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

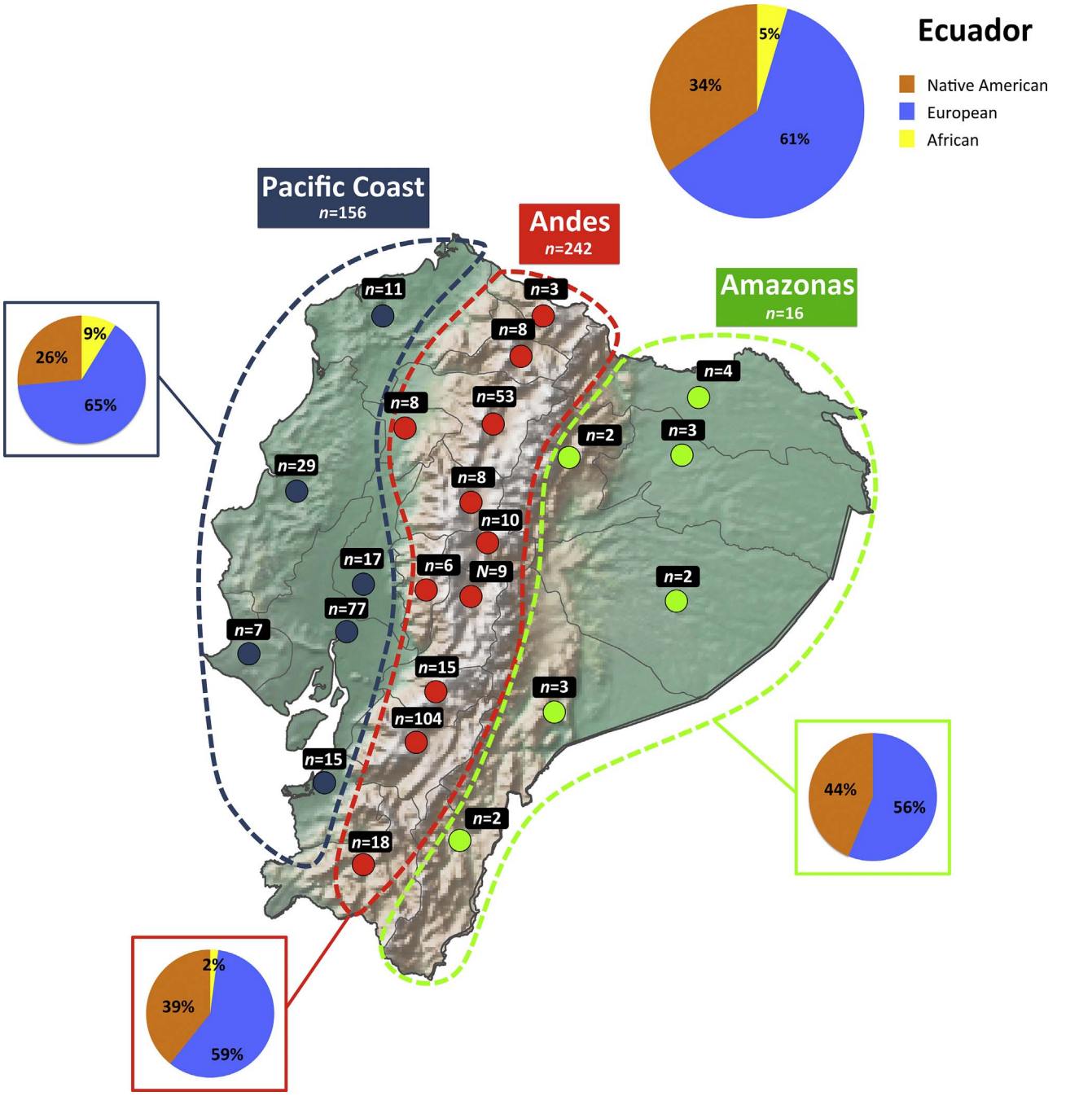


**Fig. 1.** Map of Ecuador showing the sampled regions grouped by provinces (**Table S1**). The pie charts show the proportions of inferred continental ancestry of Y-chromosome profiles. The map on the background has been taken from the public domain http://www.naturalearthdata.com (made with Natural Earth).

American component (according to the census), it shows diversity values that are higher than the average in other Native American groups (Table 1) [3].

As expected, the PPY23 panel shows higher values of *HD* and *DC* in all population partitions when compared to the cPPY23 panel, a trend that is also visible when exploring the full Yfiler sample set (Table 1).

When considering a longitudinal division of the country, the Pacific Coast is the region that shows the lowest values of diversity; and this pattern is consistent for the PPY23, cPPY23, and the Yfiler sets.

### 3.2. Inferred ancestry of Y-chromosome haplotypes

Ancestral components of Y-STR haplotypes were estimated by computing their probability of belonging to different ancestral reference populations using the approach described by Egeland et al. [19]; Table 2. According to this method, the distribution of probabilities indicates that most of the Ecuadorian haplotypes (90.6% of the PPY23 profiles and 84.7% of the Yfiler profiles; Fig. 2) could be assigned with high probability ( > 90%) to one of the three ancestral populations (Fig. 1). Of those haplotypes with probabilities of ancestry adscription above 90%, the partition of ancestry is as follows: 61% European, 34% Native American, and 5% sub-Saharan African. Fig. S1 provides ancestry probabilities split by ecological regions. The spatial distribution of ancestry across the country is heterogeneous. As shown in Fig. 3, European ancestry is more concentrated in the coast, around Santiago de Guayaquil, while the Native American ancestry is more present in the Andes and the Amazon regions, and the African component concentrates in the Northwest of the country, more specifically

**Table 1**
Diversity and forensic parameters for full PPY23, cPPY23 and Yfiler datasets.

| | *n* | *k* | *S* | *HD* | *M* | *DC* | *MP* |
|---|---|---|---|---|---|---|---|
| Population – PPY23 | | | | | | | |
| Ecuador – all | 269 | 264 | 258 | 0.9998 ± 0.00 | 14.77 ± 6.6 | 0.9556 | 258 |
| Ecuador – Andes | 112 | 112 | 112 | 1.0000 ± 0.00 | 14.54 ± 6.5 | 1.0000 | 112 |
| Ecuador – Pacific Coast | 141 | 137 | 133 | 0.9996 ± 0.00 | 15.03 ± 6.7 | 0.9433 | 133 |
| Ecuador – Amazon rainforest | 16 | 16 | 16 | 1.0000 ± 0.02 | 14.16 ± 6.7 | 1.0000 | 16 |
| Ecuador – Guayas province | 77 | 75 | 73 | 0.9993 ± 0.00 | 15.13 ± 6.8 | 0.9481 | 73 |
| Ecuador – Los Ríos province | 17 | 17 | 17 | 1.0000 ± 0.02 | 14.20 ± 6.7 | 1.0000 | 17 |
| Ecuador – Manabí province | 29 | 28 | 26 | 0.9974 ± 0.01 | 14.87 ± 6.8 | 0.8966 | 26 |
| Ecuador – Pichincha province | 53 | 53 | 53 | 1.0000 ± 0.00 | 14.42 ± 6.5 | 1.0000 | 53 |
| Europe (Purps et al. [5]) | *2168* | *2133* | *2098* | *1.0000 ± 0.00* | *14.51 ± 6.5* | *0.9677* | *2100* |
| Asia (Purps et al. [5]) | *823* | *787* | *759* | *0.9999 ± 0.00* | *14.81 ± 6.6* | *0.9222* | *740* |
| Africa (Purps et al. [5]) | *394* | *327* | *288* | *0.9984 ± 0.00* | *13.33 ± 6.0* | *0.7310* | *239* |
| Native America (Purps et al. [5]) | *200* | *189* | *180* | *0.9993 ± 0.00* | *14.06 ± 6.3* | *0.9000* | *175* |
| Population – cPPY23 | | | | | | | |
| Ecuador – all | 269 | 256 | 244 | 0.9996 ± 0.00 | 10.54 ± 4.8 | 0.9037 | 241 |
| Ecuador – Andes | 112 | 111 | 110 | 0.9998 ± 0.00 | 10.31 ± 4.7 | 0.9821 | 110 |
| Ecuador – Pacific Coast | 141 | 133 | 126 | 0.9991 ± 0.00 | 10.77 ± 4.9 | 0.8936 | 125 |
| Ecuador – Amazon rainforest | 16 | 16 | 16 | 1.0000 ± 0.02 | 9.991 ± 4.8 | 1.0000 | 16 |
| Ecuador – Guayas province | 77 | 75 | 73 | 0.9993 ± 0.00 | 10.82 ± 4.9 | 0.9481 | 73 |
| Ecuador – Los Ríos province | 17 | 17 | 17 | 1.0000 ± 0.02 | 10.48 ± 5.0 | 1.0000 | 17 |
| Ecuador – Manabí province | 29 | 27 | 25 | 0.9951 ± 0.01 | 10.62 ± 4.9 | 0.8621 | 25 |
| Ecuador – Pichincha province | 53 | 52 | 51 | 0.9993 ± 0.00 | 10.36 ± 4.8 | 0.9623 | 51 |
| Europe (Purps et al. [5]) | *2168* | *2034* | *1932* | *0.9999 ± 0.00* | *10.36 ± 4.7* | *0.8911* | *1849* |
| Asia (Purps et al. [5]) | *823* | *755* | *705* | *0.9997 ± 0.00* | *10.59 ± 4.8* | *0.8566* | *672* |
| Africa (Purps et al. [5]) | *394* | *305* | *260* | *0.9973 ± 0.00* | *9.045 ± 4.1* | *0.6599* | *189* |
| Native America (Purps et al. [5]) | *200* | *180* | *170* | *0.9982 ± 0.00* | *10.00 ± 4.5* | *0.8500* | *148* |
| Population – Yfiler | | | | | | | |
| Ecuador – all | 414 | 400 | 387 | 0.9998 ± 0.00 | 10.55 ± 4.8 | 0.9325 | 383 |
| Ecuador – Andes | 242 | 241 | 240 | 1.0000 ± 0.00 | 10.45 ± 4.7 | 0.9917 | 240 |
| Ecuador – Pacific Coast | 156 | 147 | 135 | 0.9991 ± 0.00 | 10.71 ± 4.9 | 0.8654 | 134 |
| Ecuador – Amazon rainforest | 16 | 16 | 16 | 1.0000 ± 0.02 | 9.991 ± 4.8 | 1.0000 | 16 |
| Ecuador – Guayas province | 77 | 75 | 73 | 0.9993 ± 0.00 | 10.82 ± 4.9 | 0.9481 | 73 |
| Ecuador – Los Ríos province | 17 | 17 | 17 | 1.0000 ± 0.02 | 10.48 ± 5.0 | 1.0000 | 17 |
| Ecuador – Manabí province | 29 | 27 | 25 | 0.9951 ± 0.01 | 10.62 ± 4.9 | 0.8621 | 25 |
| Ecuador – Pichincha province | 53 | 52 | 51 | 0.9993 ± 0.00 | 10.36 ± 4.8 | 0.9623 | 51 |
| Europe (Purps et al. [5]) | 2168 | 2034 | 1932 | 0.9999 ± 0.00 | 10.36 ± 4.7 | 0.8911 | 1849 |
| Asia (Purps et al. [5]) | 823 | 755 | 705 | 0.9997 ± 0.00 | 10.59 ± 4.8 | 0.8566 | 672 |
| Africa (Purps et al. [5]) | 394 | 305 | 260 | 0.9973 ± 0.00 | 9.045 ± 4.1 | 0.6599 | 189 |
| Native America (Purps et al. [5]) | 200 | 180 | 170 | 0.9982 ± 0.00 | 10.00 ± 4.5 | 0.8500 | 148 |

*n*: Sample size; *k*: Number of different haplotypes; *S*: Singletons; *HD*: haplotype diversity; *M*: average number of pairwise differences; *DC*: Discrimination capacity (*S*/*n*); *MP*: matching probability as in [18].

**Table 2**
Percentages of ancestry estimates from Y-STR profiles.

| | European | Native American | African |
|---|---|---|---|
| PPY23 haplotypes | | | |
| Ecuador – all | 58.5 | 36.5 | 5.0 |
| Andes | 53.1 | 43.8 | 3.1 |
| Pacific Coast | 62.6 | 30.5 | 6.9 |
| Amazon rainforest | 57.1 | 42.9 | 0.0 |
| Yfiler haplotypes | | | |
| Ecuador – all | 60.9 | 34.5 | 4.6 |
| Andes | 58.7 | 39.3 | 2.0 |
| Pacific Coast | 64.7 | 26.5 | 8.8 |
| Amazon rainforest | 56.2 | 43.8 | 0.0 |

in the Esmeraldas province (where it makes up 44% of the ancestry).

### 3.3. Multidimensional scaling analysis

An initial MDS analysis was carried out on $R_{ST}$ distances computed considering Ecuador as a single sample (Fig. 4A; Stress value = 3%). Dimension 1 clearly separates sub-Saharan African samples from the rest, although with an important level of dispersion along this dimension. Native American and Asian ones are close in this dimension, whereas Ecuadorians fall in between Native Americans and Europeans. Dimension 2 separates Asian samples from the European and Native American ones in two main poles. While the Asian population samples form a tight cluster, the European and Native American samples are more dispersed along this axis. The proximity of the Native American samples to the European ones mirrors the fact that almost all Native American groups have an important amount of European ancestry (particularly in their Y-chromosomes). Dimension 3 (Fig. S**2**) displays Ecuadorian sample more closely related to Europeans than to Native Americans, in agreement with the observation that European ancestry predominates on Ecuadorian Y-chromosomes (Table 2). In this Dimension 3, the Peruvian sample still remains closely related to Europeans, while Bolivians and Brazilians separates more clearly towards the opposite pole.

When considering the main Ecuadorian regions (Fig. 4B; Stress value = 4.2%), the MDS shows a similar pattern. The three regions appear closely related to the European and the Native American ones. In good agreement with Y-chromosome profile ancestry, it is the Pacific Coast that shows the closest proximity to the two European samples and, to a much lesser extent, to the African samples, in both Dimensions 1 and 2. Consistently, the Ecuadorian samples from Amazon and Andes fall closer to the three reference Native American samples in Dimension 1. In Dimension 2, it is the sample from Peru that is closely related to the three Ecuadorian regions, a fact that fits well with the geographic proximity of both countries and the genetic homogeneity of the Andes region [21].

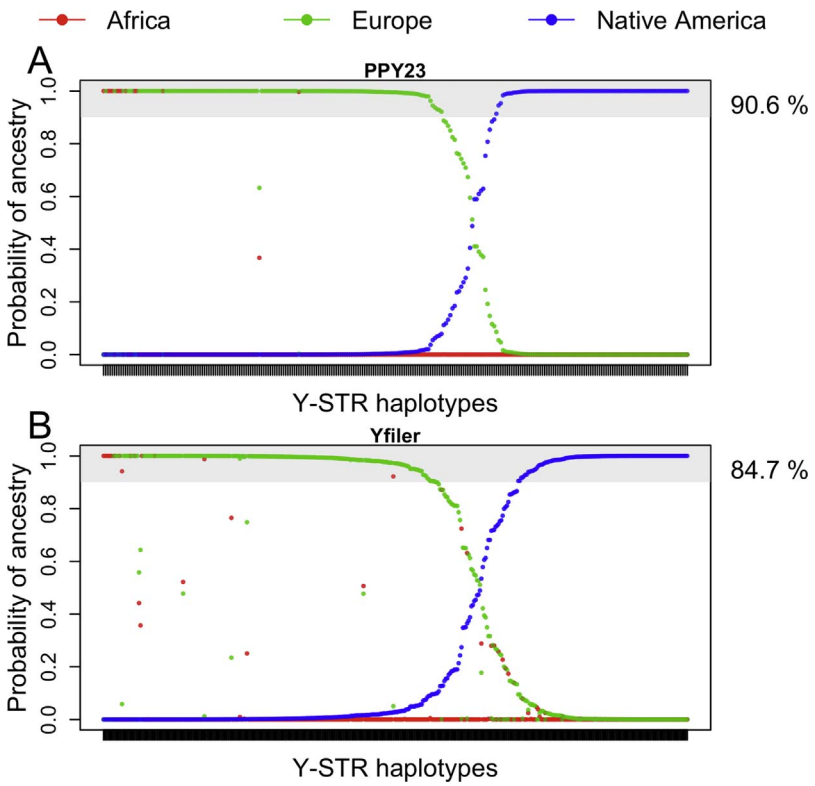


**Fig. 2.** Classification of PPY23 (A) and Yfiler (B) into three main ancestral groups. Every single profile receives three probability values that sum-up to 1, namely, probability of African (red), European (green), and Native American (blue) origin. A total of ~90% PPY23 and 84% of the haplotypes received probabilities above 90% (grey shading zone indicated at the top of the two graphs). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3.4. AMOVA

Analysis of population sub-structure was carried out using AMOVA and considering different geographic partitions and marker sets (Table 3). The data show that variation in Ecuador is moderately stratified, although, as typical of human populations, most of the variation occurs within populations. Thus, for the PPY23 panel, $R_{ST}$ is ~2.2% when considering the whole sample and when considering a longitudinal division of the country. Moreover, these values are statistically significant (Fisher exact test, 10,000 permutations). By dividing the samples longitudinally in the three main ecological regions, it is possible to observe that variation among populations within groups account for most of the among population variation. When considering the Yfiler panel, a drop of about 50% is seen in the variation among groups, although for this dataset are not statistically significant (Table 3).

## 4. Discussion

The present study represents the most comprehensive study carried out on Y-chromosome STRs in Ecuador to date. An important effort was devoted to collect samples representing the three main continental ecological regions of the country. Forensic parameters indicate that both PPY23 and Yfiler profiles have a very high diversity, and, as expected, PPY23 profiles from Ecuador have more discriminating power than Yfiler haplotypes. Population stratification of the Y-chromosome variation is not well visible when using statistical summary indices (low $F_{ST}$ or $R_{ST}$ values, high diversity indices). However, regional genetic

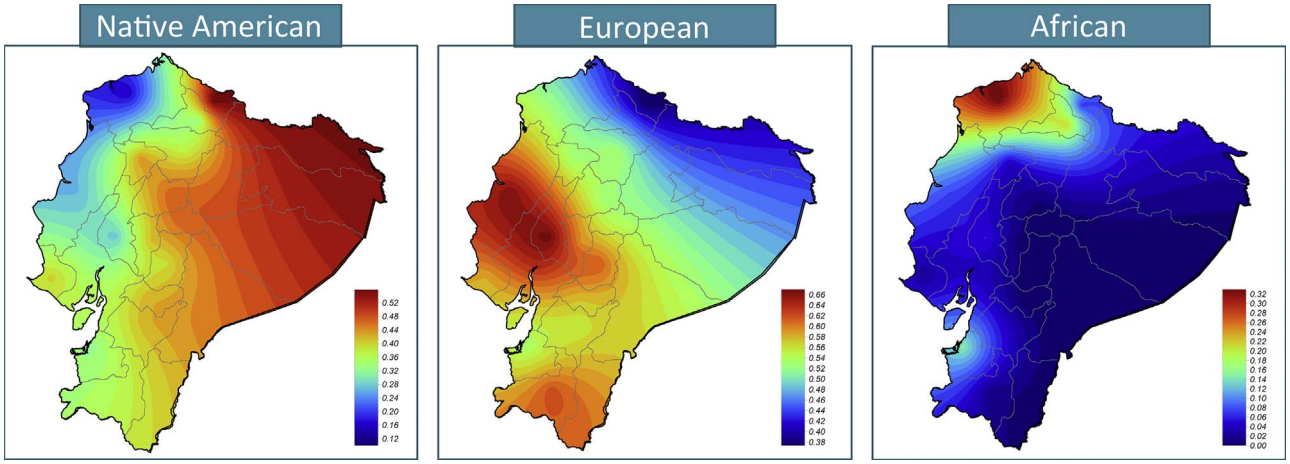


**Fig. 3.** Map of interpolated probabilities of inferred ancestry of Y-chromosome profiles. Interpolation is based on average probabilities per province. In order not to distort averages, those provinces with sample sizes below five individuals were disregarded from these representations. Yfiler profiles (instead of PPY23 data) were used in order to take advantage of the larger sample size and geographic coverage. Only those profiles receiving a probability of assignment into a particular ancestry higher than 90% were considered. The map on the background has been taken from the public domain http://www.naturalearthdata.com.

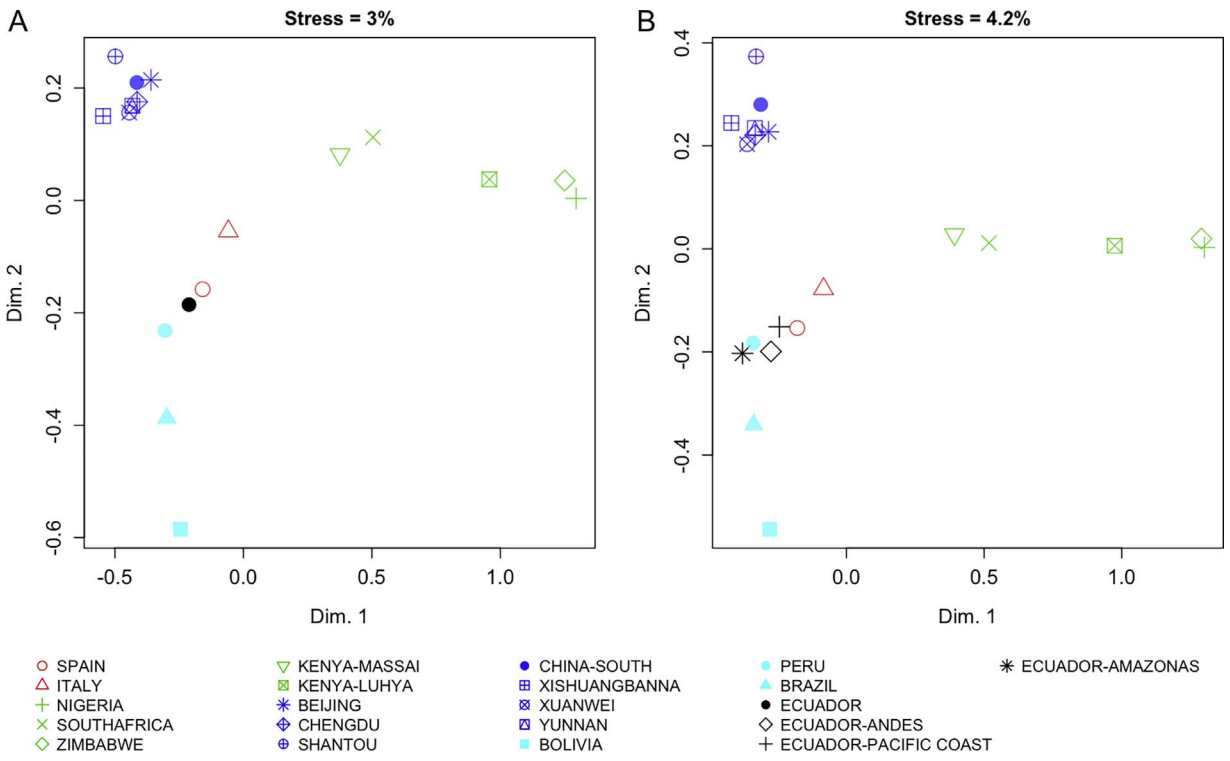


**Fig. 4.** MDS based on $R_{ST}$ distances considering Ecuador as a single sample (A), and considering the three main ecological regions in Ecuador separately (B).

differences clearly manifest when investigating geographic patterns of ancestry inferred from Y-chromosome haplotypes. This level of population sub-structure might be particularly relevant when inferring the most likely origin of a Y-chromosome haplotype.

Although using Y-chromosomal DNA markers to infer biogeographical ancestry have many limitations due to e.g. the fact that the whole Y-chromosome represents actually a single block (locus) of ancestral information, or the existence of gender-biased genetic admixture (which is particularly important in American countries), this piece of information could be relevant in particular forensic applications (e.g. police investigation) [19,22].

The data indicate that variation in Ecuador is stratified along the three main ecological regions. The sub-Saharan African component is more prevalent in the Northwest of the country, specifically in the provinces of Esmeraldas and Imbabura. This is not surprising given that this province is home to the Afro-Ecuadorian culture as a consequence of the African slave trade [23,24]; the Ecuadorian census of 2010 indicates that 43.9% of the population from this province is Afro-Ecuadorian, while the second most numerous group are the “mestizo”. In contrast, the Native American component is more prevalent in the Andean region, also in good agreement with the census; whereas European haplotypes are more prevalently established in Santiago de Guayaquil, a signal that is also visible in the interpolated maps of European Y-haplotype ancestry.

Among the limitations of the present study is the sample size of the Amazon region, although this fact seems unlikely to affect the conclusions of the present study. It is worth mentioning that sampling logistic in the Amazon is particularly complex (less densely populated region, geographical barriers, specific consents for sampling ethnic Native American groups, etc). Another point of interest for future studies would be to investigate new statistical algorithms to infer ancestry from Y-STRs. In this sense, it would be worth to know if poorly resolved ancestry estimates could be related to the algorithm employed, or rather to the noise generated by the presence of interrogated haplotypes in the reference populations used for classification (e.g. existence of European Y-haplotypes in the Native American sample sets used for classification).

Summarizing, to the best of our knowledge, this is the first time that ancestry has been inferred from Y-STRs haplotypes at a population level. The results fit very well with historical and demographic

**Table 3**
AMOVA results for different grouping schemes. Analyses were carried out by considering all the Ecuadorian provinces in a single group (“Ecuador – all”), and considering the longitudinal division of Ecuador in the three main groups (Amazon rainforest, Andes, and Pacific coast: “Ecuador – longitudinal”).

| Markers | Population set | Pops/Gr/*n* | AP/AG | APWG | WP | $R_{ST}$ | *P*-value |
|---|---|---|---|---|---|---|---|
| PPY23 | Ecuador – all | 22/1/270 | 2.20 | – | 97.80 | 0.02204 | 0.033 ± 0.001 |
| | Ecuador – longitudinal | 22/3/270 | 0.00 | 2.21 | 97.79 | 0.02214 | 0.017 ± 0.001 |
| cPPY23 | Ecuador – all | 22/1/270 | 1.02 | – | 98.98 | 0.01015 | 0.175 ± 0.003 |
| | Ecuador – longitudinal | 22/3/270 | 0.06 | 1.16 | 98.77 | 0.01227 | 0.130 ± 0.003 |
| Yfiler | Ecuador – all | 24/1/415 | 0.99 | – | 99.01 | 0.00992 | 0.084 ± 0.003 |
| | Ecuador – longitudinal | 24/3/415 | 0.25 | 0.96 | 98.79 | 0.01206 | 0.056 ± 0.002 |

Pops/Gr/*n*: number of populations/groups/individuals included in the AMOVA; AP: variation among populations; AG: variation among groups; APWG: variation among populations within groups; WP: variation within populations; Ecuador – longitudinal division: Andes (Azuay, Bolívar, Cañar, Carchi, Chimborazo, Cotopaxi, Imbabura, Loja, Pichincha, Santo Domingo, Tungurahua), Pacific coast (El Oro, Esmeraldas, Guayas, Los Ríos, Manabí, Santa Elena), Amazon rainforest (Morona Santiago, Napo, Orellana, Pastaza, Sucumbios, Zamora Chinchipe).

documentation (census) of Ecuador. This method paves the way for future studies where classical procedures (mainly based on summary statistical indices of genetic variation [25]) fail to detect population stratification. Moreover, the dataset provided in the present study contributes to fill a gap in the reference forensic database YHRD, by contributing Y-chromosome profiles that represent different areas of Ecuador and providing high resolution PPY23 haplotypes.

## Acknowledgments

The present project received support from Laboratorio Biomolecular (Ecuador) and Laboratorio de Genética Molecular (Cruz Roja Ecuatoriana, Ecuador). A.S. received support from the Instituto de Salud Carlos III (Proyecto de Investigación en Salud, Acción Estratégica en Salud): project GePEM ISCIII/PI16/01478/Cofinanciado FEDER. F.M.T. received support from project ReSVinextISCIII/PI16/01569/Cofinanciado FEDER; Consellería de Sanidade, Xunta de Galicia (RHI07/2-intensificación actividad investigadora, PS09749 and 10PXIB918184PR), Instituto de Salud Carlos III (Intensificación de la actividad investigadora 2007–2012, PI16/01569), Fondo de Investigación Sanitaria (FIS; PI070069/PI1000540) del plan nacional de I + D + I and 'fondos FEDER'. A.S. and F.M.T. received support from 2016-PG071 Consolidación e Estructuración REDES2016GI-1344 G3VIP (Grupo Gallego de Genética Vacunas Infecciones y Pediatría, ED341D R2016/021)(A.S. and F.M.T). We extend our gratitude to the donors who participated in this study.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2017.11.011.

## References

[1] U. Toscanini, F. Brisighelli, C. Llull, G. Berardi, A. Gómez, F. Andreatta, J. Pardo-Seco, A. Gómez-Carballa, F. Martinón-Torres, V. Álvarez-Iglesias, et al., Charting the Y-chromosome ancestry of present-day Argentinean mennonites, J. Hum. Genet. (2016).

[2] U. Toscanini, F. Brisighelli, F. Moreno, J.A. Pantoja-Astudillo, E.A. Morales, P. Bustos, J. Pardo-Seco, A. Salas, Analysis of Y-chromosome STRs in Chile confirms an extensive introgression of European male lineages in urban populations, Forensic Sci Int Genet 21 (2016) 76–80.

[3] U. Toscanini, L. Gusmão, G. Berardi, A. Amorim, Á. Carracedo, A. Salas, E. Raimondi, Y chromosome microsatellite genetic variation in two native American populations from Argentina: population stratification and mutation data, Forensic Sci. Int. Genet. 2 (2008) 274–280.

[4] J.M. Cárdenas, T. Heinz, J. Pardo-Seco, V. Álvarez-Iglesias, P. Taboada-Echalar, P. Sánchez-Diz, Á. Carracedo, A. Salas, The multiethnic ancestry of Bolivians as revealed by the analysis of Y-chromosome markers, Forensic Sci Int Genet 14 (2015) 210–218.

[5] J. Purps, S. Siegert, S. Willuweit, M. Nagy, C. Alves, R. Salazar, S.M. Angustia, L.H. Santos, K. Anslinger, B. Bayer, et al., A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, Forensic Sci. Int. Genet. 12C (2014) 12–23.

[6] D. Reich, N. Patterson, D. Campbell, A. Tandon, S. Mazieres, N. Ray, M.V. Parra, W. Rojas, C. Duque, N. Mesa, et al., Reconstructing native American population history, Nature 488 (2012) 370–374.

[7] K. Bryc, C. Velez, T. Karafet, A. Moreno-Estrada, A. Reynolds, A. Auton, M. Hammer, C.D. Bustamante, H. Ostrer, Genome-wide patterns of population structure and admixture among Hispanic/Latino populations, Proc. Natl. Acad. Sci. U. S. A. 107 (Suppl 2) (2010) 8954–8961.

[8] J.R. Homburger, A. Moreno-Estrada, C.R. Gignoux, D. Nelson, E. Sanchez, P. Ortiz-Tello, B.A. Pons-Estel, E. Acevedo-Vasquez, P. Miranda, C.D. Langefeld, et al., Genomic insights into the ancestry and demographic history of South America, PLoS Genet. 11 (2015) e1005602.

[9] S. Cardoso, M.A. Alfonso-Sánchez, L. Valverde, D. Sánchez, M.T. Zarrabeitia, A. Odriozola, B. Martínez-Jarreta, M.M. de Pancorbo, Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon, Heredity (Edinb.) 108 (2012) 609–615.

[10] Y.M.C. Paz, M.J. Guillen Sacoto, P.E. Leone, Genetics and genomic medicine in Ecuador, Mol. Genet. Genomic Med. 4 (2016) 9–17.

[11] C. Baeza, R. Guzmán, M. Tirado, A.M. López-Parra, T. Rodriguez, M.S. Mesa, E. Fernández, E. Arroyo-Pardo, Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador), Forensic Sci. Int. 173 (2007) 214–219.

[12] F. González-Andrade, D. Sánchez, B. Martínez-Jarreta, B. Budowle, Y-chromosome STR haplotypes in three different population groups from Ecuador (South America), J. Forensic Sci. 53 (2008) 512–514.

[13] A. Gaviria, M.E. Sánchez, G. Morejón, M. Vela, V. Aguirre, G. Burgos, Characterization and haplotype of 11 Y-STR loci in Ecuadorian population, Forensic Sci. Int. Genet. Suppl. Series. 4 (2013) e310–e311.

[14] M.E. Sánchez, G. Burgos, A. Gaviria, V. Aguirre, M. Vela, P.E. Leone, C. Paz-y-Miño, Y STRs mutation events in father-son pairs in Ecuadorian individuals, Forensic Sci. Int.: Genet Suppl. Ser. 5 (2015) e310–e311.

[15] M.S. Jota, D.R. Lacerda, J.R. Sandoval, P.P. Vieira, S.S. Santos-Lopes, R. Bisso-Machado, V.R. Paixão-Cortes, S. Revollo, Y.M.C. Paz, R. Fujita, et al., A new sub-haplogroup of native American Y-chromosomes from the Andes, Am. J. Phys. Anthropol. 146 (2011) 553–559.

[16] M. Geppert, M. Baeta, C. Núñez, B. Martínez-Jarreta, S. Zweynert, O.W. Cruz, F. González-Andrade, J. González-Solorzano, M. Nagy, L. Roewer, Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America, Forensic Sci. Int. Genet. 5 (2011) 100–104.

[17] L. Roewer, M. Nothnagel, L. Gusmão, V. Gomes, M. González, D. Corach, A. Sala, E. Alechine, T. Palha, N. Santos, et al., Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, PLoS Genet. 9 (2013) e1003460.

[18] N. Nei, Molecular Evolutionary Genetics1987, Columbia University Press, New York, 2016.

[19] T. Egeland, H.M. Bøvelstad, G.O. Storvik, A. Salas, Inferring the most likely geographical origin of mtDNA sequence profiles, Ann. Hum. Genet. 68 (2004) 461–471.

[20] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Res. 10 (2010) 564–567.

[21] M. Bodner, U.A. Perego, G. Huber, L. Fendt, A.W. Rock, B. Zimmermann, A. Olivieri, A. Gomez-Carballa, H. Lancioni, N. Angerhofer, et al., Rapid coastal spread of first Americans: novel insights from South America's Southern Cone mitochondrial genomes, Genome Res. 22 (2012) 811–820.

[22] A. Salas, H.-J. Bandelt, V. Macaulay, M.B. Richards, Phylogeographic investigations: the role of trees in forensic genetics, Forensic Sci. Int. 168 (2007) 1–13.

[23] A. Salas, M. Richards, M.V. Lareu, R. Scozzari, A. Coppa, A. Torroni, V. Macaulay, Á. Carracedo, The African diaspora: mitochondrial DNA and the Atlantic slave trade, Am. J. Hum. Genet. 74 (2004) 454–465.

[24] A. Salas, M. Richards, M.V. Lareu, B. Sobrino, S. Silva, M. Matamoros, V. Macaulay, Á. Carracedo, Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA, Am. J. Phys. Anthropol. 128 (2005) 855–860.

[25] U. Toscanini, M. García-Magariños, G. Berardi, T. Egeland, E. Raimondi, A. Salas, Evaluating methods to correct for population stratification when estimating paternity indexes, PLoS One 7 (2012) e49832.