Forensic Science International: Genetics 33 (2018) 59–65

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

The geographic mosaic of Ecuadorian Y-chromosome ancestry

U. Toscanini[a,b,1], A. Gaviria[c], J. Pardo-Seco[b,d,e], A. Gómez-Carballa[b,d,e], F. Moscoso[b,f], M. Vela[c], S. Cobos[c], A. Lupero[c], A.K. Zambrano[c], F. Martinón-Torres[d,e], A. Carabajo-Marcillo[f], R. Yunga-León[f], N. Ugalde-Noritz[f], A. Ordoñez-Ugalde[b,f,g], A. Salas[b,*,1]

[a] Pricai-Fundación Favaloro, Buenos Aires, Argentina
[b] Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, and GenPoB Research Group, Instituto de Investigaciones Sanitarias (IDIS), Hospital Clínico Universitario de Santiago, Galicia (SERGAS), Spain
[c] Centro de Investigación Genética y Genómica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad Tecnológica Equinoccial, Quito, 1701129, Ecuador
[d] Translational Pediatrics and Infectious Diseases, Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain
[e] GENVIP Research Group, Instituto de Investigación Sanitaria de Santiago, Galicia, Spain[2]
[f] Laboratorio Biomolecular, Cuenca, Ecuador
[g] Neurogenetics Group, FPGMX-IDIS, Santiago de Compostela, Spain

ARTICLE INFO

Keywords:
Ecuador
Y-chromosome
Ancestry
Population-stratification

ABSTRACT

Ecuadorians originated from a complex mixture of Native American indigenous people with Europeans and Africans. We analyzed Y-chromosome STRs (Y-STRs) in a sample of 415 Ecuadorians (145 using the AmpF*l*STR® Yfiler™ system [Life Technologies, USA] and 270 using the PowerPlex®Y23 system [Promega Corp., USA]; hereafter Yfiler and PPY23, respectively) representing three main ecological continental regions of the country, namely Amazon rainforest, Andes, and Pacific coast. Diversity values are high in the three regions, and the PPY23 exhibits higher discrimination power than the Yfiler set. While summary statistics, AMOVA, and R_{ST} distances show low to moderate levels of population stratification, inferred ancestry derived from Y-STRs reveal clear patterns of geographic variation. The major ancestry in Ecuadorian males is European (61%), followed by an important Native American component (34%); whereas the African ancestry (5%) is mainly concentrated in the Northwest corner of the country. We conclude that classical procedures for measuring population stratification do not have the desirable sensitivity. Statistical inference of ancestry from Y-STRS is a satisfactory alternative for revealing patterns of spatial variation that would pass unnoticed when using popular statistical summary indices.

1. Introduction

Ecuador is located in the Northwest of South America, bordering with Colombia to the North, with Peru to the South and East, and with the Pacific Ocean to the West. In continental Ecuador, there are three main ecological regions: (i) the Pacific coast, (ii) the Andes ("Sierra"), and (iii) the Amazon rainforest. The full mainland territory is politically divided into 23 different provinces. The population of Ecuador is approximately 14 million people (INEC; http://www.inec.gob.ec/estadisticas/). According to the official 2010 census, Ecuadorians self-identify as "mestizos" (71.9%), "montubios" (7.4%; referred in Ecuador as a kind of mestizo people living in the coastal countryside that originated from a mixture of Native Americans, Europeans, and Africans), "Afro-descendant" people (7.2%; "Afroecuatorianos"), Native Americans (7%; "Indígenas") with about 14 ethnic groups, and European descendants (6.1%; "blancos"). As in other neighboring countries (mainly, Bolivia and Peru) the Native American component is large, compared to other South American countries that have a higher European (e.g. Argentina, Chile, Uruguay, etc) or African component (Brazil, Colombia). The most numerous indigenous groups are the Amazonian Quichua or Quechua (> 122,000 people), the Andean Quichua (> 605,000), and the Shuar from the Amazonian region (> 79,700). Most Europeans arrived to Ecuador in the 16th century with the Spanish conquest, giving rise to the initial "mestizo" population.

In contrast to other countries of South America [1–5], Ecuador is underrepresented in the scientific literature on genetic studies. There are only a few articles focused on the analysis of specific Native

* Corresponding author.
E-mail address: antonio.salas@usc.es (A. Salas).
[1] Equal contribution.
[2] www.genvip.org.

https://doi.org/10.1016/j.fsigen.2017.11.011
Received 28 June 2017; Received in revised form 16 November 2017; Accepted 18 November 2017
Available online 22 November 2017
1872-4973/

American groups in a more anthropological context [6–9], or sporadically in biomedical [10] or forensic genetic studies. Baeza et al. [11] analyzed 15 Y-chromosome STRs in a population sample (n = 120) from the capital city, Quito, and the publication reported indices of forensic interest. González-Andrade et al. [12] analyzed a sample of 102 Mestizo, Kichwa and African American individuals from Ecuador; however, only allele frequencies of the 12 Y-STR haplotypes are reported in the publication, thus limiting the usefulness of this dataset in forensic and anthropological studies. Gaviria et al. [13] reported 11 Y-STR haplotypes in a sample of Ecuadorians. Finally, Sánchez et al. [14] analyzed 11 Y-STR haplotypes in confirmed father-son pairs in order to investigate mutation rates. A few studies were also carried out on Native Americans (including Ecuadorian indigenous people); these studies however focused on particular Native American lineages and mainly aimed at revealing the phylogenetic and phylogeographic features of the most common Native haplogroup Q [15,16]. In addition, the study by Roewer et al. [17] focused on Y-chromosome variation of South Native American populations, including a sample from Waorani Ecuadorians (Pastaza province), and reported distinctive haplogroup features (related to haplogroup C3; now renamed as C2 [defined by SNP M217]) in this population.

In the aforementioned articles, haplotypes were not always reported and/or the set of markers analyzed was very limited (below the minimal Y haplotype dataset: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 & DYS385) considering the more recent standards in the field [5]. Currently, the reference Y-STR database in forensic genetics, namely, the Y-haplotype Reference Database (YHRD; https://yhrd.org) contains a few hundred haplotypes (e.g. 750 minimal haplotypes; June 6, 2017) representing a variety of population groups in Ecuador, but e.g. no PPY23 profiles.

Here we provide the most comprehensive dataset of Y-STRs for forensic and anthropological use, contributing a total of 415 Y-chromosome profiles (270 PPY23) and covering the most representative geographic areas of the country. In addition, we also aimed to disentangle the level of population stratification existing in Ecuadorians.

2. Material and methods

2.1. Samples

A total of 414 unrelated samples were randomly recruited by the laboratories of Cruz Roja Ecuatoriana (Quito; Ecuador) and Laboratorio Biomolecular (Cuenca; Ecuador) from 86 locations in mainland Ecuador plus one single haplotype sampled from Galápagos island (Table S1). The mainland locations represent the 23 continental provinces from Ecuador, which fall into the three main ecological regions of the country distributed along a North-South axis (Fig. 1): Pacific coast (n = 156), Andes (n = 242), and Amazon rainforest (n = 16). Note that Galápagos, located in the country's insular region, would represent the province 24 of Ecuador.

Of the total recruited samples, 145 were genotyped for the Yfiler panel (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385ab, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATAH4), while 270 were genotyped for the PPY23 panel (which includes six Y-STRs in addition to those in the Yfiler, namely DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643).

All the haplotypes were deposited in the YHRD database (https://yhrd.org) under accession number YA004261 and are reported in Table S2. Samples were genotyped following manufacturer's recommendations.

Written informed consent was obtained from all the donors.

2.2. Statistical analyses

Forensic parameters were computed for the Ecuador sample considered as a whole and by regions following a division that represents the main ecological areas of the country, namely, Pacific coast, Andes, and Amazon rainforest.

Population comparisons were carried out using reference populations obtained from the dataset in Purps et al. [5]. The selected populations represent the three main ancestral groups assumed to be present in Ecuadorian Y-chromosomes: sub-Saharan Africans (n = 394; represented by Zimbabwe [n = 55]; Kenya [n = 144], and South Africa [n = 114]); Native Americans (n = 200; Bolivia [n = 56], Peru [n = 83], and Brazil [n = 61]); and Europeans (n = 2168; represented by Spain [n = 706], and Italy [n = 1462]). Central Asians (n = 823; represented by China), were used for some analyses. For all the datasets considered, DYS389II alleles were recoded as the difference between the number of repeats at DYS389II and the number of repeats at DYS389I; the recoded marker was used for calculations instead of the original codification, as done previously [5]. Haplotypes bearing microvariants and/or duplications (11 and 3 respectively in our sample), as well as null alleles, were excluded from AMOVA and *Rst* distances estimations; DYS385ab locus was not considered for these analyses. Haplotype diversity (*HD*) was calculated as an analogous of the gene diversity [18]. Discrimination capacity (*DC*) was estimated as the ratio between the number of unique haplotypes and the total number of haplotypes.

All the computations were undertaken by partitioning the full data set into three main marker sets (and excluding the single haplotypes sampled in Galápagos): PPY23 haplotypes (n = 269), PPY23 haplotypes collapsed to Yfiler STRs (hereafter cPPY23), and Yfiler haplotypes (n = 414 [269 + 145]).

It is possible to infer the most likely ancestral origin of Y-STR profiles using population sets that represent the main continental ancestries (the reference populations described above: sub-Saharan Africans, Europeans, and Native Americans). This statistical inference can be obtained using the algorithm described in Egeland et al. [19]; a procedure previously employed in other population contexts and using Y-chromosome haplotypes [1]. Briefly, the method employs a combined PCA-QDA approach: first, a principal component analysis (PCA) reduces the dimension of the data; next, a quadratic discrimination analysis (QDA) performs the classification. We considered the number of principal components (PC) accounting for > 80% of the variation (37 for PPY23 and 24 for Yfiler).

Population differences among regions in Ecuador were inferred using analysis of molecular variance (AMOVA). Genetic distances between pairs of populations in Ecuador and with the reference populations were quantified by R_{ST}. AMOVA and genetic distances were computed using the Arlequin v.3.5.1.3 software [20].

In order to facilitate interpretation of inter-population genetic distances, we carried out a Kruskal's non-metric Multidimensional Scaling (MDS) analysis based on the R_{ST} distances and using the *isomds* function as implemented in MASS package of the statistical software R (www.r-project.org). The MDS plot was build using the Ecuadorian samples and the reference ancestral populations described above.

The spatial representation in a geographical context of Y-chromosome ancestral probabilities were carried out using SAGA v. 4.0.1 (http://www.saga-gis.org/) and the Kriging method.

3. Results

3.1. Forensic parameters

Various forensic parameters were computed for all the Ecuadorian samples, considering the sample as a whole and various regional divisions. All parameters were also calculated for different marker sets (Table 1).

Europe is the reference continental population that shows the highest levels of *HD* and *DC*, and Ecuador shows comparably high levels of these parameters. Although Ecuador has an important Native