Mol Biol Rep (2014) 41:1533–1543
DOI 10.1007/s11033-013-2999-z

Analysis of admixture and genetic structure of two Native American groups of Southern Argentinean Patagonia

Andrea Sala · Daniel Corach

Received: 22 May 2013 / Accepted: 28 December 2013 / Published online: 14 January 2014

Abstract Argentinean Patagonia is inhabited by people that live principally in urban areas and by small isolated groups of individuals that belong to indigenous aboriginal groups; this territory exhibits the lowest population density of the country. Mapuche and Tehuelche (*Mapudungun* linguistic branch), are the only extant Native American groups that inhabit the Argentinean Patagonian provinces of Río Negro and Chubut. Fifteen autosomal STRs, 17 Y-STRs, mtDNA full length control region sequence and two sets of Y and mtDNA-coding region SNPs were analyzed in a set of 434 unrelated individuals. The sample set included two aboriginal groups, a group of individuals whose family name included Native American linguistic root and urban samples from Chubut, Río Negro and Buenos Aires provinces of Argentina. Specific Y Amerindian haplogroup Q1 was found in 87.5 % in Mapuche and 58.82 % in Tehuelche, while the Amerindian mtDNA haplogroups were present in all the aboriginal sample contributors investigated. Admixture analysis performed by means of autosomal and Y-STRs showed the highest degree of admixture in individuals carrying Mapuche surnames, followed by urban populations, and finally by isolated Native American populations as less degree of admixture. The study provided novel genetic information about the Mapuche and Tehuelche people and allowed us to establish a genetic correlation among individuals with Mapudungun surnames that demonstrates not only a linguistic but also a genetic relationship to the isolated aboriginal communities, representing a suitable proxy indicator for assessing genealogical background.

Keywords Argentinean Patagonia · Mapuche · Tehuelche · Autosomal, mt-DNA and Y-genetic markers · Genetic admixture · *Mapudungun*

Daniel Corach and Andrea Sala are members of National Research Council of Argentina (Carrera del Investigador del Consejo Nacional de Investigaciones Científicas y Tecnológicas, CONICET).

Electronic supplementary material The online version of this article (doi:10.1007/s11033-013-2999-z) contains supplementary material, which is available to authorized users.

A. Sala (✉) · D. Corach (✉)
Servicio de Huellas Digitales Genéticas (UBA) and Cátedra de Genética y Biología Molecular, School of Pharmacy and Biochemistry, Universidad de Buenos Aires, Junin 956 (1113), Buenos Aires, Argentina
e-mail: asala@ffyb.uba.ar

D. Corach
e-mail: dcorach@yahoo.com

A. Sala · D. Corach
Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina

Introduction

It has been demonstrated that the extant population of Argentina is heterogeneous and the result of admixture events involving Native Americans, Europeans and Sub-Saharan Western Africans, whose contributions differ in accordance with the geographical regions considered [1–3]. The Patagonian region of Argentina is delimited to the north by the Colorado River and to the south by Tierra del Fuego; it comprises seven provinces inhabited by 2,089,100 people in an extension of 903,446 km^2 (population density 2.31 inhabitants/km^2). Most urban cities have reduced population sizes, only four have over 100,000 inhabitants, all of whom where settled during the first half of the twentieth century. In 1878, Argentinean National Army intensified the hostilities against the tribal groups

and was called the Second Campaign to the Desert (the First Campaign took place in 1830 but was less extended). Their intention was to exterminate all native tribes and to deprive them of their lands with the excuse of extend the national territories. Between 1821 and 1899 over 20,000 natives were killed; some of them survived and some others moved to urban areas forced to work as farm servants or to become national army soldiers [4].

Nowadays, Mapuche and Tehuelche represent the only aboriginal groups that inhabit the region. The former group is, at present, the most numerous of Argentina with over 113,000 members (census 2001) and is one of the 30 different aboriginal groups that are presently recognized in the country. Their extended distribution range includes the Patagonian provinces of Neuquén, Río Negro, Chubut, and Santa Cruz and to a lesser extent, the Pampean provinces of Buenos Aires, La Pampa and Córdoba. Patagonian Mapuche populations consist of over 78,000 individuals (approximately 13 % of the total aboriginal population) 74.8 % of whom are Mapuche or first-generation descendants, whereas a small proportion (17 %) live in small isolated communities. The origin of the Mapuche peoples remains unclear, but it is accepted that they received an important influence from the Chilean Araucanian; therefore, their origins might be the result of admixture processes among populations inhabiting both sides of the Southern Andes [5, 6]. Before the dispersion that allowed expanding their territories, the Mapuche developed a culture and a language called *Mapudungun* (*mapu* meaning "earth" and *dungun* "to speak"). Its linguistic affinities have been correlated with the Quechuan linguistic branch [7, 8]. Ruhlen and Greenberg [9] classified it within the Andean Southern linguistic sub branch. Its origin probably precedes the arrival of Spanish settlers in Southern South America in the early sixteenth century and was restricted to its oral expression, as most Amerindian languages.

Family lineages adopted words to identify their family members. Before evangelization, Mapuche names were composed of two elements: a prefix equivalent to the given name and a suffix to designate the ancestral name; therefore, the name *Calfucurá* meaning Blue (*calfú*) Stone (*curá*), includes the equivalent to a given name and a family or ancestral name. As part of the evangelization process, baptism was imposed and priests used a Catholic name and the combined original name and surname as a unique word, for instance, Calfucurá in the above example. These ancestral names were called "*kuga*" or "*cuga*" and were transmitted by the father to his children. However, there are different transmission models proposed: matrilineal [10], patrilineal [11], and selective, in which the father transmitted the *kuga* to some of his children born from his first wife and the other wives transmitted their family names to their descendants [12].

The Tehuelche inhabit Chubut and Santa Cruz provinces, and Buenos Aires metropolitan areas in a reduced number. Based on the last census, the total number of Patagonian individuals belonging to the Tehuelche people is 4,351, 7 % of whom are still living in isolated communities. The term Tehuelche is a modification of the composed Mapudungun word *Chehuelche*, *cheuel* meaning brave and *che*, people. Their original language could have belonged to the "*tshon*" root, according to Lehmann-Nitsche [13]. Their anthropological classification is confusing and variable according to the different authors; for Escalada and Casamiquela [14–16] the Tehuelche belong to the "*Tehuelche Complex*". This Complex includes the *Septentrional Tehuelches* (inhabiting the Pampas and Northern Patagonian areas) and the *Meridional Tehuelches* (*Patagones* in Southern Patagonia and *Onas* in Tierra del Fuego). During the eighteenth century, the Tehuelche were affected by two migratory events: the arrival of the Spanish settlers, through the north, and of the Araucans from Chile through the Andes (a process called "*araucanization*") in the west; the invasion also imposed the Mapuche language to the Tehuelche. These two events resulted in the execution of many "caciques" (leaders) and the limitation in their cultural development. It also gave rise to an admixed population involving Araucanian men and Tehuelche women.

The aim of this work was to characterize Mapuche and Tehuelche individuals inhabiting in aboriginal communities by investigating DNA polymorphisms located in autosomal, mitochondrial and Y-chromosomes in order to establish their genetic relationship with a set of urban population samples from Río Negro, Chubut and Buenos Aires provinces. Admixture analysis was conducted in order to establish the relationship of these people with extant population of urban areas of southern Argentina, with special attention to people with *Mapudungun* surnames. This investigation could provide novel information about two of the most important southern south Native American groups of Argentina.

Materials and methods

Subjects

Most samples were collected from healthy voluntary donors participating in paternity tests, after signing a written consent statement. The Bioethics Committee of the School of Pharmacy and Biochemistry of the University of Buenos Aires, Argentina, approved the research project.

Native American samples

Mapuche people ("MA", $N = 39$) inhabiting three communities: Cerro Policía, Aguada Guzman and LoncoVaca

Table 1 Description of the sample sets considered for the study

Community—Locality	Geographical location	Ethnicity	Reference	N^a	Code[b]
Cerro Policía, Aguada Guzmán, Lonco Vaca Arriba, Río Negro province	68°37′W, 39°10′S;68°57′W, 39°30′S and 69°05′W, 40°11′S, respectively	Mapuche	Present study	39	MA
El Chalía, Chubut province	70°15′W, 45°41′S	Tehuelche	Present study	28	TE
Río Negro and Chubut provinces, Argentina	Randomly sampled	*Mapudungun* aboriginal surnames	Present study	67	AS
Buenos Aires province, Argentina	Randomly sampled in metropolitan urban areas	Euro-Argentine	Present study	100	BA
Río Negro province	Randomly sampled in metropolitan urban areas	Euro-Argentine	Present study	100	RN
Chubut province	Randomly sampled in metropolitan urban areas	Euro-Argentine	Present study	100	CHU
Anecón Grande, Río Negro province	70°22′W, 41°20′S	Mapuche	[43]	39	MG
Trapa Trapa, BioBioprovince, Chile	71°16′W; 37°43′S	Pehuenche	[44]	24	PE
Huapi Island, Valdivia province, Chile	72°25′W; 40°15′S	Mapuche	[44]	34	HU
Puerto Williams, Antartica province, Chile	67°40′W; 55°S	Yaghan	[44]	15	YG

[a] Number of individuals in the sample set

[b] Refers to the code employed in the present study

Arriba, Río Negro province, and Tehuelche ("TE", $N = 28$) from the locality of El Chalía, Chubut province. These samples were kindly provided in 1993 by Dr. Raul Carnese and were previously characterized by different genetic marker systems [17]. When these samples were obtained ethical issues were not so stringent as nowadays, hence no written consent statement was signed.

Urban samples

- *Aboriginal surname samples*: Sixty-seven samples from Chubut and Río Negro provinces of unrelated males bearing aboriginal surnames (belonging to the *Mapudungun* language), coded as "AS" ("AS", $N = 67$).
- *European surname samples*: from Chubut ("CHU", $N = 100$), Río Negro (RN, $N = 100$) and Buenos Aires ("BA", $N = 100$) provinces. Once selected, all the samples were treated anonymously. Table 1 summarizes the sample size, the collection sites as well as the previously published results used for comparisons (see Supplementary Map 1 for geographical information).

Analytical methods

DNA extraction

DNA was extracted either from blood samples spotted onto Whatman 3MM paper using conventional protocols or from liquid saliva as previously described [18].

Autosomal STRs

A set of 15 STRs, including Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenin, Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818 combined within PowerPlex 16 System (Promega Corp. Madison, USA) was analysed according to the manufacturer's protocol.

Y-chromosome STRs

Mapuche and Tehuelche samples were analyzed by means of 17 Y-STRs including: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385I/II, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 and Y GATAH4. This set is included within the commercial Y-Filer kit (Life Technologies, Applied Biosystems, Foster City, USA). The AS samples were typed only for the minimal Y-haplotype, which includes DYS385I/II, DYS389I and II, DYS19, DYS390, DYS391, DYS392 and DYS393. Amplification and electrophoresis conditions were carried out as previously described [19]. All haplotypes included in this study have been deposited in the Y-Chromosome Haplotype Reference Database (YHRD) [20].

D-Loop sequencing

The complete D-Loop region was amplified using primers L15971 and H00599 [21]. Amplicons were purified using columns (QiaquickPCR purification System-Qiagen,

Germany) or ExoSap (USB Corp, USA). The sequencing reactions were performed with the Big Dye Terminator System v1.1 (Life Technologies, Applied Biosystems, Foster City, USA) according to the supplier's protocol. Sequencing reaction products were purified from residual dye terminators by means of alcohol precipitation. All sequences were performed with both forward and reverse primers (with at least 4–6 primers for each sample) and electropherograms were visualized and edited with the Sequencher v4.8 (Gene Codes Corporation, USA) software.

Capillary Electrophoresis analysis

Capillary Electrophoresis analysis was carried out in an automated sequencer ABI3100-*Avant* (Applied Biosystems, Foster City, USA) either for fragment or sequence analysis.

SNP analysis

The Y-SNPs: M3 C/T (Q1a3a1 haplogroup -"hg"-), M269 T/C (hg R1b1a2) and U179 G/A (hg I), and the mtDNA-SNPs: 8027 G/A and 12007 G/A (that define hgA2), 3547 A/G (hgB2), 14308 T/C (hgC) and 2092 C/T (hgD2), were analyzed by Real Time PCR followed by High Resolution Melting Analysis as previously described [22] in Native American samples. The SNP M3 was also typed in AS group.

Statistical analysis

Autosomal STRs

Allele frequency, power of discrimination (PD) and power of exclusion (PE) were obtained with PowerStatsV12. Hardy–Weinberg equilibrium (HWE), gene diversity, Exact Test of population Differentiation and heterozygosities were calculated with Arlequin v3.1 [23].

Population heterogeneity was analyzed by means of an AMOVA test (Arlequin v3.1), the hierarchical groups selected for the analysis was: MA + TE versus AS group. Normalized Slatkin genetic distance matrix [24] was obtained with the help of Arlequin 3.1 and graphed as a Multi-Dimensional Scaling plot (MDS) by using XLSTAT (Addinsoft Corp) software.

Supplementary Table 1 shows autosomal genotyping data of the groups included in the present study.

Y-chromosome polymorphism

Haplotype data were analyzed with Arlequin v3.1. Haplotype diversity was calculated according to Nei [25]. With regard to locus DYS389, the length of DYS389II, expressed in number of repeat units, was subtracted to that of DYS389I [26]. The YHRD database [20] was used to search worldwide haplotype frequency distribution. Diversity parameters including dw_{min}, mw_{max}, mw_{min}, mb_{min} and db_{max} were calculated as previously described [27].

Supplementary Table 2 provides haplotypic data of the groups included in the present study.

Admixture analysis

The admixture analysis of AS group based on Y-chromosome STRs was carried out with Admix 2_0 software [28, 29]. Two parental populations were selected: a set of aboriginal samples composed of Mapuche and Tehuelche isolates—all of them belonging to Q1a3a1 haplogroup—was considered as parental population 1 and a set of individuals from Buenos Aires area (BA) as parental population 2 (selected on the basis that this population is those who have had more interaction with European immigrants). Mutation rate was taken from Zhivotovsky et al. [30]. Time of admixture event was 150 years, corresponding to the time when a massive interaction between aboriginals and military forces took place during the epilogue of the Second Campaign to the Argentinean Patagonian Desert in 1879.

Admixture analysis based on autosomal STRs data was carried out using Structure 2.0 software [31] and involved the same parental and test populations as described above. Admixture and correlated allele frequencies models were selected. Burn-in was set to 20,000 iterations, followed by 30,000 Markov Chain Monte Carlo steps.

Mitochondrial DNA

Sequences were aligned with DNA Alignment Software v1.1.3.0 (www.fluxus-engineering.com). Haplotype frequencies, molecular diversity indices and Fst (significant level 0.05) were calculated using Arlequin v3.1. MDS, between MA, TE, MG, PE, HU and YG groups (reference in Table 1) were plotted with XLSTAT (Addinsoft Corp) software. Median-joining networks were obtained with the Network 4.6.1.0 program [32].

Supplementary Table 3 provides mitochondrial haplotypes of Native American groups and AS group included in the analysis.

Surname transmission model

With the aim of evaluating the three possible surname transmission models, individuals included within AS were scrutinized by comparing the *kuga* or suffix of their family names with the Y-STRs haplotypes. Those cases in which a

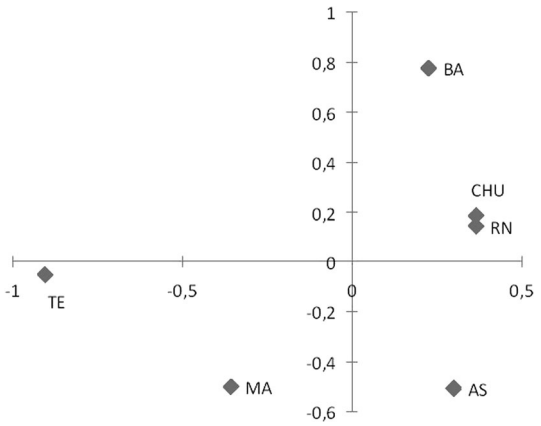

Fig. 1 MDS of autosomal STRs of linearized Slatkin distance showing the relationship between BA (Buenos Aires), CHU (Chubut), RN (Río Negro), MA (Mapuche), TE (Tehuelche) and AS (Aboriginal Surnames) for autosomal STRs data. Kruskal stress 1.104E−4

common suffix matched an identical haplotype might reflect the paternal transmission of this cultural trait.

Results and discussion

Autosomal STRs

A total of 15 STRs were analyzed in the Mapuche and Tehuelche samples. Supplementary Table 1 describes the allele frequency and the genetic parameters that characterized each locus in both Amerindian groups. Frequency distributions for the markers THO-1, CSF1PO, vWA, D13S317, D7S820 and D16S539 do not show significant differences from previously published data [17]. In both groups the highest discriminative power was observed in Penta E. All loci met HWE, with the exception of D18S51 and D8S1179 loci in the Tehuelche group. After applying the Bonferroni correction [33], none of the analyzed markers departed from HWE. Observed heterozygosity ranged from 0.59 (THO-1) to 0.95 (D18S51; D13S317 and PENTA E) in Mapuche, and 0.57 (THO-1 and D16S539) to 1 (Penta E) in Tehuelche. The average gene diversity was similar in both groups (0.759 ± 0.385 in Mapuche and 0.746 ± 0.381 in Tehuelche, respectively).

In relation to the AS group, the gene diversity was higher than that in the aboriginal groups (0.774 ± 0.402) and the observed heterozygosity ranged from 0.606 (D5S818) to 0.892 (D8S1179). The Exact Test of population differentiation was not statistically significant in any of the three groups.

Normalized Slatkin genetic distance was represented in a two-dimensional MDS plot (Fig. 1) including urban samples (BA, RN, CHU and AS groups) and native groups: Mapuche (MA) and Tehuelche (TE). All distances were significant, except those between Río Negro and Chubut. The AS group fell equidistant between MA and the cluster RN-CHU, while the largest distance was observed between TE and BA. The shortest distance was observed between CHU and RN, whereas the longest was observed between BA and both aboriginal groups.

Statistically non-significant differences were observed when the AMOVA analysis was conducted among groups (MA + TE vs. AS) and not among populations within groups ($p = 0.660 \pm 0.015$ and 0.156 ± 0.008, respectively).

Admixture analysis was performed using aboriginal samples (Mapuche and Tehuelche) and the Buenos Aires sample as parental populations. Selection of $k = 2$ was based upon historical information. These parental populations were compared with the urban populations of Chubut and Río Negro and the sample set whose surnames included Mapudungun words. Results are summarized in Table 2. A higher Native American contribution is detectable in the AS group, followed by Chubut and finally Río Negro urban populations. The parental aboriginal groups denoted 82–17 % Native American to European proportion while in the Buenos Aires sample the figures were almost the inverse, 81 % European to 19 % Native American, in complete agreement with Corach et al. [2] and Avena et al. [3].

Table 2 Admixture analysis based on autosomal STRs using the STRUCTURE software

Population	Cluster 1	Cluster 2
Tehuelche	0.828	0.172
Mapuche	0.821	0.179
Aboriginal Surname	0.572	0.428
Chubut	0.427	0.573
Río Negro	0.344	0.656
Buenos Aires	0.190	0.810

Cluster 1: composed of a gene pool of Native American isolated communities

Cluster 2: mostly composed of European contribution

Y-chromosome analyses

Supplementary Table 2 depicts the haplotypes observed in the Tehuelche and Mapuche samples. In total, 13 out 17 different haplotypes were observed in the Tehuelche group and 14 out 16 in the Mapuche. Two haplotypes were shared between the two groups (MY1-TY7 and MY2-TY12, respectively). These two haplotypes (considering minimal haplotype plus DYS437, DYS438 and DYS439 loci) were not observed in the worldwide haplotype database (release 39) [20].

Haplotype diversity was 0.958 ± 0.036 in MA and 0.948 ± 0.04 in TE. The analysis of the SNP M3-Q3 that defines the Native American haplogroup Q1a3a1 was present in 10/17 (58.82 %) in the Tehuelche haplotypes and in 14/16 (87.5 %) in the Mapuche samples.

In the AS group, 54 out of 67 haplotypes were different; 48 of which were unique. Haplotype diversity was 0.987 ± 0.066. Fifty nine percent (40/67) of the samples belonged to haplogroup Q1a3a1.

In 27/67 (40.29 %) of the samples we found no correlation between surname transmission and Native American ancestral haplogroups. These samples exhibited non-Native American Y-specific chromosome lineages in combination with aboriginal surnames. In contrast, 96 % of these samples (26/27) belonged to Native American matrilineages, strongly suggesting that maternal surname transmission played a key role in the maintenance of aboriginal family names. Only one sample with an Amerindian surname showed non-Amerindian matri or patrilineal ancestral haplogroups.

No shared surname was detected within the sixty-seven individuals with Mapuche surnames. The sample set denoted 31 different suffixes associated with the ancestral names or "*kuga*". Twenty of them were represented once, two twice, three three times, one four times, four five times, one six times and one nine times. With the exception of the last group, all other individuals showed different Y-STR haplotypes (data not shown). In the latter group, individuals shared the lineage name *lef* (meaning fast), although their prefixes were different, resulting in nine different family names. All of them exhibited an Amerindian mtDNA hg, and five the Y chromosomal Q1a3a1 hg; three of them shared an identical Y-STR haplotype, supporting the "*kuga*" transmission. The limited number of individuals sharing the *kuga* and Y-chromosome haplotype might suggest that the selected-paternal surname transmission model proposed by Jimenez [34] is nowadays infrequent as this condition was exhibited by only three out of 67 individuals (4.47 %).

Although special attention has been paid to the *Mapudungun* as a language, it should be noted that it might represent a "*Koiné*" or common language since many different ethnic groups as the Tehuelche, Pehuenche, Ranqueles and Pampa descendants may have adopted the language after Mapuche invasions and language imposition [35]. Accordingly, the prevalence of Mapuche linguistic elements within their current family names not necessarily indicate that they belong to the Mapuche group, but to other ethnicities that were influenced by the Mapuche during a period of over 300 years [16].

Haplotype diversity in the AS group was slightly higher than in the aboriginal groups, considering the minimal haplotype (AS: 0.988 ± 0.006; MA: 0.958 ± 0.036; TE: 0.948 ± 0.04, respectively). Non-significant genetic distances (Rst) were observed between the AS and TE or MA groups, when Y-chromosome data were compared. Figure 2 depicts the relationship between Y-chromosome haplotypes in BA, Río Negro (RN), Chubut (CHU), MA, TE and AS. The figure obtained shows a similar picture than that obtained for autosomal genotypes (Fig. 1).

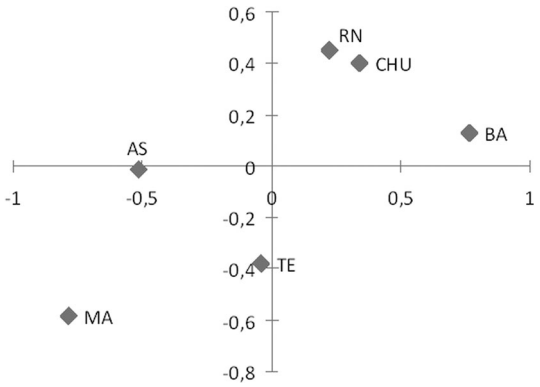

Fig. 2 MDS of minimal Y-STRs haplotype genetic distances. Kruskal stress 1.103E−4

Table 3 presents the matching probabilities within and among groups, taking into account the minimal Y STR-haplotype. The Tehuelche group (0.1073) displays the highest maximum probability of finding a match within the population (mw_{max}) followed by the Mapuche and AS groups. The maximum probability of obtaining dissimilar haplotypes (db_{max}) when sampling two individuals from MA and AS, MA and TE and AS and TE are 0.9813, 0.9852 and 0.9947, respectively. These results are consistent with the haplotype sharing observed between the analyzed populations (maximum shared haplotypes were observed between Mapuche and AS as was mentioned above). The parameter mw_{max}/mb_{min} gives an estimate for the upper limit of how many times more likely it is to find a match within a population rather than between two populations. Its high value in the Tehuelche group in relation to the AS group is remarkable. The lowest value was obtained in the AS group with respect to the Mapuche group, in which the results indicate that it is approximately equally probable to find a match within or between groups. Similar conclusions were drawn from mtDNA results (data not shown). According to these analyses, there is a similar probability for finding shared mitochondrial or Y-haplotypes in the AS group than between the AS and MA groups. This finding could be interpreted as a closer genetic relationship between AS and MA than AS and TE.

The admixture analysis from Y-STR data demonstrated that the contribution of parental population 1 (aboriginal) to the Aboriginal Surname sample was 58.7 % while the contribution of parental population 2 (BA sample) was

Table 3 Y-minimal haplotype matching probabilities within and among Mapuche, Tehuelche and Aboriginal Surname groups

	MA	TE	AS
N	16	17	67
dw_{min}	0.898	0.892	0.973
mw_{max} $(1 - dw_{min})$	0.101	0.107	0.027
mb_{min}	0.018 MA/AS	0.014 TE/MA	0.018 AS/MA
	0.014 MA/TE	0.005 TE/AS	0.005 AS/TE
$db_{max} = 1 - mb_{min}$	0.981 MA/AS	0.985 TE/MA	0.994 AS/TE
	0.985 MA/TE	0.994 TE/AS	0.981 AS/MA
mw_{max}/mb_{min}	5.433 MA/AS	7.25 TE/MA	1.443 AS/MA
	6.860 MA/TE	20.24 TE/AS	5.094 AS/TE

MA Mapuche, *TE* Tehuelche, *AS* Aboriginal Surname, *N* number of individuals in the data set

mw_{max} maximum probability of finding a match within the population, db_{max} maximum probability of obtaining dissimilar haplotypes, mw_{max}/mb_{min} estimates the probability to find a match within/between two populations (see "Materials and methods" section for more details)

Table 4 Admixture analysis based on Y-STRs data using Admix-2

Population	mY1	mY2
Aboriginal Surname	0.587	0.413
Chubut	0.272	0.728
Río Negro	0.047	0.953

mY1: composed of a gene pool of Native American isolated communities

mY2: composed of the Buenos Aires sample

Performed with Admix 2_0 software; Bootstrapped 1,000 times

41.3 % (Table 4). A more predominant European male contribution was observed in the urban people from Chubut (72.8 %) and Río Negro (95.3 %).

Mitochondrial DNA data

All aboriginal samples denoted Amerindian specific haplogroups. A total of 39 unrelated Mapuche individuals were analyzed by means of the entire mitochondrial D-Loop sequence. The haplogroup distribution was hgA2 = 12.82 %; hgB2 = 35.90 %, hgC1 = 25.64 % (C1b = 70.0 % and C1d = 30.0 %, respectively) and hgD = 25.64 % (D1g = 50.0 %; D1j = 20.0 %, D4h3 = 20.0 % and D4 = 10.0 %, respectively). Haplotype diversity was 0.983 and nucleotide diversity was: 0.014 ± 0.007. Thirty-one out of 39 different haplotypes were observed and 26 haplotypes were unique. One sample exhibited the Huetar deletion motif, with the characteristic 106–111 deletion in association with haplogroup hgD1 [36–40].

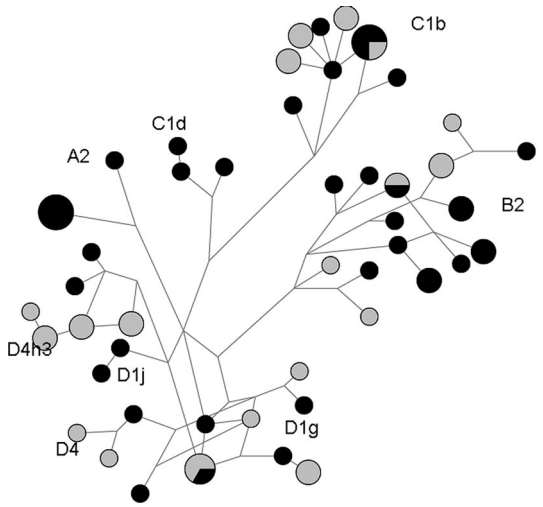

Fig. 3 Median-joining network connecting Tehuelche (in *grey*) and Mapuche (in *black*) haplotypes. Node size is proportional to the frequency (frequency range from 1 to 4 individuals)

Tehuelche haplogroup distribution was as follow: 21.43 % hgB2, 25.0 % hgC1 (subhaplogroup C1b) and 53.57 % hgD (D1g = 40.0 %; D4h3 = 46.670 % and D4 = 13.33 %, respectively), with a total of 19/28 different haplotypes. Ten were unique. No samples belonging to hgA2 were detected. Haplotype diversity was 0.976 and nucleotide diversity was 0.0118 ± 0.0061.

Supplementary Table 3 depicts the haplotypes observed in both groups. Sub-haplogroups were defined in accordance with the coding region SNPs and non-coding mutations present in the mitochondrial control region, according to Perego et al. [41] and Bodner et al. [42].

Only 3 out of 50 different haplotypes were shared by both tribal groups (hgs B2, C1 and subhaplogroup D1g). The median-joining (M-J) network is displayed in Fig. 3, in which the relationship between the Mapuche and Tehuelche haplotypes is shown. Fst genetic distance was not significant ($p > 0.05$).

Mitochondrial DNA results were compared with those of other groups previously described, including one Mapuche group belonging to Anecon Grande (MG) in Río Negro province [43] and the groups analyzed by Moraga et al. [44] corresponding to three Native American groups from Chile: Pehuenche (PE), Mapuche (HU) and Yaghan (YG) (see Table 1). Sequence analysis was restricted to hypervariable regions I and II (fragment considered for HVRI: 16050–16400 and for HVRII: 70–300). Table 5 depicts the haplogroup distribution in the different groups. As it can be observed, the haplogroup distribution in both Mapuche samples from Argentina (MA and MG) is almost identical, while in the Tehuelche samples the hgD frequency is in concordance with the Chilean groups. Figure 4 depicts the obtained M-J network restricted to HVRI and

Table 5 mt DNA haplogroups distribution

	hgA (%)	hgB (%)	hgC (%)	hgD (%)	Non Amerindian hg (%)	References
Mapuche (MA)	12.82	35.89	25.64	25.64	0	Present study
Tehuelche (TE)	0	21.43	28.57	50.0	0	Present study
Aboriginal Surname (AS)	4.48	28.36	35.82	26.86	4.48	Present study
Mapuche (MG)	15.38	38.46	20.51	25.64	0	[43]
Mapuche (HU)	0	7.1	44.1	48.7	0	[44]
Pehuenche (PE)	2.8	10.5	41.0	45.7	0	[44]
Yaghan (YG)	0	0	47.6	52.4	0	[44]

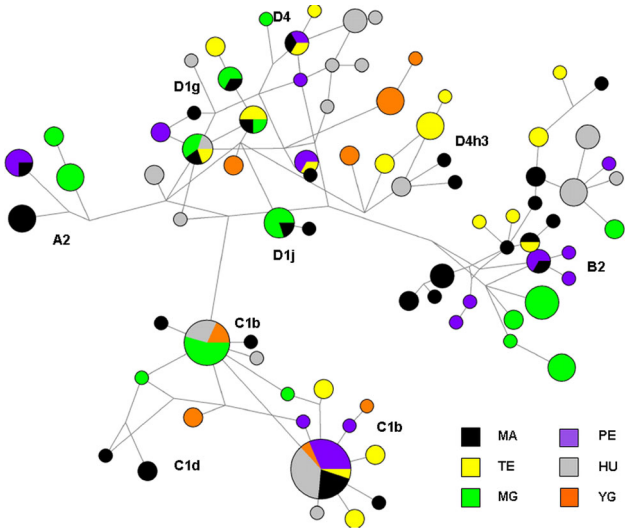

Fig. 4 Median-joining network for mtDNA haplotypes of South American Natives. Areas of the *circles* are proportional to the number of individuals. References are indicated in Table 1. Analysis was restricted to HVRI and HVRII

HVRII. Shared haplotypes were observed in hgC and hgD between the Argentinean and Chilean groups, and to a lesser extent, between MA and PE in hgA and hgB.

In Fig. 5, a MDS plot of the Fst distances shows the relationship between these groups. The Mapuche (MA) group shows no significant differences regarding TE, MG and PE, and shows significant distances with respect to HU and YG. No significant differences were observed between TE and HU (Mapuche from Chile). This result could coincide with the process of Araucanization experienced by the Tehuelches when trans-Andean Araucans arrived in Argentina (a process of admixture between Araucanian men and Tehuelche women).

In 67 samples of individuals that were classified according to their surnames (AS group) 55 different haplotypes were observed; 45 of which were unique. Molecular indices obtained were: haplotype diversity 0.9918; mean number of pairwise differences 13.22 ± 6.02 and nucleotide diversity (average over loci) 0.012 ± 0.006.

Haplogroup distribution was hgA2 4.48 %, hgB2 28.36 %; hgC1 35.82 % (all belonging to the C1b subhaplogroup); hgD1 26.86 % (all D1g) and 4.48 % denoted non-specific Native Amerindian haplogroups (3/67, two of hgU5a and one hgH). Significant differences (Fst) were observed between AS when compared to TE and MA.

Discussion

A vast territory located south of the Colorado River in Argentina and the Bio-Bio River in Chile was the land that belonged to diverse Amerindian groups up to the end of the nineteenth century. These communities were displaced by military campaigns on both sides of the Andes.

At present, some isolated communities persist and maintain their cultural traditions including their language, the *Mapudungun*. This work provides additional information about the genetic characteristics of two Patagonian ethnic groups. Available information is scarce and fragmentary, especially concerning genetic attributes, although the Mapuche were the first South American Amerindian peoples to be investigated by means of mtDNA polymorphisms [43]. At present, new contributions are shedding light on these Patagonian groups [41, 42].

The members of isolated communities showed Native American mitochondrial lineages. A reduced number of Mapuche individuals showed hgA2, which was absent in the Tehuelche.

The analysis of the haplotypes showed interesting attributes such as the insertion 291.1A observed in 80 % (4/5) of hgA2 haplotypes. This frequent haplotype in the Mapuche (64T, 73G, 146C, 153G, 235G, 263G, 291.1A, 315.1C, 523d, 524d, 16111T, 16223T, 16290T, 16319A, 16356C, 16362C) was observed and previously reported in one sample from Río Negro province [45]. Additional information provided by autosomal STRs analysis allowed detecting a possible first-degree relationship between two of the hgA samples. These samples belong to two different communities (LoncoVaca and Aguada Guzman, respectively). Internal migrations between the communities could

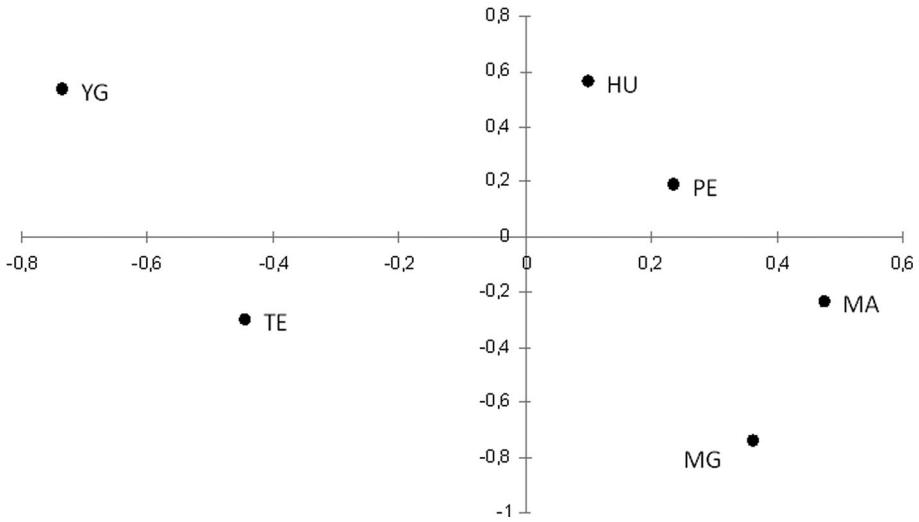

Fig. 5 MDS obtained from F*st* distances for mtDNA haplotypes. References are indicated in Table 1

explain this observation. This situation reduces the observed frequency of hgA2 from 12.82 to 10.25 %.

Haplogroup B2 also showed clear peculiarities such as the combination of two point mutations, 470G and 499A, in the hypervariable region III (HVRIII) that were observed in 93 % of the Mapuche and 67 % of the Tehuelche; additionally, in the Mapuche the variant 16249C was observed in 43 % of the hgB2 but was absent in the Tehuelche. In hgC1, all Mapuche and Tehuelche individuals exhibited the combination 493G, 523d, 524d within HVRIII. Mutation 16241G was associated to hgD in 20 % of Mapuche and 47 % in Tehuelche. These individuals also exhibited the mutated positions 16301T and 16342C, which could be assigned to haplogroup D4h3, according to previous publications [41, 46]. One Mapuche sample (belonging to hgD1g, see Supplementary Table 3) exhibited a 6 bp deletion (106–111d). This motif described in two Aymara individuals from northern Chile [36] associated with hgD, was also detected in Chibcha-speaking groups from Costa Rica and Panama associated with hgA (identified by Santos et al. [39], who coined the term "Huetar deletion motif"). Kolman et al. [47] described the same motif in Ngöbe Amerindians from Panamá, also associated to hgA samples. More recently, Gaya-Vidal et al. [48] found this deletion in one Quechua sample from Bolivia, in association with hgA and Bobillo et al. [45] found it in three hgD urban samples (two from Buenos Aires and one from Corrientes province) in a population study conducted in the Argentinean population. These findings could indicate that the 6-bp deletion is not specific to any Native American haplogroup but could have multiple origins, as proposed by Merriwether et al. [37].

A high proportion of Y-Native American chromosome lineages (haplogroup Q1a3a1) were found in the Mapuche sample (87.5 %). This result contrasts with those reported by Blanco-Verea et al. [49], where the authors found that in a sample of 76 individuals considered as Mapuche, only 29 % belonged to haplogroup Q1a3a1 while 71 % of them denoted non-Amerindian Y-chromosomes. According to the geographical information provided, the sample set corresponds to an urban population (Esquel city, northwest of Chubut province) and the results could reflect a higher degree of admixture with the non-Amerindian population. In our study, only two Mapuche individuals belonged to the non-Native American Y-Haplogroup: one to R1b1b and the other to Clade E (the first one was defined based on SNP typing and the last one on the Haplogroup Predictor [50]). In the Tehuelche group, haplogroup Q1a3a1 was present in 58.82 % of the samples. The samples with non-Native American haplogroups belong to R1b, I and G2a haplogroups (the first two were defined based on SNPs typing and the last one was inferred with the help of the Haplogroup Predictor). Only two haplotypes were shared between the two communities.

The analysis of the AS group allowed to detecting higher gene diversity values from autosomal, Y-chromosome and mtDNA than those obtained in the MA or TE groups. In the AS group, 95.52 and 59.70 % of the samples belong to Native American mitochondrial or Y-chromosome specific lineages, respectively. Excluding the polymorphisms observed between positions 300–315 of mtDNA-HVRII, seven mtDNA haplotypes were shared between MA and AS and four between TE and AS. Five Y-haplotypes were shared between AS and MA and two haplotypes were common in the three groups. In addition, the most frequent minimum haplotype in AS (13-13-31-23-10-14-13-12,15) shows a frequency of 9 % and is related to one of the most frequent haplotype in MA (13-13-31-23-9-14-13-12,15) whose frequency is 12.5 %. Both haplotypes differ by a single mutation step in locus DYS391. Previous research in the general population of Argentina showed that specific Native American haplogroup Q1a3a1 is

present in only 5 % of the samples [2], showing a clear difference with the values observed in the AS group.

The admixture analysis carried out with autosomal and Y-STRs showed a higher degree of admixture in those individuals whose surnames contained *Mapudungun* words, where the Native American ancestry is approximately 57 %, while in the urban population from Chubut, Río Negro or Buenos Aires provinces this component tends to decrease, as previously reported [2, 3]. In the MDS representation of autosomal and Y-chromosome results, the AS group is located in between Chubut and Río Negro populations and the aboriginal population, in concordance with admixture analysis results.

In conclusion, this work allowed us to characterize two Argentinean Patagonian ethnicities by means of genetic markers used in forensic identifications. The results obtained were compared with a random group selected on the basis of their surnames as cultural proxy indicator, provided they included *Mapudungun* words. Admixture analysis carried out by means of Y-STRs as well as autosomal STRs allowed us to establish a correlation between genetic information with linguistic ones. Nevertheless, further analysis including ancestral informative markers would be necessary to precisely determine the degree of admixture attained. The data obtained constitute a novel population genetics information that could provide in the future to draw the final genetic figure of the Native American groups of Argentina.

Acknowledgments We would like to thank Dr. RF Carnese for kindly providing the aboriginal samples during the early '90s and Dr. L. Saccomanno and S. Vannelli for helping in the sampling process of the Patagonian samples. This work was partly supported, by grants UBACyTB047, 2008–2010; UBACyT 2011-2014, Code 200202001 00744, to DC and grant PIP 11420090100232-CONICET to AS.

References

1. Corach D, Marino M, Sala A (2006) Relevant genetic contribution of Amerindian to the extant population of Argentina. Prog Forensic Genet 11(1288):397–399
2. Corach D, Lao O, Bobillo MC, van Der Gaag K, Zuniga S, Vermeulen M, van Duijn K, Goedbloed M, Vallone PM, Parson W, deKnijff P, Kayser M (2010) Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA. Ann Hum Genet 74:65–76
3. Avena S, Via M, Ziv E, Perez-Stable EJ, Gignoux CR, Dejean C, Huntsman S, Torres-Mejia G, Dutil J, Matta JL, Beckman K, Burchard EG, Parolin ML, Goicoechea A, Acreche N, Boquet M, Rios Part Mdel C, Fernandez V, Rey J, Stern MC, Carnese RF, Fejerman L (2012) Heterogeneity in genetic admixture across different regions of Argentina. PLoS ONE 7(4):e34695
4. Martinez Sarasola C (2005) Nuestros paisanos los indios. EmecéEdts, Buenos Aires
5. Dana LP (2006) Indigenous peoples in Chile. Int J Entrepreneurship Small Bus 3:779–786
6. García F, Moraga M, Vera S, Henríquez H, Llop E, Aspillaga E, Rothhammer F (2006) mtDNA microevolution in Southern Chile's archipelagos. Am J Phys Anthropol 129:473–481
7. Loukotka C (1968) Classification of South American Indian languages. Latin American Studies Center, University of California Press, Berkeley
8. Greenberg JH, Turner CG, Zegura SL, Campbell L, Fox JA, Laughlin WS, Szathmary EJE, Weiss KM, Woolford E (1986) The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. Curr Anthropol 27:477–497
9. Ruhlen MJ, Greenberg J (2007) An Amerindian etymological dictionary. Stanford University Press, Stanford
10. Latchan RE (1922) La organización social y la creencia de los antiguos araucanos. Rev Museo Hist Nat 3:245–868
11. Faron LC (1956) Araucanian Patri-Organization and the Omaha system. Am Anthropol 58:435–456
12. Silva Galdanes O (1984) Los Araucanos Prehispanos ¿Un Caso de Doble Filiación? Boletín del Museo Regional de la Araucanía Temuco 1:41–46
13. Lehmann-Nitsche R (1921) El grupo lingüístico Alakaluf de los canales magallánicos. Revista del Museo de La Plata, Buenos Aires
14. Escalada F (1949) El complejo tehuelche. Estudios de etnografía patagónica. Instituto Superior de Estudios Patagónicos Press, Buenos Aires
15. Casamiquela R (1965) Rectificaciones y ratificaciones hacia una interpretación definitiva del panorama etnológico de la Patagonia y área septentrional adyacente. Instituto de Humanidades: Universidad Nacional del Sur Press, Bahía Blanca
16. Casamiquela R (1969) Un nuevo panorama etnológico del área pampeana y patagónica adyacente. Pruebas etnohistóricas de la filiación tehuelche septentrional de los guaraníes. Museo Nacional de Historia Natural Press, Santiago de Chile
17. Sala A, Penacino G, Carnese R, Corach D (1999) Reference database of hypervariable genetic markers of Argentina: applications for molecular anthropology and forensic casework. Electrophoresis 208:1733–1739
18. Quinque D, Kittler R, Kayser M, Stoneking M, Nasidze I (2006) Evaluation of saliva as a source of human DNA for population and association studies. Anal Biochem 15:272–277
19. Marino M, Sala A, Corach D (2007) Genetic attributes of the YHRD minimal haplotype in 10 provinces of Argentina. Forensic Sci Int Genet 1:129–133
20. Y Chromosome Haplotype Reference Database. http://www.yhrd.org
21. Brandstätter A, Niederstatter H, Parson W (2004) Monitoring the inheritance of heteroplasmy by computer-assisted detection of mixed basecalls in the entire human mitochondrial DNA control region. Int. J. Legal Med. 118:47–54
22. Zuccarelli G, Alechine E, Caputo M, Bobillo MC, Corach D, Sala A (2011) Rapid screening for Native American mitochondrial and Y-chromosome haplogroups detection in routine DNA analysis. Forensic Sci Int Genet 5:105–108
23. 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
24. Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 41:264–279
25. Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York
26. Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L et al (1997) Evaluation of

Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133
27. Brinkmann C, Forster P, Schürenkamp M, Horst J, Rolf B, Brinkmann B (1999) Human Y chromosomal STR haplotypes in a Kurdish population sample. Int J Leg Med 112:181–183
28. Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. Mol Biol Evol 15:1298–1311
29. Dupanloup I, Bertorelle G (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol Biol Evol 18:672–675
30. Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L (2004) The effective mutation rate at y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74:50–61
31. Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959
32. Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations. Genetics 141:743–753
33. Bonferroni CE (1936) Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8:3–62
34. Jimenez JF (2002) Matrilinealidad versus patrilinealidad. La obra de Félix José de Augusta y la polémica acerca de la filiación entre los Reche Mapuche. Centro de Documentación Patagónica Eds, Bahía Blanca
35. Censabella M (2007) Las lenguas indígenas de la Argentina. Universidad de Buenos Aires Eudeba Eds, Buenos Aires
36. Merriwether DA (1993) Mitochondrial DNA variation in South American Indians. PhD Dissertation, University of Pittsburgh, Pittsburgh
37. Merriwether DA, Ferrell RE, Rothhammer F (1995) mtDNA D-loop 6-bp deletion found in the Chilean Aymara: not a unique marker for Chibcha-speaking Amerindians. Am J Hum Genet 56:812–813
38. 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
39. Santos M, Barrantes R (1994) D-Loop mtDNA deletion is a unique marker of Chibchan Americans. Am J Hum Genet 55:413–414
40. Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong QP, Myres NM, Salas A, Semino O, Bandelt HJ, Woodward SR, Torroni A (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr Biol 19:1–8
41. Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gomez-Carballa A, Huber G, Zimmermann B, Corach D, Babudri N, Panara F, Myres NM, Parson W, Semino O, Salas A, Woodward SR, Achilli A, Torroni A (2010) The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. Genome Res 20:1174–1179
42. Bodner M, Perego UA, Huber G, Fendt L, Röck A, Zimmermann B, Olivieri A, Gómez-Carballa A, Lancioni H, Angerhofer N, Bobillo MC, Corach D, Woodward SR, Salas A, Achilli A, Torroni A, Bandelt HJ, Parson W (2012) Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. Genome Res. doi:10.1101/gr.131722.111
43. Ginther C, Corach D, Penacino GA, Rey JA, Carnese RF, Hutz MH, Anderson A, Just J, Salzano FM, King MC (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. Exs 67:211–219
44. 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
45. 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
46. Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ (2008) The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. PLoS ONE 3(3):e1764
47. Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngobe Amerinds of Panama. Genetics 140:275–283
48. Gayà-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, Villena M, Vasquez R, Bravi CM, Dugoujon J-M (2011) mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: different stories and special genetic traits of the Andean Altiplano populations. Am J Phys Anthropol 145:215–230
49. Blanco-Verea A, Jaime JC, Brión M, Carracedo A (2010) Y-chromosome lineages in native South American population. Forensic Sci Int Genet 4:187–193
50. Haplogroup Predictor. http://www.hprg.com/hapest