz.

Raymond M, Rousset F. 1995. GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. J Hered 86:248–249.

Resano M, Esteban E, González-Pérez E, Vía M, Athanasiadis G, Avena S, Goicoechea A, Bartomioli M, Fernández V, Cabrera A, Dejean C, Carnese F, Moral P. 2007. How many populations set foot through the Patagonian door? Genetic composition of the current population of Bahía Blanca (Argentina) based on data from 19 Alu polymorphisms. Am J Hum Biol 19:827–835.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779.

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.

Roy AM, Carroll ML, Kass DH, Nguyen SV, Salem AH, Batzer MA, Deininger PL. 1999. Recently integrated human Alu repeats: finding needles in the haystack. Genetica 107:149–161.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Sandoval J, Delgado B, Rivas L, Bonilla B, Nugent D, Fujita R. 2004. Variants of mtDNA among islanders of the lake Titicaca: highest frequency of haplotype B1 and evidence of founder effect. Rev Peru Biol 11:161–168.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin: a software for population genetics data analysis. Ver 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.

Stanish C. 2001. The origin of state societies in South America. Annu Rev Anthropol 30:41–64.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071.

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.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M, Pääbo S, Watson E, Risch N, Jenkins T, Kidd KK. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern humans origins. Science 271:1380–1387.

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, Kidd KK. 1998. A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402.

Tschopik H. 1963. The Aymara. In: Steward JH, editor. Handbook of South American Indians, Vol. 2. New York: Cooper Square. p 501–573.

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.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, Reddy PG, Das PK, Batzer MA, Jorde LB. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res 13:1607–1618.