Forensic Science International: Genetics 6 (2012) 504–509

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Forensic Population Genetics - Original Research

Updated Brazilian STR allele frequency data using over 100,000 individuals: An analysis of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX and vWA loci

Vitor Rezende da Costa Aguiar [a,1], Eldamária de Vargas Wolfgramm [a,1], Frederico Scott Varella Malta [b,2], Adriana Gonçalves Bosque [b,2], Amanda de Castro Mafia [b,2], Vanessa Cristina de Oliveira Almeida [b,2], Fabiola de Andrade Caxito [b,2], Victor Cavalcanti Pardini [b,2], Alessandro Clayton Souza Ferreira [b], Iúri Drumond Louro [a,*]

[a] Núcleo de Genética Humana e Molecular, Departamento de Ciências Biológicas, Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Av. Marechal Campos 1468, Campus de Maruípe, CEP 29040-090 Vitória, ES, Brazil
[b] Laboratório Hermes Pardini, Departamento de Genética Molecular, Avenida das Nações, 2448 Portaria A, 32200-000 Vespasiano, MG, Brazil

ARTICLE INFO

Article history:
Received 8 February 2011
Received in revised form 17 June 2011
Accepted 5 July 2011

Keywords:
Autosomal STRs
Allelic frequencies
Population data
Brazil

ABSTRACT

The Brazilian population is one of the most heterogeneous populations of the world, formed mainly by an admixture of European, African and Native American populations. Brazil is the fifth largest country in the world (8,511,960 km^2), being divided into five geographical regions. This study provides population genetic data of up to 137,161 unrelated individuals representing the entire Brazilian territory. Allelic frequencies and other population data analysis are reported for the 15 autosomal STR loci included in the PowerPlex® 16 kit (Promega Corporation, Madison, WI, USA). In order to guarantee that individuals were not related, we have considered only F1 data from couples undergoing paternity testing. The number of individuals genotyped for each locus was: CSF1PO (113,526); D3S1358 (135,133); D5S818 (135,181); D7S820 (137,136); D8S1179 (134,211); D13S317 (137,161); D16S539 (136,942); D18S51 (136,739); D21S11 (130,014); FGA (135,839); Penta D (110,333); Penta E (128,055); TH01 (112,695); TPOX (123,102); vWA (127,415). Allele sizes ranged from 1 to 48.2. Statistic parameters (PD, PIC and Ho; considering values ≥0.75) suggest that markers D13S317, D16S539, D18S51, D21S11, D7S820, D8S1179, Penta D, Penta E, TH01, FGA and vWA were more informative for genetic identification purposes in the Brazilian population.

1. Population

Brazil is the fifth largest country in the world and occupies an area of 8.5 million km^2. The country is divided into 26 States and one Federal District, which are grouped into five macroregions (North, North-East, Central-West, South-East and South). Its present population exceeds 190 million Portuguese speaking people, in contrast to all other Latin American nations in which Spanish is the official language, reflecting their colonization pattern. The Brazilian population was originated by successive migration waves, starting with an initial arrival of approximately 500,000 Portuguese colonizers (mostly men) between the years of 1500 and 1800, which met the local Amerindian population, estimated at 2.4 million at the time [1]. The Portuguese-Amerindian admixture started soon after the arrival of the first colonizers. At the same time period, the slave trade brought approximately 4 million Sub-Saharan Africans to Brazil, providing the second major ethnical contribution to the Brazilian population [1]. From then on, other waves of immigration occurred, especially the influx of nearly 6 million officials from different countries, mainly Italy, Portugal, Spain, Germany, Syria, Lebanon and Japan, adding even more complexity to the already multiethnic highly admixed Brazilian population [1].

2. Sampling

STR data were collected of up to 137,161 unrelated individuals of F1 generation of couples undergoing paternity testing in the Hermes Pardini Laboratory, Minas Gerais, Brazil. These individuals were representative of all 5 geographical regions and the Federal District

* Corresponding author. Tel.: +55 27 3335 7251; fax: +55 27 335 7250.
E-mail address: iurilouro@yahoo.com (I.D. Louro).
[1] First and second authors contributed equally to this manuscript.
[2] Tel.: +55 31 3629 4852; fax: +55 31 3629 4872.

1872-4973/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2011.07.005

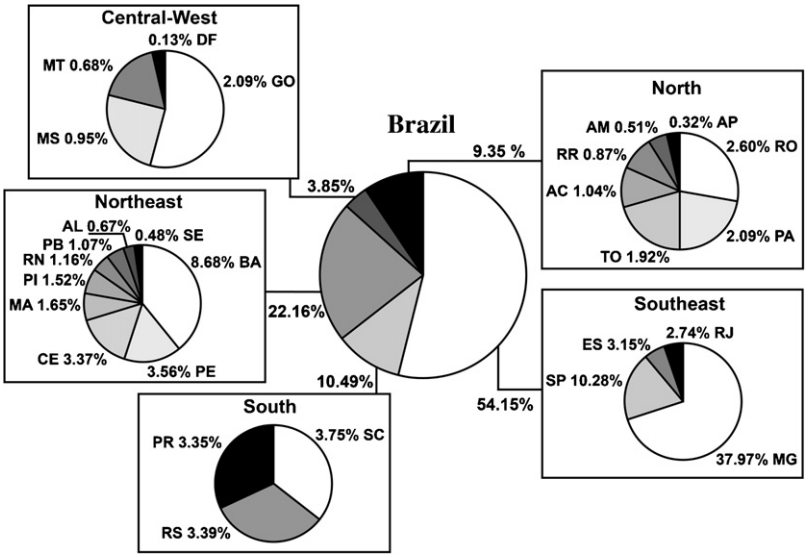

Fig. 1. Dataset Composition. Brazilian macroregions with their States are shown. Percent values represent specific contributions to the complete dataset. Southeast: ES: Espírito Santo; MG: Minas Gerais; RJ: Rio de Janeiro; SP: São Paulo. South: PR: Paraná; RS: Rio Grande do Sul; SC: Santa Catarina. Central-West: DF: Distrito Federal; GO: Goiás; MS: Mato Grosso do Sul; MT: Mato Grosso. North: AC: Acre; AM: Amazonas; AP: Amapá; PA: Pará; RO: Rondônia; RR: Roraima; TO: Tocantins. Northeast: AL: Alagoas; BA: Bahia; CE: Ceará; MA: Maranhão; PB: Paraíba; PE: Pernambuco; PI: Piauí; RN: Rio Grande do Norte; SE: Sergipe.

of Brazil (Fig. 1). In order to guarantee that individuals were not related, we have considered only F1 data from couples undergoing paternity testing. The number of individuals genotyped for each locus was: CSF1PO (113,526); D3S1358 (135,133); D5S818 (135,181); D7S820 (137,136); D8S1179 (134,211); D13S317 (137,161); D16S539 (136,942); D18S51 (136,739); D21S11 (130,014); FGA (135,839); Penta D (110,333); Penta E (128,055); TH01 (112,695); TPOX (123,102); vWA (127,415).

3. STR typing

Peripheral blood was collected on Whatman® FTA Elute paper. After thoroughly drying the paper, a 3 mm punch was washed in 500 μl of sterile water, transfered to a new tube and incubated with 30 μl of sterile water at 95 °C for 30 min, at which point the paper punch was discarded and template DNA was ready for amplification. The PowerPlex®16 System Kit (Promega Corporation, Madison, WI, USA) was used for STR typing according to the manufacturers' recommended protocols. PCR fragments were separated by capillary electrophoresis on a MegaBACE™ 1000 (GE Healthcare, United Kingdom). Allele calling was done by the Fragment Profiler software. All steps were carried out according to laboratory internal control standards and kit controls.

4. Statistical analysis

Allelic frequencies, Hardy–Weinberg equilibrium (HWE), observed heterozygosity (Ho) and expected heterozygosity (He) were calculated by the Arlequin software, version 3.11 [2]. Other statistical parameters such as matching probability (MP), power of discrimination (PD), polymorphic information content (PIC), power of exclusion (PE) and typical paternity index (TPI) were analyzed using the PowerStats software, version 12 (Promega Corporation, Madison, WI, USA) [3]. Results were considered to be significant at $P < 0.05$. Our results were compared to other Brazilian Populations data [4], as well as populations from Sub-Saharan Africa (Angola [5], Guinea-Bissau [6] and Mozambique [7]), Eurasia (Portugal [8–12], Spain [13–16], Italy [17–19] and Germany [20]) and East Asia (Japan [21]) described in the Autosomal STR DNA Database website [22]. These comparisons were computed by means of analysis of molecular variance (AMOVA) and Pairwise analyzed (*Fst*) using the Arlequin software (v3.11) [2].

5. Results and conclusion

The participation of each macroregion and State in the final composition of the dataset is shown in Fig. 1. Allelic frequencies and other statistical parameters of forensic interest for all 15 STR loci of the PowerPlex®16 System are shown in Table 1. All markers showed a high degree of genetic polymorphism, with observed heterozygosity (Ho) values ranging from 0.77223 (D3S1358) to 0.89750 (Penta E), except CSF1PO, D5S818 and TPOX, which presented Ho values ≤ 0.75. PD values were all higher than 0.75 and PIC values were higher than 0.70, except for CSF1PO, D5S818 and TPOX loci. Analysis always resulted in *P*-values < 0.00001.

Because the Brazilian population is the result of centuries of multiethnic crossings among Ameridians, Africans and Europeans, we compared our results with some of those populations. Pairwise analysis (*Fst*) demonstrated that the Brazilian population was more closely related to the Europeans for markers D3S1358, D7S820, D16S539 and TH01 and more similar to Sub-Saharan Africans according to markers FGA, Penta D, Penta E, D18S51, D21S11 and vWA. Comparing our results with other Brazilian population [4] showed a high degree of concordance for markers D18S51, CSF1PO and D5S818. AMOVA analysis showed marker TPOX with the highest index of variation among individuals of different populations (24.35%), as opposed to marker Penta E that showed the lowest index (5.96%). The lowest fixation index (*Fst*) was observed for marker Penta E (*Fst* coefficient = 0.05963). Pairwise and AMOVA analysis results are shown in Table 2.

This work is the largest STR frequency analysis of the Brazilian population up to date, reflecting the entire Brazilian territory and studying over 100,000 unrelated individuals for all loci represented in the Powerplex®16 Promega Kit. Therefore we recommend these data to be used in genetic identification analysis of Brazilian individuals henceforth.This paper follows the ISFH recommendations relating to STR allele designation [23,24] and the guidelines for publication of Forensic Population Genetics requested by the journal [25].

Ethical standards

This work was in accord with the national research ethic regulations and was approved by the Ethics Committee of the