Forensic Science International: Genetics 7 (2013) 537–542

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians

Tanja Heinz [a,1], Vanesa Álvarez-Iglesias [a], Jacobo Pardo-Seco [a], Patricia Taboada-Echalar [a], Alberto Gómez-Carballa [a], Antonio Torres-Balanza [b], Omar Rocabado [b], Ángel Carracedo [a], Carlos Vullo [c,d], Antonio Salas [a,1,*]

[a] Unidade de Xenética, Instituto de Ciencias Forenses and Departamento de Anatomía Patolóxica e Ciencias Forenses, Grupo de Medicina Xenómica (GMX), Facultade de Medicina, Universidade de Santiago de Compostela, 15872, Galicia, Spain
[b] Instituto de Investigaciones Forenses, Fiscalía General del Estado Plurinacional de Bolivia, La Paz, Bolivia
[c] Equipo de Argentina de Antropología Forense, Córdoba, Argentina
[d] Laboratorio de Inmunogenética y Diagnóstico Molecular, Córdoba, Argentina

ARTICLE INFO

Article history:
Received 25 December 2012
Received in revised form 10 May 2013
Accepted 30 May 2013

Keywords:
Bolivia
AIMs
Ancestry
SNPs
Indels

ABSTRACT

We have genotyped 46 Ancestry Informative Markers (AIMs) in two of the most populated areas in Bolivia, namely, La Paz (Andean region; $n = 105$), and Chuquisaca (Sub-Andean region; $n = 73$). Using different analytical tools, we inferred admixture proportions of these two American communities by comparing the genetic profiles with those publicly available from the CEPH (Centre d'Etude du Polymorphisme Humain) panel representing three main continental groups (Africa, Europe, and America). By way of simulations, we first evaluated the minimum sample size needed in order to obtain accurate estimates of ancestry proportions. The results indicated that sample sizes above 30 individuals could be large enough to estimate main continental ancestry proportions using the 46 AIMs panel. With the exception of a few individuals, the results also indicated that Bolivians showed a predominantly Native American ancestry with variable levels of European admixture. The proportions of ancestry were statistically different in La Paz and Chuquisaca: the Native American component was 86% and 77% (Mann–Whitney *U*-test: un-adjusted *P*-value = 2.1×10^{-5}), while the European ancestry was 13% and 21% (Mann–Whitney *U*-test: un-adjusted *P*-value = 3.6×10^{-5}), respectively. The African ancestry in Bolivians captured by the AIMs analyzed in the present study was below 2%. The inferred ancestry of Bolivians fits well with previous studies undertaken on haplotype data, indicating a major proportion of Native American lineages. The genetic differences observed in these two groups suggest that forensic genetic analysis should be better performed based on local databases built in the main Bolivian areas.

1. Introduction

The ethnic composition of Bolivia includes a great diversity of cultures. About 30.7% of Bolivians are Quechua-speaking (1.6 million) and 25.2% are Aymara-speaking (1.3 million). Other important ethnic groups are the Chiquitano (2%; about 110,000 Bolivians) and the Guaraní (1.5%; about 78,000). The census indicates that about 55% of the population is Native (5 million), while 30% are 'mestizo' (a mix of indigenous people and Europeans or those of European descent) and 15% are 'white' (which are thought to be descendants of Europeans).

In the context of human population genetics, a large proportion of the genotyping efforts carried out in South American populations have been devoted to the analysis of mitochondrial DNA (mtDNA) molecules [1–7] (see also other none mtDNA studies [8]). Also, in Bolivia, most of the genetic studies to date have been carried out on the mtDNA [9–13], and only a few have utilized autosomal SNPs [14,15]. These studies revealed that the main mtDNA ancestry of Bolivians is Native American, with very little contribution from Europeans and Africans. The mtDNA, however, provides only a partial picture of the population ancestry of an individual given that this molecule realistically represents only a single locus. Moreover, admixture is particularly complex in Bolivia, as it is the case for most of the American countries, like e.g. in Brazil [16–19].

* Corresponding author at: Unidade de Xenética, Instituto de Medicina Legal and Departamento de Anatomía Patolóxica e Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, Calle San Francisco sn, 15872 Santiago de Compostela, Galicia, Spain. Tel.: +34 982 820 000; fax: +34 881 812 459.
E-mail address: antonio.salas@usc.es (A. Salas).
[1] These authors contributed equally to this work.

1872-4973/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2013.05.012

In order to contribute to a better understanding of the genomic ancestral structure of Bolivians, we have analyzed two representative samples from two of the most important demographic areas of the country, the departments of La Paz and Chuquisaca. The city of La Paz is the seat of the government of Bolivia and it is also the name of the La Paz department. It is the second largest city in the country and it is located at an elevation more than 3600 m above sea level. Chuquisaca is a department located in the center-south of the country that borders on the Department of Cochabamba, Tarija, Potosí and Santa Cruz. The main city of this department is Sucre, which is the official capital of Bolivia. Chuquisaca is traversed by the main cordillera of the mountain range of the Andes, but part of the department lies within the basin of the Amazon River and other parts within the basin of the Río de La Plata.

The aim of the present study is to characterize the admixture proportions of two of the main departments in Bolivia by genotyping a panel of 46 ancestry informative markers (AIMs). An important effort of the present study was devoted to a simulation study specifically designed to investigate the dependence of ancestry inferences on sample sizes. The results are of special interest to the field of forensic genetics since they provide the basis for future studies aimed at understanding the patterns of variability in Bolivia and their implications in routine forensic casework.

2. Materials and methods

2.1. Population samples

Our study was performed using 734 DNA samples in total: (1) 178 samples from Bolivian populations and (2) 556 reference samples from the Human Genome Diversity Cell Line Panel, HGCP-CEPH [20]. Samples from Bolivia came from two areas: Chuquisaca in the Sub-Andean region ($n = 73$) and La Paz in the Andean region ($n = 105$). The samples of the HGCP-CEPH panel came from populations of four different continents: Africa ($n = 105$), Europe ($n = 158$), America ($n = 64$), and East Asia ($n = 229$). In preliminary analyses, East Asia (data also collected from the CEPH) was included in the group of classification samples in order to model the impact of modern East Asian immigrants in Bolivia; however, admixture analysis indicated that East Asian does not have a significant contribution to our Bolivian samples (data not shown). Therefore, East Asia was eliminated from the final analyses shown in the present study.

DNA was extracted from bloodstains using standard phenol-chloroform procedures. The geographical origins and the genotyping data of all the samples analyzed in the present study are indicated in Table S1.

2.2. AIM-INDEL genotyping

All of the DNA samples were genotyped for 46 INDEL markers [21]. These markers have shown previous ability to distinguish between populations belonging to four of the main continental regions. The AIMs were genotyped in a 46-multiplex PCR amplification, followed by capillary electrophoresis. Each AIM-Indelplex PCR amplification was performed with 5 μl 2× Qiagen Multiplex PCR Master Mix, 10× Primer Mix, and 0.5 μl DNA (concentration between 0.5 and 5 ng/μl) in a final volume of 10 μl. PCR thermocycling conditions were: initial step at 95 °C for 15 min; 28 cycles at 94 °C for 30 s, 60 °C for 90 s, and 72 °C for 60 s; and the final step at 72 °C for 60 min. Following amplification, 0.8 μl PCR product was added to 11.5 μl Hi-Di Formamide (Applied Biosystems) and 0.3 μl Liz-500 Size Standard (Applied Biosystems). DNA fragments were separated according to size on a 3130 Genetic Analyzer (Applied Biosystems) and analyzed using the GeneMapper software (Applied Biosystems).

2.3. Simulations on ancestry

A simulation experiment has been designed in order to estimate the dependence of ancestry inferences on sample sizes. For each of the two Bolivian sets of genetic profiles, we randomly take 1000 sub-samples of variable sizes (from 5 to 70 profiles; in stepwise increments of 5 and taken without replacement). Thus, for example, we obtained 1000 sub-samples of size 5, also 1000 sub-samples of size 10, and so on until a maximum sample size of 70. For each of the sub-samples we computed ancestry proportions as indicated in the next section. Continental ancestry estimates were obtained as the mean values calculated from the 1000 sub-samples of each sample window (i.e. mean values for the 1000 sub-samples of size 5, and so on) and bootstrapping intervals were built accordingly (for each sample window). Simulations were carried out using R 2.13.0 [22].

2.4. Statistical analysis

We conducted our analysis of population structure using Admixture v. 1.22 [23]. Admixture was run using default parameters, and cross validation errors were obtained in order to determine the most likely K value. Admixture values were represented as bar plots using R 2.13.0 [22].

Plink v.1.07 [24] was used to obtained Identity By State (IBS) values between individuals. IBS values were used to carry out two-dimensional Principal Component Analysis (PCA).

In addition, three-way predictions of ancestral origin (Africa, Europe, and Native Americans) of Bolivian profiles were also carried out using Snipper (http://mathgene.usc.es/snipper/), as previously reported [25]. Snipper uses SNP data collected from HapMap populations as training sets (http://hapmap.ncbi.nlm.nih.gov). Prediction was based on maximum likelihood.

3. Results

3.1. Estimates of continental ancestry in Bolivia

Fig. 1 shows a map indicating where the Bolivian samples analyzed in the present study were sampled; the pie charts indicate the values of main continental ancestry obtained when using the total sample from La Paz and Chuquisaca (see below).

A bar plot of ancestry values were built for all the individuals used in the present study (Fig. 2); the analyses were carried out using different values of K. According to the cross validation procedure implemented in Admixture, the optimum K value for the data analyzed was three (Fig. S1). These three components perfectly separate the SNP profiles belonging to each of the main ancestral continental populations: African, European, and Native American (Fig. 2). Average values of Native American, European and African ancestry are 86%, 12.5%, and 1.5%, in individuals from La Paz and 76.8%, 21.4%, and 1.8% in individuals from Chuquisaca; respectively. These proportions were statistically different between these two populations for the Native American (86% vs. 77%; Mann–Whitney U-test: un-adjusted P-value = 2.1×10^{-5}), and the European components (13% vs. 21%; Mann–Whitney U-test: un-adjusted P-value = 3.6×10^{-5}).

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.fsigen.2013.05.012.

Ancestry values are however variable when examined individually (Fig. 3). There are 20 and 6 individuals from La Paz and Chuquisaca, respectively, having ~100% of Native ancestry. The