

The Genetic Legacy of the Pre-Colonial Period in Contemporary Bolivians

Patricia Taboada-Echalar^{1,9}, Vanesa Álvarez-Iglesias^{1,9}, Tanja Heinz^{1,9}, Laura Vidal-Bralo¹, Alberto Gómez-Carballea¹, Laura Catelli³, Jacobo Pardo-Seco¹, Ana Pastoriza¹, Ángel Carracedo¹, Antonio Torres-Balanza², Omar Rocabado², Carlos Vullo^{3,4}, Antonio Salas^{1,*9}

1 Unidad de Xenética, Instituto de Ciencias Forenses and Departamento de Anatomía Patológica e Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, Galicia, Spain, **2** Instituto de Investigaciones Forenses, Fiscalía General del Estado Plurinacional de Bolivia, La Paz, Bolivia, **3** Equipo Argentino de Antropología Forense, Córdoba, Argentina, **4** Laboratorio de Inmunogenética y Diagnóstico Molecular, Córdoba, Argentina

Abstract

Only a few genetic studies have been carried out to date in Bolivia. However, some of the most important (pre)historical enclaves of South America were located in these territories. Thus, the (sub)-Andean region of Bolivia was part of the Inca Empire, the largest state in Pre-Columbian America. We have genotyped the first hypervariable region (HVS-I) of 720 samples representing the main regions in Bolivia, and these data have been analyzed in the context of other pan-American samples (>19,000 HVS-I mtDNAs). Entire mtDNA genome sequencing was also undertaken on selected Native American lineages. Additionally, a panel of 46 Ancestry Informative Markers (AIMs) was genotyped in a sub-set of samples. The vast majority of the Bolivian mtDNAs (98.4%) were found to belong to the main Native American haplogroups (A: 14.3%, B: 52.6%, C: 21.9%, D: 9.6%), with little indication of sub-Saharan and/or European lineages; however, marked patterns of haplogroup frequencies between main regions exist (e.g. haplogroup B: Andean [71%], Sub-Andean [61%], Llanos [32%]). Analysis of entire genomes unraveled the phylogenetic characteristics of three Native haplogroups: the pan-American haplogroup B2b (originated ~21.4 thousand years ago [kya]), A2ah (~5.2 kya), and B2o (~2.6 kya). The data suggest that B2b could have arisen in North California (an origin even in the north most region of the American continent cannot be disregarded), moved southward following the Pacific coastline and crossed Meso-America. Then, it most likely spread into South America following two routes: the Pacific path towards Peru and Bolivia (arriving here at about ~15.2 kya), and the Amazonian route of Venezuela and Brazil southwards. In contrast to the mtDNA, Ancestry Informative Markers (AIMs) reveal a higher (although geographically variable) European introgression in Bolivians (25%). Bolivia shows a decreasing autosomal molecular diversity pattern along the longitudinal axis, from the Altiplano to the lowlands. Both autosomes and mtDNA revealed a low impact (1–2%) of a sub-Saharan component in Bolivians.

Citation: Taboada-Echalar P, Álvarez-Iglesias V, Heinz T, Vidal-Bralo L, Gómez-Carballea A, et al. (2013) The Genetic Legacy of the Pre-Colonial Period in Contemporary Bolivians. PLoS ONE 8(3): e58980. doi:10.1371/journal.pone.0058980

Editor: Carles Laluetza-Fox, Institut de Biologia Evolutiva - Universitat Pompeu Fabra, Spain

Received: December 23, 2012; **Accepted:** February 12, 2013; **Published:** March 20, 2013

Copyright: © 2013 Taboada-Echalar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The research leading to these results has received funding from the People Program (Marie Curie Actions) of the European Union's Seventh Framework Program FP7/2007-2013/ under REA grant agreement n° 290344, and the grants from the "Ministerio de Ciencia e Innovación" (SAF2008-02971) and from the Plan Galego IDT, Xunta de Galicia (EM 2012/045) given to A.S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: antonio.salas@usc.es

9 These authors contributed equally to this work.

Introduction

The Republic of Bolivia is located in central-south America. It is bordered by Peru to the West, Chile to the southwest, Paraguay and Argentina to the South, and Brazil to the North and East. Before the arrival of the Europeans, the Andean region of the country was an important part of the Inca Empire, the largest state in Pre-Columbian America, although the Inca civilization arose from the highlands of Peru in the early 13th century. The Spaniards discovered the silver mines of Potosí in 1544 and soon began enslaving Natives as workers in the mines. The Spanish Empire conquered the region in the XVI century, and, during the colonial period, this territory was called Upper Peru. In the XVII century, the Spanish began bringing in African slaves in high numbers to help work in the mines, an institution that would last

until abolition in 1826 (the independence of the country would arrive in 1825).

Nowadays, Bolivia is politically divided into nine departments and its geography varies from the high mountains in the Andes (West) to the eastern lowlands (Llanos), situated within the Amazon Basin.

About 10.2 million people live in Bolivia (Instituto Nacional de Estadística de Bolivia; INE; <http://www.ine.gov.bo/>). The country harbors a great cultural diversity. The number of individual languages known in Bolivia is 45; of those, 37 are living languages, one is a second language without mother-tongue speakers, and seven have no known speakers [1,2]. The main language spoken today in Bolivia is Spanish, but there are also other important pre-Hispanic languages, such as Quechua (inherited from the Incan Empire; spoken by >1.6 million inhabitants), and Aymara (1.3

million inhabitants). The Quechua-speaking peoples inhabit mostly the (Sub)Andean valleys of Cochabamba and Chuquisaca and some mountain regions in Potosí and Oruro, while Aymara is mainly spoken in the high plateau (Altiplano) of the departments of La Paz, Oruro and Potosí (the area around Lake Titicaca). There are also other ethnic groups in the East, mainly living in the Llanos (including the Bolivian Amazon areas) e. g. Chiquitanos (>110,000 inhabitants), Guaraníes (>78,300 inhabitants; living mainly on the border with Paraguay), Moxeños (>76,000 inhabitants); these groups mainly occupy the departments of Santa Cruz, Beni, Pando and Tarija. The statistics vary slightly according to the different sources (see e.g. also [3]).

The ethnic composition of Bolivia includes a great diversity of cultures but most indigenous peoples have assimilated a ‘mestizo’ culture. The Amerindian population accounts for approximately 55%; the remaining population is believed to be admixed with Europeans and Africans. There are more than 30 ethnic groups in Bolivia, the largest being Quechua- (about 1,500,000) and Aymara-speaking (25%). Several pseudo-ethnic terms are commonly used in Bolivia to self-designate their ancestries, such as ‘Mestizos’ (considered to be a mixture of Native Bolivians and Europeans), ‘Blancos’ (‘Whites’; considered to be descendants of Europeans or ‘Criollos’), ‘Afro-Bolivians’, Asians, and others (mainly Europeans from Germany, France, Italy, Portugal, and a minority of people coming from neighboring countries such as Argentina, Brazil, Chile, Colombia, etc.). The ‘Afro-Bolivians’ are descendants of African enslaved people, live mainly in the department of La Paz, and are concentrated in the provinces of Nor Yungas and Sud Yungas. Asian individuals are mainly Japanese (about 14,000) and Chinese (4,600).

Analysis of mtDNA variation has been used to explore demographic patterns in different regions of America, helping to unravel the origin of different ethnic groups, and the impact of modern migrations [4–9]. Some South American regions have received more attention in the literature than others; e.g. Argentina [10–12], Brazil [13,14], Colombia [15–17], etc. However, limited genotyping efforts have been dedicated to Bolivians, and those have mainly focused on analysis of a few Native American Bolivians. Bert et al. [18] analyzed the mitochondrial DNA diversity in three different populations from the Llanos de Moxos ($n = 54$) located in the lowlands of the Amazonian basin; the data indicated a higher genetic diversity in this locality than that observed in other American populations. In 2004, Dornelles et al. [19] analyzed a sample of Ayorea ($n = 91$) individuals living in two Bolivian and one Paraguayan (neighboring) communities; the data suggested the effect of strong genetic drift in this population, significantly reducing the amount of variability. More recently, Corella et al. [20] analyzed several small samples from the Bolivian Piedmont, including Chimane, Mosen, Aymara and Quechua ($n = 100$); the results suggest high genetic diversity in the area and high levels of inter-group variability. Afonso-Costa et al. [21] analyzed 111 individuals collected from La Paz, and the data was analyzed in a forensic context. Finally, Gayà-Vidal et al. [22] analyzed a sample of Aymaras and Quechuas from Bolivia ($n = 189$); according to the authors, the data support a past common origin of the Altiplano populations in the ancient Aymara territory with independent although related histories with the Peruvian Quechuas.

Autosomal SNPs have been analyzed in only few studies. Galanter et al. [23] analyzed a panel of ancestry informative markers (AIMs) in a collection of pan-American samples, including a few from Bolivia. These authors found that most of the Bolivians have a main Native American component (ranging from 90% to 98%), with the exception of the Yungas province, showing a

dominant African nature. Recently, Watkins et al. [24] carried out a genome-wide analysis of 28 Bolivians, and although the results agreed with Galanter et al. in that Bolivian genomes have predominant Native American features, they estimated a higher European component (12%) in these peoples.

The goal of the present study is to explore the variability of the Bolivian populations globally, including non-Native Bolivians from rural and urban populations, as well as those representing main departments and ecological regions (from the high mountains to the Llanos), and comparing them to previous studies that were carried out on a more local scale and to Native populations. Here, we have carried out the largest sampling of Bolivian populations to date ($n = 720$). The data have been meta-analyzed jointly with previously available data from Bolivia and a large pan-American database. Whole genome sequencing was also carried out on selected mtDNAs in order to investigate Native American branches that have not been investigated before. In addition, AIMs have been genotyped in order to infer patterns of main continental ancestries that have contributed to the recent history of the country.

Materials and Methods

Sample collection

A total of 720 samples were recruited for the present study. They represent the three main regions of the country: a) Andean or Altiplano ($n = 240$); b) Sub-Andean ($n = 204$); and c) the Llanos ($n = 276$) which to a large extent corresponds to the Bolivian Amazonian Basin. These samples were collected in the following Bolivian departments: Beni ($n = 102$), Chuquisaca ($n = 89$), Cochabamba ($n = 103$), La Paz ($n = 253$), Pando ($n = 90$), and Santa Cruz ($n = 83$). Birthplaces and other geographic information are given in **Table S1**. Additional samples ($n = 490$) were collected from the literature: La Paz ($n = 110$; [21]), Chimane ($n = 10$), Mosen ($n = 10$), Aymara ($n = 10$) and Quechua ($n = 16$) [20], Ayoreo ($n = 91$; [19]), and Trinitario ($n = 12$), Yuracare ($n = 15$), Ignaciano ($n = 15$), and Movima ($n = 12$), all of which are around the small area of Llanos de Moxos [18], Aymara ($n = 96$) and Quechua ($n = 93$) [22].

A subset of the total samples ($n = 178$), representing the departments of La Paz ($n = 105$), and Chuquisaca ($n = 73$), were genotyped for a panel of 46 ancestry informative markers (AIMs) in Heinz et al. [25]. Additional Bolivian samples selected from the total were genotyped *de novo* ($n = 420$) and merged with those that were previously genotyped.

Ethics statement

Written informed consent was obtained from all sample donors. Analysis of mtDNA sequences was approved by the institutional review boards of Santiago de Compostela (Spain). Moreover, the study conforms to the Spanish Law for Biomedical Research (Law 14/2007- 3 of July).

DNA sequencing of the control region and entire genomes

Samples were PCR amplified and sequenced for HVS-I region, as described previously [26]. Entire genome sequencing was done as described in [27,28].

We have followed the phylogenetic approach to scan the sequences as an *a posteriori* sequence quality control using the principles described in [29–31]. This filter was also applied to the data collected from the literature.

Nomenclature of mtDNA variants are referred against the revised Cambridge Reference sequence or rCRS [32,33], and

haplogroup nomenclature follows Phylotree Build 15 (www.phylotree.org; [34]); see also [35]. For the sake of inter-population comparisons and population summaries, we make the simplification that all A, B, C, and D haplotypes correspond to the Native American branches and not to the East Asian ones [36], given that specific mtDNA SNPs were not available. Therefore, along the text and figures, we used haplogroup labels according to the level of phylogenetic resolution used in the present study; e.g. B4 instead of B2; although it is most likely that all B4 belong to the Native American haplogroup A2. **Table S1** provides however the most accurate haplogroup classification according to the level of phylogenetic resolution obtained in the present study. Entire genomes generated in the present study are publicly accessible *via* GenBank with accession numbers KC503925 to KC503933.

AIM-INDEL genotyping

Bolivian samples were genotyped for 46 INDEL markers [37] in a multiplex PCR amplification and capillary electrophoresis process. Each AIM-Indelplex PCR amplification was performed with 5 μ l 2x Quiagen Multiplex PCR Master Mix, 10x Primer Mix, and 0.5 μ l DNA (concentration between 0.5 – 5 ng/ μ l) in a final volume of 10 μ l. PCR thermocycling conditions were: initial temperature of 95°C for 15 min; 28 cycles at 94°C for 30 sec, 60°C for 90 sec, and 72°C for 60 sec; final step at 72°C for 60 min. Following amplification, 0.8 μ l PCR product was added to 11.5 μ l Hi-Di Formamide (Applied Biosystems) and 0.3 μ l Liz-500 Size Standard (Applied Biosystems). DNA fragments were separated according to size using a 3130 Genetic Analyzer (Applied Biosystems) and were analyzed with GeneMapper (Applied Biosystems).

Statistical analysis and molecular dating

Haplotype (H) and nucleotide (π) diversities, and mean number of pairwise differences (M) were calculated using DnaSP v.5 software [38]. Arlequin 3.5.1.2 [39] was used to compute AMOVA (Analysis of Molecular Variance) and the significance of the covariance components associated with different levels of genetic structure were tested on haplotype frequencies applying a non-parametric permutation procedure. Population pairwise F_{ST} values, between/within population average nucleotide pairwise differences, and Nei's inter-population distances, were also computed using Arlequin 3.5.1.2 [39].

Diversity indices, phylogeographic inferences and inter-population comparisons were carried out using the sequence range 16090 to 16365, since this is the most commonly reported segment in the literature. Problematic variation located around 16189, which was usually associated with length heteroplasmy, e.g., 16182C or 16183C, was ignored.

Fisher's exact test and Pearson's chi-square test were undertaken using the R package (<http://www.r-project.org/>); a significant nominal value of $\alpha = 0.05$ was considered.

Maximum parsimony trees were built for the complete genomes obtained in the present study and those collected from the literature. For each cluster, the time to the most recent common ancestor (TMRCA) was calculated by computing the averaged distance (ρ) of all the haplotypes in a clade to the respective root haplotype. Heuristic estimates of the standard error (σ) were calculated from an estimate of the genealogy, as done in [40]. Hotspot mutations (16182C, 16183C and 16519) were excluded from calculations (as usual). The corrected evolutionary rate proposed by Soares et al. [41] was used to convert mutational distances into years. The TMRCA values obtained in the present study could be slightly over-estimated as indirectly suggested by estimates obtained in the literature regarding the entrance of the

first Paleo-Indians into the American continent [4,5,9]. Therefore, the TMRCA values obtained here should be validated using a larger number of entire genomes.

Analysis of population structure was undertaken using the software STRUCTURE 2.3.4 [42–44]. Both, burn-in and Markov Chain Monte Carlo (MCMC) repetition were set to a length of 100,000. Parameters were selected as indicated in Heinz et al. [25]. Furthermore, the reference samples obtained from the Human Genome Diversity Cell Line Panel, HGCP-CEPH [45], were used to assist in clustering; this panel constitutes a collection of 556 reference samples representing four main continents: Africa ($n = 105$), Europe ($n = 158$), America ($n = 64$), and East Asia ($n = 229$). Each run was repeated five times from $K = 2$ to $K = 7$. Structure Harvester (<http://taylor0.biology.ucla.edu/structureHarvester/>) was used to estimate optimal K values. CLUMPP 1.1.2 [46] and Distruct 1.1 [47] were used to prepare data for visualization as bar plot representations. R 2.13.0 [48], together with the SNPassoc package [49], was used to run two- and three-dimensional Principal Component Analysis (PCA). We compared and evaluated the results obtained from both approaches.

Snipper [50] (<http://mathgene.usc.es/snipper/>) was used to make four-way predictions of ancestral origin (Africa, Europe, East Asia, and Native Americans) of Bolivian profiles. SNP data collected from HapMap populations were also used as training sets (<http://hapmap.ncbi.nlm.nih.gov>). Prediction was based on maximum likelihood.

Results

Mitochondrial DNA molecular diversity in Bolivia

Several diversity indices have been computed considering different hierarchical levels (**Table 1**). When analyzed by main regions, all seemed to show similar diversity values; with the Department of La Paz harboring the lowest haplotype diversity, and Chuquisaca being the region with the lowest nucleotide diversity. The highest haplotype diversity was found in the North (Pando) but the highest nucleotide diversity is observed in the South (Santa Cruz). Therefore, there is no obvious correlation between departments and molecular diversity as measured by way of statistical summary indices. When examining the diversity by rural and urban populations, it was observed that rural populations harbor higher nucleotide and haplotype diversity than urban populations.

The most apparent geographic pattern was found when examining molecular diversity values by main ecological region. The Andean followed by the Sub-Andean regions had substantial lower diversity values than the Llanos. Therefore, the diversity was found to increase longitudinally, from the high mountains to the lowlands of the Llanos.

Diversity was particularly low in some ethnic groups compared to general urban and rural populations. For instance, the Ayoreo (from Bolivia and Paraguay) and the Aymara had extremely low diversity values compared to the average values observed in Native Bolivians and Bolivians in general (**Table 1**), a fact that is most likely due to strong founder effect in the case of the Ayoreo [19], while in the Aymara discussed by Corella et al. [20] it was most likely due to the small sample size analyzed (the values were compared with those obtained with the Aymara population from Gayá-Vidal et al. [22] using a larger sample size). Our sample from general rural and urban Beni showed significantly higher diversity values than those observed from other Native American groups from the same department (e.g. the Piedmont populations

Table 1. Diversity indices in Bolivian mtDNAs and main American and African regions.

Populations groups	Reference	n	k	S	H	π	M
Department							
Beni	Present study	102	63	71	0.982±0.002	0.02470±0.00130	6.818
Chuquisaca	Present study	89	57	59	0.971±0.009	0.01588±0.00099	5.718
Cochabamba	Present study	103	71	77	0.981±0.007	0.01855±0.00086	6.677
La Paz	Present study	253	121	81	0.960±0.007	0.01864±0.00086	5.106
Pando	Present study	90	54	60	0.983±0.006	0.01977±0.00068	7.117
Santa Cruz	Present study	83	49	46	0.978±0.006	0.02618±0.00097	7.226
All Bolivia	Present study	720	306	134	0.976±0.002	0.02237±0.00041	6.130
Rural vs. Urban							
Rural	Present study	189	108	92	0.980±0.004	0.02440±0.00069	6.709
Urban	Present study	531	240	121	0.975±0.003	0.02225±0.00049	6.098
Regions							
Andean	Present study	240	116	78	0.961±0.007	0.01856±0.00088	5.086
Sub-Andean	Present study	204	118	90	0.973±0.006	0.02215±0.00083	6.092
Llanos	Present study	276	141	100	0.982±0.003	0.02458±0.00059	6.759
Native groups							
Aymara-speakers (Beni)	[20]	10	5	9	0.667±0.163	0.00499±0.00251	1.800
Moseten (Beni)	[20]	10	8	18	0.956±0.059	0.01761±0.00209	6.356
Quechua-speakers (Beni)	[20]	16	7	17	0.692±0.124	0.01094±0.00358	3.950
Chimane (Beni)	[20]	10	8	16	0.933±0.077	0.01422±0.00212	5.133
Aymara-speakers	[22]	96	39	48	0.956±0.009	0.01645±0.00142	4.523
Ayoreo	[19]	91	8	10	0.473±0.061	0.00626±0.00104	2.260
Ignaciano	[18]	15	11	23	0.933±0.054	0.01773±0.00237	6.400
Movina	[18]	12	8	12	0.894±0.078	0.00814±0.00174	2.940
Quechua-speakers	[22]	93	40	48	0.946±0.012	0.01997±0.00118	5.471
Trinitario	[18]	12	11	22	0.985±0.040	0.01884±0.00212	6.803
Yuracare	[18]	15	11	22	0.952±0.040	0.01852±0.00151	6.686
All Native		380	114	84	0.946±0.006	0.02201±0.00043	6.008

The indices were computed using the common segment of the HVS-I region from position 16090 to 16365.

NOTE: n = sample size; K = number of different haplotypes; S = number of segregating sites; H = haplotype diversity; π = nucleotide diversity; M = average number of pairwise differences (mismatch observed mean).

doi:10.1371/journal.pone.0058980.t001

of Moseten, Chimane, etc [20], the Llanos populations analyzed in [18]; see **Table 1**).

There are only few haplotypes in Bolivia that appears more than two or three times in the whole dataset, again indicating the high mtDNA diversity of the country (**Figure S1 and Figure S2**). When evaluating the differentiation between departments, it was again observed that comparisons of other departments with Santa Cruz were the ones that displayed the highest F_{ST} values (**Figure S3**). The influence of this department is also indirectly observed when studying F_{ST} values between the main ecological regions, indicating that the comparisons between Llanos and the other two areas showed the largest values (**Figure S4**).

Finally, **Figure S5 and Figure S6** show the expected (virtual) heterozygosity in the main departments and in the main regions, respectively; the dispersion of values indicated the existence of substantial mtDNA diversity between departments.

Phylogeography

The mtDNA Native American component predominates in the Bolivian population (98.4%); most of the variation can be classified

into one of main Native American haplogroups, (A: 14.3%, B: 52.6%, C: 21.9%, D: 9.6%), **Figure 1**.

There were remarkable differences in haplogroup frequencies between the main Bolivian departments (**Figure 1**). For instance, the department of La Paz had the following haplogroup composition, A: 10%, B: 71%, C: 12%, and D: 7%, which was in sharp contrast with the composition observed in the department of e.g. Beni, A: 12%, B: 30%, C: 48%, and D: 10%. The haplogroup distribution found by Afonso-Costa et al. [21] in a smaller sample from La Paz was slightly different to that found in the Bolivians from La Paz studied here, but both samples agreed regarding the high proportion of haplogroup B in this area. This geographic pattern mirrors in reality the location of the departments at different altitudes; thus, the most important differences are observed between Andean and Sub-Andean populations *versus* Llanos. For instance, haplogroup B is the predominant haplogroup in the Altiplano (71%), which decreases to 61% in Sub-Andean populations and 32% in Llanos (**Figure 1**).

As expected, Bolivia shares more haplotypes with South America than with Meso and North America (**Table 2; Table**

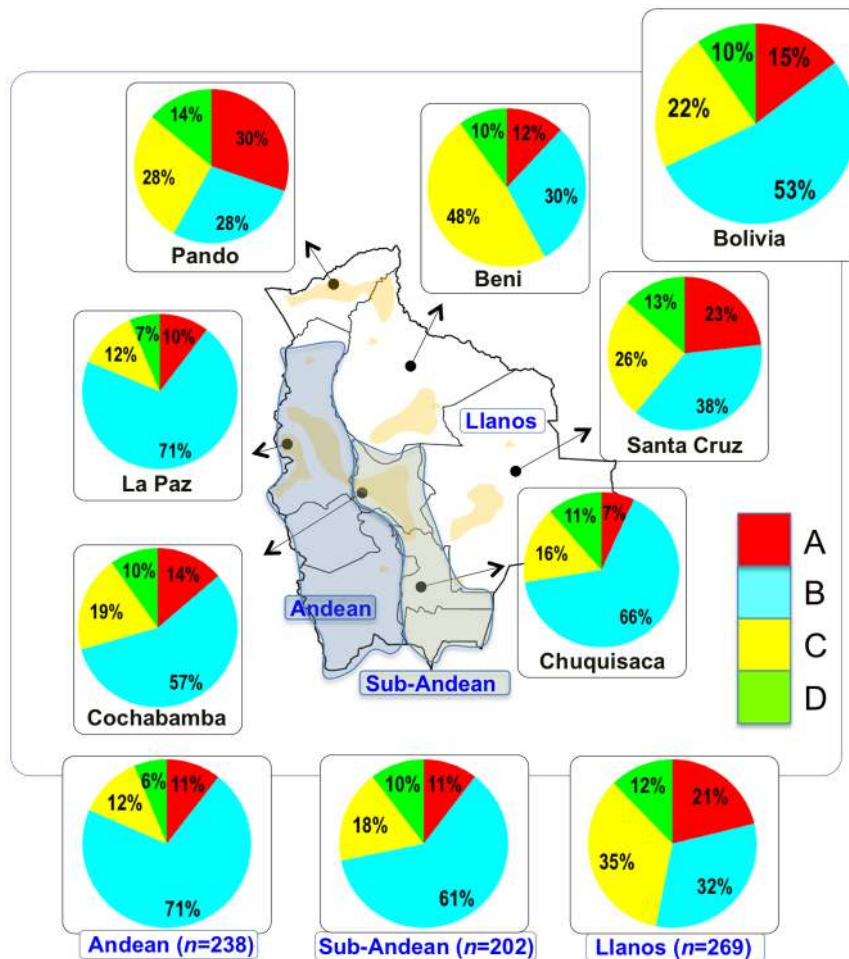


Figure 1. Map of Bolivia showing the location of the samples collected in the present study. The pie charts represent the distribution of basal Native American haplogroup frequencies in the region. doi:10.1371/journal.pone.0058980.g001

S2), although these values have to be interpreted with care due to the different sample sizes. Note however that the largest sample size is in North America, but Bolivia only shares 18% of the haplotypes with this region but 49% with other South American populations (**Table 2**). It is also worth mentioning that 48 haplotypes are shared between the three American regions; therefore, of the 68 haplotypes shared with North America, most of them are also shared with South America (**Table 2, Table S2**). A large amount of haplotypes has been only observed in Bolivia (250 out of 383 different haplotypes; 65%).

There are eight mtDNAs that most likely belong to haplogroup C1d (all carry the characteristic transition 16051); there is another mtDNA that could be allocated to the northern South American cluster C1d2a, which does not carry transition 16091 but carries transition 16209 instead. C1d has been reported as one of the founding Paleo-Indian mtDNA lineages of the American continent [5] and is therefore found along the double continent. As has been previously described [5], C1d did not necessarily follow a main Pacific coastal route once it reached South America; this would explain why this lineage appears more in the Llanos than in the Andean region.

The five mtDNAs belonging to D4h3a found in our Bolivian samples most likely arrived at about the same time as C1d but probably followed a different route. According to [4], D4h3a

appeared to spread along the Pacific coast. In agreement with this hypothesis, the five D4h3a Bolivians were observed in the Andean region, and three out of the five D4h3a lineages belong to the Peruvian sub-clade D4h3a3.

According to expectations, no members of the Paleo-Indian founder X2a, which has a North American distribution, have been found in Bolivians. Also, in good agreement with Bodner et al. [9], we did not find any members belonging to southern cone lineages D1j and D1g, favoring the hypothesis that these two haplogroups crossed the Andes at lower latitudes (Chile towards Argentina), where they were most likely incubated.

Very recently, de Saint Pierre et al. [51] proposed two new lineages, B2i2 and C1b13, which are thought to have originated in the Southern Cone, and are closely associated with the Paleo-Mapuche people in Chile and Argentina. Members of these two lineages have only been sporadically found outside of Chile and Argentina. Here, we only found one member of C1b13, and according to the hypothesis formulated by de Saint Pierre et al. [51], this mtDNA could have arrived in Bolivia from the South instead of from the West.

Some Bolivian mtDNAs are particularly interesting given that they appear at relatively high frequencies in this region and neighboring areas but are absent in the rest of the continent. For instance, the motif 16189 16217 16290 appears six times within

Table 2. Shared haplotypes between Bolivia and other American regions.

	N	SH with Bolivia	H	H ₁
North America	7551	68	1875	1085
Meso-America	2792	68	925	612
South-America	5727	187	1587	1009
Bolivia	1127	383	383	250

Only the segment 16024 to 16385 was considered for comparison. **N** = sample size; **SH** = shared haplotypes; **H** = number of different haplotypes; **H₁** = number of unique haplotypes per region.
doi:10.1371/journal.pone.0058980.t002

haplogroup B4 in different regions from Bolivia; it also appeared in four Andean Peruvian Quechua individuals [52]. The motif 16173 16192 16223 16298 16325 16327 16346 (haplogroup C) appears twelve times in Bolivia, and all of them in the Llanos. Curiously, this motif was found also in another individual from La Paz in [21] and another one from Chaco, in the North of Argentina, close to the Bolivian Llanos [11].

Haplogroup L (L referring to all mtDNA branches excluding macro-haplogroups M and N) mtDNAs represent the sub-Saharan exclusively maternal genetic component of the country, which in this case only accounts for 1% of the total mtDNA pool (**Table S1**). The proportion of African recent ancestry in Bolivia is very low compared to other South American countries that were more influenced by the African slave trade [16,17,53]. There are only seven different haplotypes belonging to some of the typical sub-Saharan sub-clades. Individual #BNI19 belongs to L0a1b2, which is spread in all of sub-Saharan Africa as well as in America, but carries a distinctive transition (within L0a1) at position 16271 (which gets a score of nine in the mutation list of Soares et al. [41]). The L1b (#BNI75) and L1c1a1 (#LaPaz467) profiles most likely come from the West-Central African region. The L1c3b1a and the L3f1b4a profiles appear in Cabinda, Angola and Mozambique [54–57], but also in West-Central, in Gabon [58]. As expected [53,59], the most likely geographic origin of the African Bolivian profiles is west-central Africa, but also encompasses the Southwest and the Southeast.

It is also curious that the European mtDNA component represents less than 1% of the Bolivian population. With the level of resolution of the HVSI segment, it is not possible to assign European lineages to a particular region in Europe; the four sequences observed here belong to haplogroups H1af, K and X1.

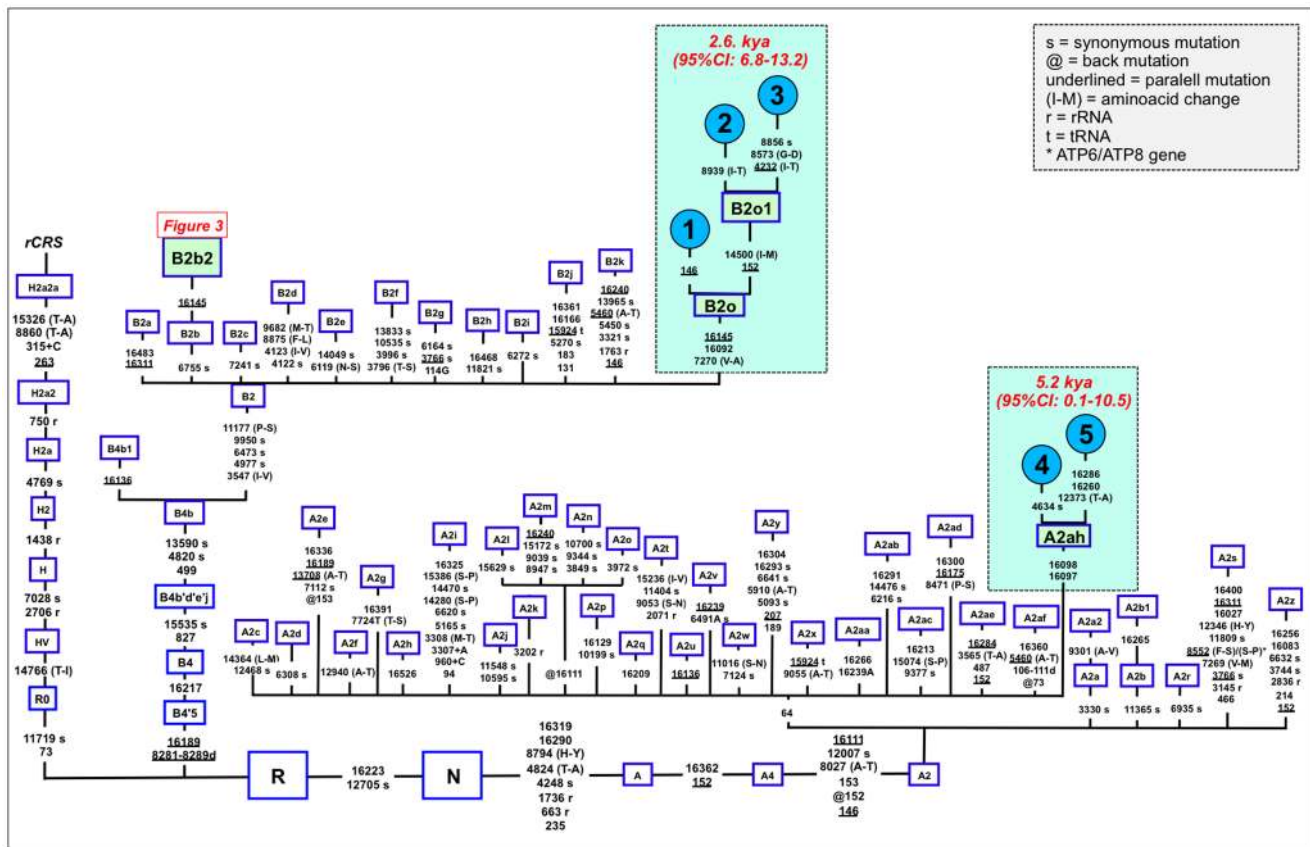


Figure 2. Maximum parsimony tree of the main branches characterizing haplogroups A2 and B2, indicating the new branches generated in the present study, namely, A2ah and B2o. The mutations are displayed along the branches: numbering is according to the rCRS [71]: a mutation alone indicates a transition while an upper case letters (A, C, G, T) indicates a transversion; “+” and “d” refer to insertions and deletions, respectively; “s” refers to a synonymous variant while non-synonymous variants are indicated by way of the amino acid change; a star indicates a variant located in the *ATP6* or *ATP8* gene; a prefix @ indicates a back mutation; and underlined positions indicate parallel mutations. Sample ID (see Table S3) for the entire genomes generated in the present study are as follows: #1: COBIJA577, #2: COBIJA604, #3: STACRUZ261, #4: BENI86, and #5: STACRUZ221.
doi:10.1371/journal.pone.0058980.g002

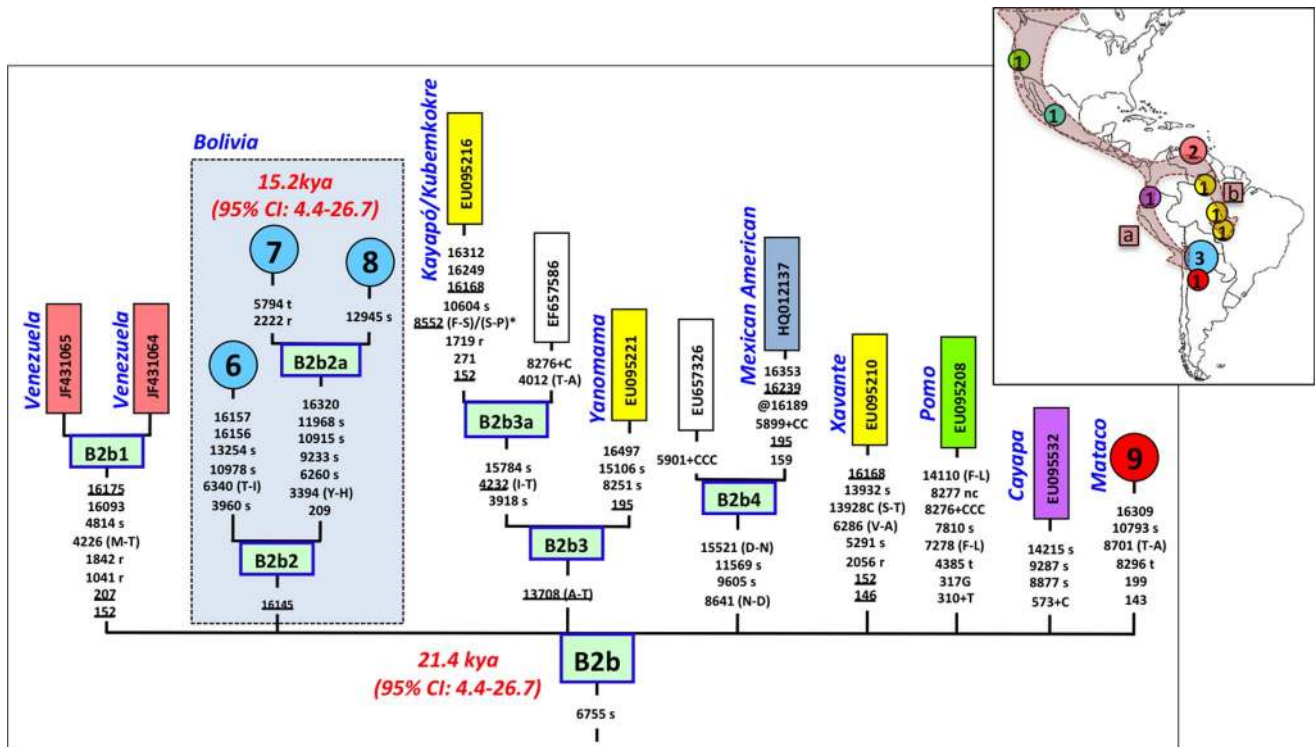


Figure 3. Maximum parsimony tree of haplogroup B2b. The map on the top right inset indicates the geographic location of the B2b genomes represented in the tree; circles are proportional to the sample size. The arrows indicate the two tentative migration routes (a and b in the map) of this lineage (see main text) along the American continent. References for entire genomes are as follows (see **Table S1** and **S3**): #6: BENI90, #7: STACRUZ258, and #8: COBIJA110, #9 Mco-10 (present study, **Table S1** and **S3**). For more information, see the legend of **Figure 2**. doi:10.1371/journal.pone.0058980.g003

An Asian origin for the profile C16104T C16223T cannot be disregarded given that two exact haplotype matches appeared in China [60] but any in Europe (in our in-house database of more than 26700 HVS-I profiles).

Entire genomes from Native American Bolivians

In order to investigate the Native American lineages in our Bolivian samples further, we selected those showing a distinctive pattern from the control region data. Thus, we chose two mtDNAs carrying the tandem variants T16097C A16098G on top of the sequence motif for haplogroup A2, and a group of six mtDNAs all carrying transition G16145A on top of the B2 basal motif (**Table S3**). Entire genome sequencing revealed some interesting phylogeographic features of these mtDNAs.

The tandem transitions T16097C A16098G conform *per se* a very distinctive motif that is found very rarely in mtDNA databases. Within haplogroup A2, this motif alone (without any other control or coding region variant) defines a novel branch of the Native American phylogeny, here named as A2ah (**Figure 2**). The two genomes show variability within this clade and, given its very restrictive location within Bolivia, this most likely point to an origin in Bolivia or the surrounding territories. We further investigated A2ah in public databases of entire mtDNA genomes. We found just one entire genome (JQ702082) that carries the transition A16098G on top of the A2 motif but it differs substantially from the two genomes found in Bolivia (apart from lacking the transition T16097C); therefore, its inclusion within A2ah should be considered to be very dubious. According to the entire genome sequences available, it can be said that haplogroup A2ah originated about 5,200 thousand years ago (kya) (although with a large 95%CI: 0.1–10.5). We additionally investigated the

phylogeographic characteristics of these lineages by looking at the abundant amount of data available in the literature on control region segments. We only detected nine mtDNAs matching the motif of A2ah. Four out of ten were observed in the data reported by Behar et al. [61] but geographic information was not available. Two other HVS-I mtDNAs were found in the ‘Hispanic’ North American subset of the SWGDAM database [62]; three other A2ah mtDNAs were observed in South America, two among the Brazilian samples of Alves-Silva et al. [13], and one in the Toba from Gran Chaco (North Argentina) [63]. In Bolivia, only four A2ah members were found (0.5%), all of which were observed in the Llanos (three in Santa Cruz and one in the Beni department). In general, the HVS-I data suggest that this lineage could have originated in Bolivia or some place in Central South America. The two members observed in the ‘Hispanic’ samples from the SWGDAM could perfectly represent recent immigrations into

Table 3. AMOVA computed based on haplotype pairwise differences of Bolivian populations (significant tests: 20,022 permutations; adjusted *P*-value<0.0000).

	Within Populations	Among Populations
Rural vs. Urban	99.85	0.15
Andean vs. Sub-Andean vs. Llanos	93.15	6.85
Departments	93.44	6.56
Provinces	93.60	6.40

doi:10.1371/journal.pone.0058980.t003

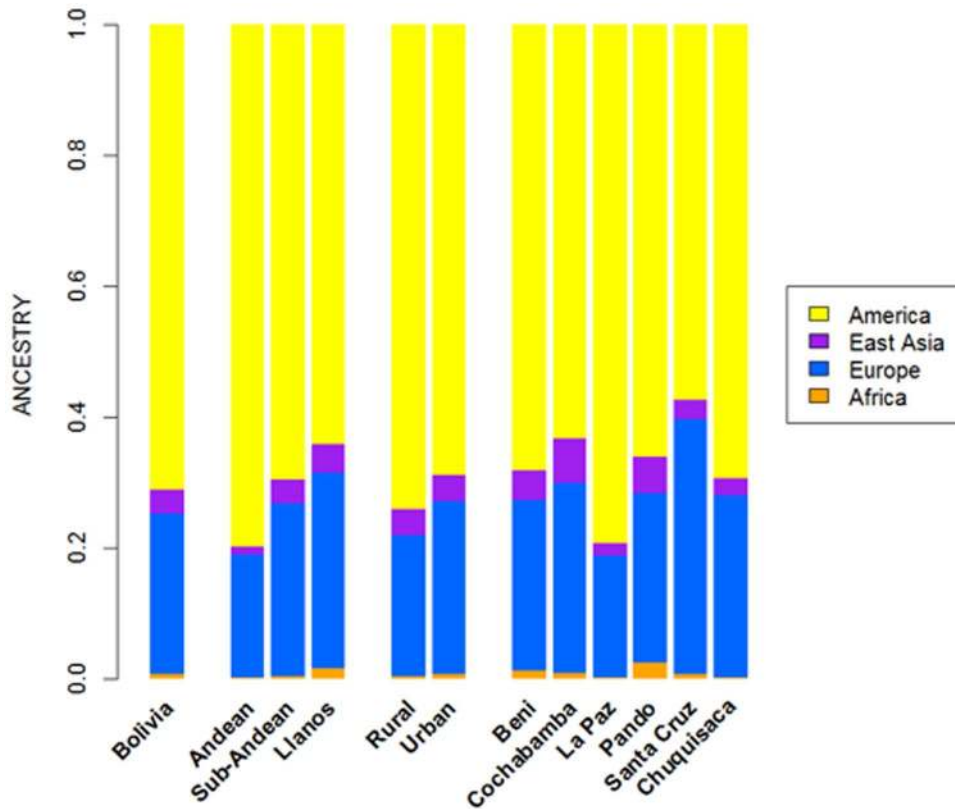


Figure 4. Average continental ancestry of Bolivians analyzed using a panel of 46 AIMs. These values are obtained from the STRUCTURE analysis and the optimal value $K = 4$. doi:10.1371/journal.pone.0058980.g004

USA. The motif T16097C A16098G was also observed in Mainland Scotland (Western Islands; Isla of Skye) and two other samples in [61], but none of the mtDNAs carried the variation defining haplogroup A2.

A phylogenetic conflict exists when trying to reconstruct the most parsimonious tree of the six B2 mtDNAs carrying transition G16145A. Three of the genomes (#6, #7 and #8 in **Figure 3**) carried the synonymous transition G6755A, while three other genomes (#1, #2 and #3) lacked G6755A but carried the non-synonymous transitions T7270C and T16092C instead (**Figure 2**). Given that G16145A seems to have a much higher mutation rate than G6755A (22 *versus* 2 mutational hits in Soares et al. [41], and 24 *versus* 5 in Phylotree; respectively), we decided to resolved the phylogeny as shown in **Figure 2** and **Figure 3**.

Thus, the motif T7270C T16092C G16145A defines a new branch of the Native American phylogeny named here as B2o, which is represented by three Bolivian genomes (**Figure 2**). No other entire genome was observed in public resources belonging to B2o. Dating the phylogeny of B2o (**Figure 2**) indicates that this haplogroup originated about 2.6 kya (95% CI: 6.8–13.2). Searching control region databases, we observed only two HVS-I candidate B2o sequences: one was observed in San Martin de Pangoa in Perú, a small town on the eastern slope of the Andes inhabited by Quechua and Nmatsiguenga people [64], while the other B2o sequence was observed in Guam (an island located in the western Pacific Ocean, territory of USA). The latter, however, carries the motif T16189C T16217C A16247G C16261T, which is very common in the Micronesia, Australia, etc. [65,66] and points to a different clade of the worldwide phylogeny, haplogroup B4a. In Bolivia, we found seven B2o representatives (1%), most of

them in the Llanos (five in Pando and one in Santa Cruz departments), and one in a Sub-Andean locality of La Paz. Therefore, altogether the data suggest that this lineage has also originated in Bolivia or some nearby region, most likely located in the Andes.

The entire genomes of the three Bolivian mtDNAs carrying G16145A but lacking T16092C, all share the transition G6755A. By searching the databases on entire genomes for the transition G6755A, we observed 10 additional mtDNAs belonging to this branch. In reality, this clade had already been described in the recent literature [6,7] and was named B2b. However, its internal variation was never analyzed in detail. **Figure 3** shows the phylogeny of B2b and the geographical location where the entire genomes were observed. Geographic and/or ethnic affiliation was available for eleven entire genomes (including the Bolivian ones): one Pomo (North California, USA) [8], one Mexican American [67], two Venezuelans from Pueblo Llano [68], one Yanomama (Brazil, Venezuela) [8], one Cayapa (Ecuador) [6], one Kayapó/Kubemkore (South Amazonia, Brazil) [8], one Xavante (Brazil) [8], and three Bolivians (present study). We additionally detected one entire genome in our DNA databank in a Mataco Native from North Argentina, which was added to the phylogeny of **Figure 3** (#9). According to the phylogeny of B2b in **Figure 3**, B2b appeared about 21.4 kya (95% CI: 4.4–26.7). The Bolivian sub-clade B2b2 is much younger, approximately 15.2 kya (95% CI 4.4–26.7). In the large HVS-I database, we only observed 14 B2b candidates, almost overlapping the territories represented by the entire genomes. In Bolivia, we counted seven B2b candidates: one Andean (La Paz), four Sub-Andean (Cochabamba, La Paz and Chuquisaca), and two in the Llanos (Beni and Santa Cruz).

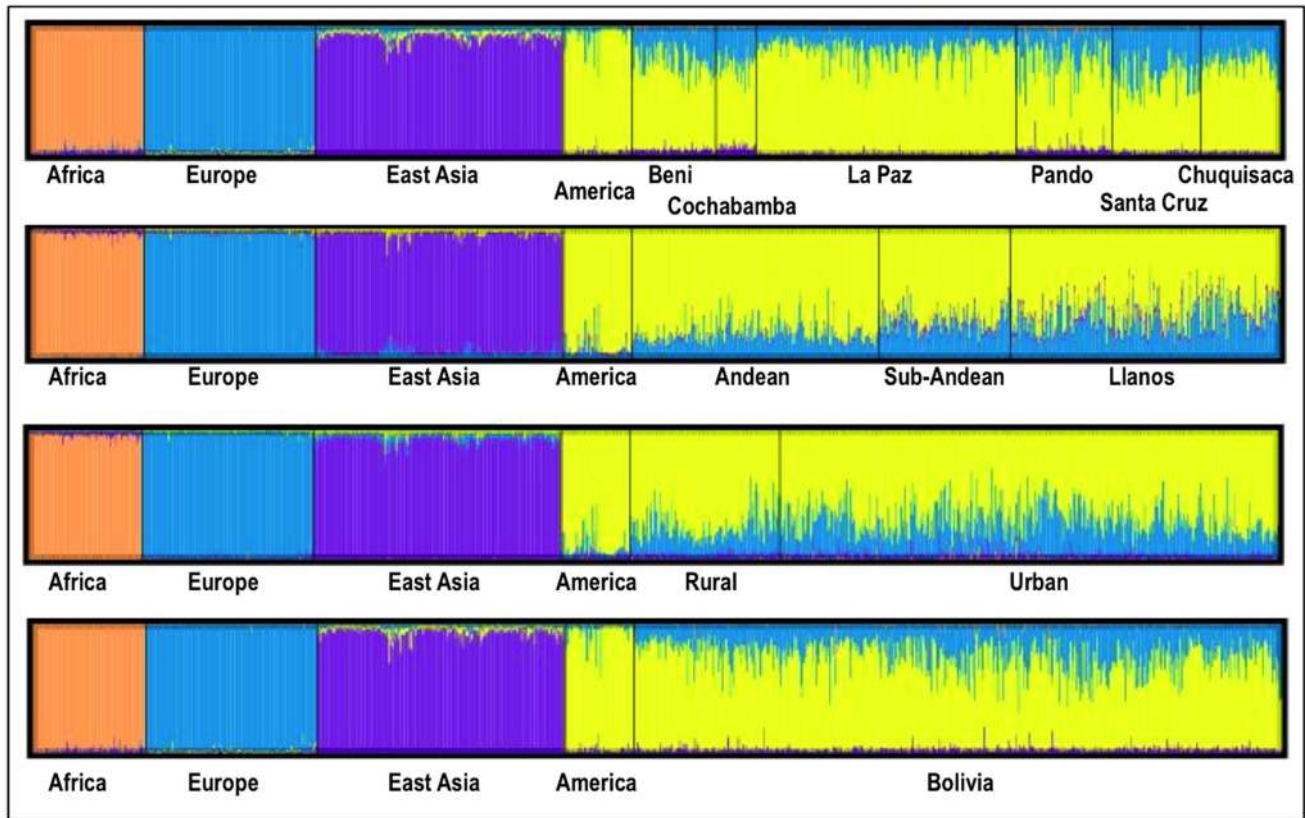


Figure 5. STRUCTURE analysis of Bolivians based on the 46 AIMS panel genotyped in the present study. Each bar plot represent analysis using different grouping schemes for the Bolivian samples. Only the results for the optimal $K = 4$ are represented (see complete analysis in **Text S1**).

doi:10.1371/journal.pone.0058980.g005

AMOVA analysis of mtDNA profiles

AMOVA was undertaken considering different hierarchical levels; by Department, Rural *vs.* Urban, and Andean *vs.* Sub-Andean *vs.* Llanos. As expected, the within-population variation accounted for most of the variance (ranging from 93–100%; **Table 3**), independently of the population sub-division employed.

AMOVA carried out by departments and ecological region provided the highest values of among-population variance (ranging from 10.48 to 11.58% of the total variance). The values are similar given the correlation that exists between these two classification criteria. The rural *vs.* urban division indicates virtually no correlation among group subdivision. The analysis carried out by province did not increase the among-population variance, indicating again that the main factor influencing within-population variation was the different altitudes in the country.

Continental ancestry in Bolivians

A panel of 46 AIMS was genotyped in Bolivians in order to infer their main continental ancestry. Data from main continental reference samples (Africa, East Asia, Europe and America) were used as classification sets.

Analysis of ancestry was carried out using STRUCTURE. **Figure 4** summarizes the average continental ancestries observed in Bolivians under different grouping schemes, while **Figure 5** shows the STRUCTURE bar-plots. The analysis showed that, on average, 71% of the component in the total Bolivian sample is Native American, followed by 25% of European ancestry. When examining the ancestry by departments, La Paz was the region

that showed the highest Native American ancestry (79%) (**Figure 4** and **Figure 5**), and, therefore, the lowest European component (19%). On the other side is the department of Santa Cruz, with 57% of Native American ancestry and 39% European. The African component was very low in all of the departments, showing the highest value in the department of Pando (2.5%), in the North. When examining by ecological regions, it was also evident that the Native American component is higher in the mountainous West (80%), and decreases progressively eastward: sub-Andean (70%) and Llanos (64%). The differences are less apparent when examining rural *vs.* urban areas (74% and 69% of Native American component, respectively).

The STRUCTURE bar-plot of **Figure 5** indicates the ancestral membership for each individual in the source population and the Bolivian samples. It clearly shows that most of the individuals have a Native American ancestry, with only a few exceptions. The high European component of Santa Cruz compared to other departments is also evident in this bar-plot, and is also evident in the Andeans. When looking at the most extreme values, we found that there are only a few that have European ancestry above 50% in most of the departments. It is, however, rare to see individuals with high African membership; the highest values correspond to 23% in one individual from La Paz and another one from Pando (17%).

The membership of individuals into the East Asian cluster is significantly higher in some individuals, as is also evident in the STRUCTURE plots. However, this perhaps mirrors the fact that the separation between the Native American component and the

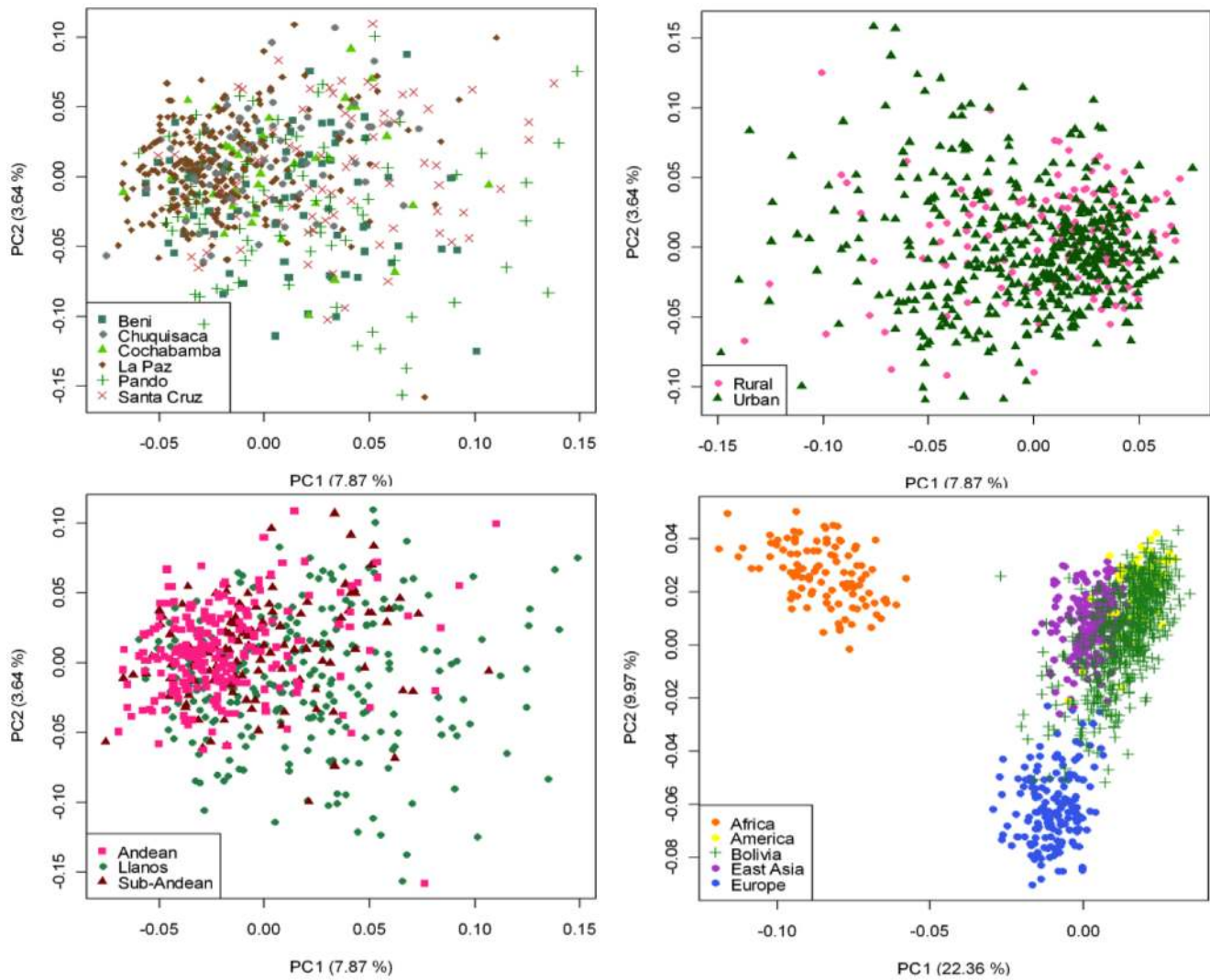


Figure 6. PCA of Bolivian profiles based on the 46-AIMs panel genotyped in the present study. Each figure shows different grouping schemes mainly aiming to show the within-Bolivian diversity. doi:10.1371/journal.pone.0058980.g006

East Asian one is not perfect. The overlap between Native Americans and East Asians can be observed more clearly in the PCA analysis (see below). This fact has been already observed in Pereira et al. [37]. Thus, most of the East Asian component would be captured by the Native American cluster in cases using a three-group structure analysis (Native American, African, and Europe).

The PCA agrees well with the results observed in STRUCTURE. The Bolivian profiles group mainly with the Native American ones from HapMap, with only some samples showing a projection towards the European cluster (Figure 6). The minor African component detected in the STRUCTURE analysis seems not to be relevant in the PCA; only one sample shows some affinity to the African cluster; it is in fact the same sample with a 23% African ancestry in STRUCTURE. The PCA carried out by departments (Text S1) showed that the department of La Paz (Andes) was more tightly grouped than the profiles from other departments; in the other pole are the departments of Santa Cruz and Pando, showing more dispersed patterns (Llanos). This west-east pattern is more evident when observing the PCA by main regions (Figure 6): the Andean profiles are more tightly grouped

in one pole of the plot, while the profiles from Llanos are in the other pole and are more dispersed, with the sub-Andean profiles occupying an intermediate position between East and West. This dispersion is also mirrored when examining the standard deviation (SD) of the Native American membership values in the main regions: (i) Andean: mean = 79.7%, SD = 0.086%; (ii) sub-Andean: mean = 69.6%, SD = 0.09; and (iii) Llanos: mean 64.2%, SD = 0.137% (Table S4); given that the department of La Paz is basically an Andean department, it also shows the lowest SD (0.85%).

Finally, measuring ancestry using AIMS depends on many factors, such as the use of different panels of SNPs (not only different SNPs, but also different numbers of them), different classification test samples, sample sizes, etc. Twelve of the samples genotyped in the present study for the 46-indel panel were also genotyped in Galanter et al. [23], thereby offering a good opportunity to investigate the ability of different AIM panels to estimate percentages of continental ancestry. As shown in Figure 7, the ancestry memberships obtained using Indels lead to a decreased Native American ancestry (and proportional increased

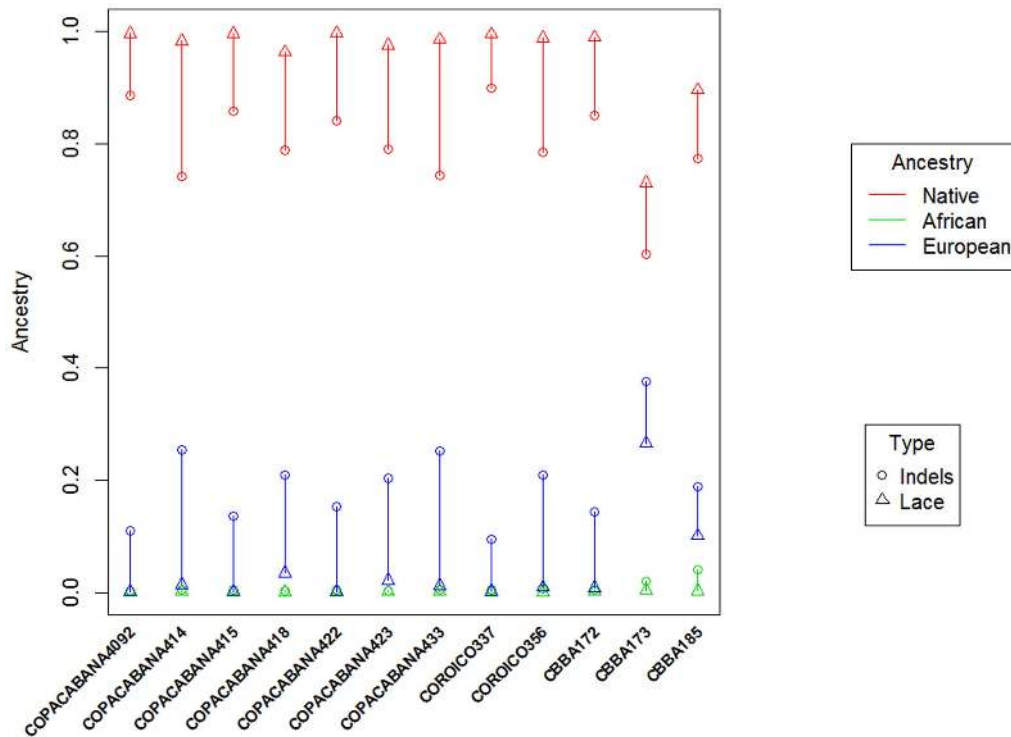


Figure 7. Percentages of main continental ancestry inferred using different AIMs: the 46-Indels panel analyzed in the present study versus the 446-SNP LACE panel analyzed in [23]. The information is only available for 12 individuals from Bolivia. The vertical bars only highlight the difference of ancestry observed using different panels for each individual
doi:10.1371/journal.pone.0058980.g007

European ancestry) compared to the estimates obtained using the LACE panel (on average, the difference is about 15%). Given the way in which the LACE panel was designed (in order to balance the informative content of the different AIMS included in the panel), it seems logical to believe that the Indels panel would tend to over-estimate the proportion of European ancestry in Bolivians. Note, however, that the genome-wide study of Watkins et al. [24] carried out in few Bolivians estimated a higher European component (12%) compared to the LACE. Further genome-wide studies on a larger Bolivian sample will allow better estimates of continental ancestries to be obtained.

Discussion

Bolivia shows a mainly Native American mtDNA component (98%). Only 1.5% of the profiles have Sub-Saharan mtDNA ancestry. The impact of Europeans in the Bolivian mtDNA pool is minimal (0.5%), which also contrasts with other South American locations [11,27]. Although the Native American mtDNA component predominates in the country, there is a highly diverse and geographic stratification in the country. In a large geographical scale, the largest difference corresponds to Llanos *vs.* the Andean and Sub-Andean regions. The political definition of the departments overlap quite well with the main latitudes observed in the country, which would explain the correspondence observed when carrying out different statistical analysis. Molecular diversity is extremely low in some indigenous group compared to urban and other rural populations, suggesting the existence of important episodes of isolation and genetic drift in Native communities. AMOVA carried out at Bolivian populations to different hierarchical levels allowed a better understanding of the spatial

geographic patterns of variability in large regions. The results agree that most of the variation accounts within populations, but that the major differentiation occurs between the Andes and the Llanos. In addition, some mtDNA phylogeographic features indicate the presence of common lineages between different Andean regions in Peru and Bolivia, most likely testifying for the common demographic past during the Inca's Empire period. This continuity in the Andean region was also observed in the Aymaras and Quechuas analyzed by Gayá-Vidal et al. [22].

By way of sequencing the entire genome of nine mtDNAs (eight from Bolivia and one from a northern Argentinean Native Mataco), and compiling a large amount of data from the literature, we could shed light on three Native American lineages: A2ah, B2o and B2b. The three haplogroups are rare in Bolivia (ranging from 0.5 and 1%) and are very rare continentally. The three clades show significant diversity within the Bolivian country, indicating that they most likely evolved locally after their arrival from other neighboring regions. The clade for which there are more data available (literature and present study) is B2b. This haplogroup most likely arose in North California; however, as suggested by its TMCRA (~21.4 kya; see also [67]) and its phylogeographic characteristics, it could have been originated even in the north most region of the American continent, constituting a new minor Paleo-Indian founder. The data also indicate that B2b traveled South, most likely following the Pacific coastline (the main route followed by the First Americans) [4,5,9,69]. It probably entered the South American sub-continent following two different paths: (i) travelling further south following the Pacific side, then reaching Ecuador, Peru, and Bolivia; and (ii) following firstly an eastwards direction, and secondly southwards crossing the Amazon basin in Venezuela and then Brazil (or firstly

bordering the Atlantic coast southwards). At the very least, these lineages arrived in northern Argentina, but it is likely that further sampling will probably detect B2b in the most southern edge of the continent. The data indicate that the Bolivian sub-lineage B2b2 arrived in this region about 15,000 years ago.

Ancestry analyses carried out on a panel of AIMs indicated that Bolivians have a substantial European ancestry (although a possibility exists that this proportion could be over-estimated; see above), which varies substantially between departments, thus indicating a differential regional impact of the European Colonialism in Bolivia. The African component is very low (in agreement with the mtDNA) as compared to other South American populations, such as the Caribbean coast [17,70], Colombia [16], and Brazil [13], but is comparable to others such as Argentina [10,11,27]. It must be taken into account that the present study did not include samples from the Yungas, a small province located within the department of La Paz that seems to concentrate the main African component in Bolivians. The African component of the Yungas people is not only inferred from their distinguishable African cultural and biological features, but also from the genetic inferences made using panels of AIMs analyzed in a small group of people from this region [23].

The present study represents the largest and most comprehensive analysis of genetic variation investigated to date in rural and urban Bolivians from the point of view of mtDNA and autosomal data. The results indicated that even those Bolivians that did not self-identify as belonging to a particular Native America ethnic group still preserved a main Native American genetic character in their genomes. Deep analyses of selected mtDNA genomes also indicate the presence of lineages that appears to be autochthonous to these peoples; provided that these lineages have not been identified in large American databases. The Native American component of Bolivians is significantly higher when observing the mtDNA instead of the autosomes. This is a common feature in other South and Central American regions (e.g. Argentina [10], Panama [2]) where the maternal Native American component has been found to be much better preserved in the maternal specific genome, a fact that is generally explained by the higher proportion of European males arriving to America than females.

Supporting Information

Figure S1 Frequency of different haplotypes (alleles in figure) by department.
(TIFF)

Figure S2 Frequency of different haplotypes (alleles in figure) by ecological region.
(TIFF)

References

- Lewis MP (2009) *Ethnologue. Languages of the world*. Dallas, Texas: SIR International.
- Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, et al. (2012) Deciphering the mitochondrial gene pool of modern panamanians. *PLoS One* 7: e38337.
- Molina Barrios R, editor (2005) *Los pueblos indígenas de Bolivia: diagnóstico sociodemográfico a partir del censo del 2001*. Santiago de Chile: Naciones Unidas.
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, et al. (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19: 1–8.
- Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, et al. (2010) The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20: 1174–1179.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, et al. (2007) Beringian standstill and spread of Native American founders. *PLoS ONE* 2: e829.
- Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, et al. (2008) The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* 3: e1764.
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, et al. (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82: 583–592.
- Bodner M, Perego UA, Huber G, Fendt L, Röck AW, et al. (2012) Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res* 22: 811–820.
- Salas A, Jaime JC, Álvarez-Iglesias V, Carracedo Á (2008) Gender bias in the multi-ethnic genetic composition of Central Argentina. *J Hum Genet* 53: 662–674.
- Bobillo MC, Zimmermann B, Sala A, Huber G, Rock A, et al. (2010) Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. *Int J Legal Med* 74: 65–76.

Figure S3 Fst values between departments.
(TIFF)

Figure S4 Fst values between ecological regions.
(TIFF)

Figure S5 Expected (virtual) heterozygosity by departments.
(TIFF)

Figure S6 Expected (virtual) heterozygosity by main ecological regions.
(TIFF)

Table S1 Mitochondrial DNA sequencing data for the Bolivian samples analyzed in the present study. Note that for a large proportion of sequences the range is generally larger than what it is usually considered to be HVS-I and HVS-II.
(XLSX)

Table S2 Shared haplotypes between Bolivia and other American locations.
(XLSX)

Table S3 Information on entire genome sequences obtained in the present study and collected from the literature and GenBank.
(XLSX)

Table S4 Continental ancestry values (standard deviations in brackets) obtained using STRUCTURE for the Bolivian samples averaged through five different iterations (see Material and Methods for more information). The minimum and the maximum values are also given.
(XLSX)

Text S1 Additional PCA and structure analyses carried out in the Bolivian samples based on 46 AIMs.
(DOCX)

Acknowledgments

We would like to thank the donors for their participation in the present project. The authors have declared that no competing interests exist.

Author Contributions

Draft the manuscript: AS. Agreed on the final version of the manuscript: PTE VÁI TH LVB AGC LC JPS AP ATB OR CV AS. Conceived and designed the experiments: AS. Performed the experiments: PTE VÁI TH LVB AGC LC AP. Analyzed the data: PTE VÁI TH LVB AGC JPS AS. Contributed reagents/materials/analysis tools: AC ATB OR CV AS. Wrote the paper: AS.

12. Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, et al. (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.
13. Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt H-J, et al. (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67: 444–461.
14. Bortolini MC, Salzano FM, Zago MA, Da Silva Junior WA, Weimer Tde A (1997) Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am J Phys Anthropol* 103: 147–156.
15. Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, et al. (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* 67: 1287–1295.
16. Salas A, Acosta A, Álvarez-Iglesias V, Cerezo M, Phillips C, et al. (2008) The mtDNA ancestry of admixed Colombian populations. *Am J Hum Biol* 20: 584–591.
17. Salas A, Richards M, Lareu MV, Sobrino B, Silva S, et al. (2005) Shipwrecks and founder effects: Divergent demographic histories reflected in Caribbean mtDNA. *Am J Phys Anthropol* 128: 855–860.
18. Bert F, Corella A, Gene M, Perez-Perez A, Turbon D (2004) Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. *Ann Hum Biol* 31: 9–28.
19. Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, et al. (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.
20. Corella A, Bert F, Perez-Perez A, Gene M, Turbon 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.
21. Afonso-Costa H, Carvalho M, Lopes V, Balsa F, Bento AM, et al. (2010) Mitochondrial DNA sequence analysis of a native Bolivian population. *J Forensic Leg Med* 17: 247–253.
22. Gayá-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, et al. (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.
23. Galanter JM, Fernández-López JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, et al. (2012) Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet* 8: e1002554.
24. Watkins WS, Xing J, Huff C, Witherspoon DJ, Zhang Y, et al. (2012) Genetic analysis of ancestry, admixture and selection in Bolivian and Totonac populations of the New World. *BMC Genet* 13: 39.
25. Heinz T, Álvarez-Iglesias V, Taboada-Echalar P, Gómez-Carballa A, Torres-Balanza A, et al. (2012) Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians. *Forensic Sci Int Genet* submitted.
26. Cerezo M, Bandelt H-J, Martin-Guerrero I, Ardanaz M, Vega A, et al. (2009) High mitochondrial DNA stability in B-cell chronic lymphocytic leukemia. *PLoS One* 4: e7902.
27. Catelli ML, Álvarez-Iglesias V, Gomez-Carballa A, Mosquera-Miguel A, Romanini C, et al. (2011) The impact of modern migrations on present-day multi-ethnic Argentina as recorded on the mitochondrial DNA genome. *BMC Genet* 12: 77.
28. Brisighelli F, Capelli C, Álvarez-Iglesias V, Onofri V, Paoli G, et al. (2009) The Etruscan timeline: A recent Anatolian connection. *Eur J Hum Genet* 17: 693–696.
29. Salas A, Carracedo Á, Macaulay V, Richards M, Bandelt H-J (2005) A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. *Biochem Biophys Res Commun* 335: 891–899.
30. Bandelt H-J, Salas A, Bravi CM (2004) Problems in FBI mtDNA database. *Science* 305: 1402–1404.
31. Bandelt H-J, Salas A, Lutz-Bonengel S (2004) Artificial recombination in forensic mtDNA population databases. *Int J Legal Med* 118: 267–273.
32. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465.
33. Salas A, Coble M, Desmyter S, Grzybowski T, Gusmao L, et al. (2012) A cautionary note on switching mitochondrial DNA reference sequences in forensic genetics. *Forensic Sci Int Genet* 6: e182–184.
34. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30: E386–394.
35. Bandelt H-J, van Oven M, Salas A (2012) Haplogrouping mitochondrial DNA sequences in Legal Medicine/Forensic Genetics. *Int J Legal Med* 126: 901–916.
36. Kong Q-P, Bandelt H-J, Sun C, Yao Y-G, Salas A, et al. (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 15: 2076–2086.
37. Pereira R, Phillips C, Pinto N, Santos C, Santos SEBd, et al. (2012) Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing. *PLoS ONE* 7: e29684–e29684.
38. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
39. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
40. Saillard J, Forster P, Lynnerup N, Bandelt H-J, Norby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67: 718–726.
41. Soares P, Ermini L, Thomson N, Mormina M, Rito T, et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84: 740–759.
42. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
43. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
44. Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources* 9: 1322–1332.
45. Cann HM, Toma Cd, Cazes L, Legrand M-F, Morel V, et al. (2002) A Human Genome Diversity Cell Line Panel. *Science* 296: 261–262.
46. Jakobsson M, Rosenberg NA (2007) CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure. *Bioinformatics* 23: 1801–1806.
47. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. (2002) Genetic Structure of Human Populations. *Science* 298: 2381–2385.
48. Team TRC (2012) R: A Language and Environment for Statistical Computing. In: The RCT, editor.
49. González JR, Armengol L, Solé X, Guinó E, Mercader JM, et al. (2007) SNPpass: an R package to perform whole genome association studies. *Bioinformatics* 23: 644–645.
50. Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, et al. (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* 1: 273–280.
51. de Saint Pierre M, Gandini F, Perego UA, Bodner M, Gómez-Carballa A, et al. (2012) Arrival of paleo-indians to the southern cone of South America: new clues from mitogenomes. *PLoS One* 7: e51311.
52. Barbieri C, Heggarty P, Castri L, Luiselli D, Pettener D (2011) Mitochondrial DNA variability in the Titiaca basin: Matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol* 23: 89–99.
53. Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, et al. (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* 74: 454–465.
54. Beleza S, Gusmao L, Amorim A, Carracedo Á, Salas A (2005) The genetic legacy of western Bantu migrations. *Hum Genet* 117: 366–375.
55. Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J (2009) On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol Biol* 9: 80.
56. Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, et al. (2002) The making of the African mtDNA landscape. *Am J Hum Genet* 71: 1082–1111.
57. Pereira L, Macaulay V, Torroni A, Scozzari R, Prata M-J, et al. (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* 65: 439–458.
58. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, et al. (2008) Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A* 105: 1596–1601.
59. Salas A, Carracedo Á, Richards M, Macaulay V (2005) Charting the Ancestry of African Americans. *Am J Hum Genet* 77: 676–680.
60. Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, et al. (2007) Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol* 134: 481–488.
61. Behar DM, Rosset S, Blue-Smith J, Balanovsky O, Tzur S, et al. (2007) The Genographic Project public participation mitochondrial DNA database. *PLoS Genet* 3: e104.
62. Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2002) The mtDNA Population Database: an integrated software and database resource for forensic comparison. *Forensic Sci Commun* 4: no 2.
63. Cabana GS, Merriwether DA, Hunley K, Demarchi DA (2006) Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. *Am J Phys Anthropol* 131: 108–119.
64. Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, et al. (2003) Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol* 20: 1682–1691.
65. Murray-McIntosh RP, Scrimshaw BJ, Hatfield PJ, Penny D (1998) Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences. *Proc Natl Acad Sci U S A* 95: 9047–9052.
66. Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, et al. (1995) Evolutionary history of the COII/tRNA(Lys) intergenic 9-base-pair deletion in human mitochondrial DNAs from the Pacific. *Molecular Biology and Evolution* 12: 604–615.
67. Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, et al. (2011) Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol Biol* 11: 293.

68. Gómez-Carballa A, Ignacio-Veiga A, Álvarez-Iglesias V, Pastoriza-Mourelle A, Ruiz Y, et al. (2012) A melting pot of multicontinental mtDNA lineages in admixed Venezuelans. *Am J Phys Anthropol* 147: 78–87.
69. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, et al. (2012) Reconstructing Native American population history. *Nature*; in press.
70. Mendizabal I, Sandoval K, Berniell-Lee G, Calafell F, Salas A, et al. (2008) Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol Biol* 8: 213.
71. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147.