Int J Legal Med (2012) 126:337–341
DOI 10.1007/s00414-011-0606-1

POPULATION DATA

Population genetic analyses of the AmpFlSTR® NGM™ in Brazil

Paulo Eduardo Raimann · Gabriela Hirschfeld-Campolongo · Simone Schumacher · Sabrina Esteves de Mattos Almeida · Cláudia Maria Dornelles da Silva · Rodrigo Rodenbusch · Alessandra Simões Bassini · Leandro Fonseca de Souza · Eugênio César Soares Nascimento · Eleusa Santana de Oliveira · Valéria Rosalina Dias E. Santos · Elza Cristina Lopes de Oliveira · Eloísa Aurora Auler-Bittencourt · Roberta Casemiro da Rocha Hirschfeld-Campolongo · Maria Luiza Almeida Prado Oliveira e Sousa · Helder Pereira de Figueiredo · Bruno Boiko Pereira de Figueiredo · Rosane Silva · Rodrigo Soares Moura-Neto

Received: 19 May 2011 / Accepted: 13 July 2011 / Published online: 18 August 2011

Abstract Population data of 15 short tandem repeat loci of the AmpFlSTR® next generation multiplex (NGM)™ were obtained from a sample of 835 individuals. The loci are the ten short tandem repeats (STRs) in the SGM Plus® Kit plus the EDNAP- and ENSFI-recommended STRs D10S1248, D22S1045, D2S441, D1S1656, and D12S391. Allele frequency and other forensically relevant statistics data were generated for the NGM loci into five current country macroregions of Brazil (North, Northeast, Central West, Southeast, and South). All the analyzed loci meet Hardy–Weinberg equilibrium expectations and no linkage disequilibrium in all pairs of loci. The observed and expected heterozygosity, power of discrimination, polymorphic information content, and the other population–

Eugênio César Soares Nascimento (in memorium).

Electronic supplementary material The online version of this article (doi:10.1007/s00414-011-0606-1) contains supplementary material, which is available to authorized users.

P. E. Raimann · G. Hirschfeld-Campolongo
Life Technologies do Brasil,
Av do Café, 277 1° andar Bloco A,
04311-000 São Paulo, SP, Brazil

S. Schumacher · S. E. de Mattos Almeida · C. M. D. da Silva · R. Rodenbusch
Laboratório de Investigação de Paternidade,
Fundação Estadual de Produção e Pesquisa em Saúde,
Av. Ipiranga, 5400, 90610-000
Porto Alegre, RS, Brazil

S. E. de Mattos Almeida
Instituto da Saúde, Centro Universitário Feevale,
Av. Dr. Maurício Cardoso, 510, 93510-250
Novo Hamburgo, RS, Brazil

P. E. Raimann · R. Rodenbusch
Programa de Pós-graduação em Biologia Celular e Molecular,
Pontifícia Universidade Católica,
Av. Ipiranga, 6681, 90619-900
Porto Alegre, RS, Brazil

C. M. D. da Silva
Curso de Pós-graduação em Diagnóstico Genético Molecular,
Universidade Luterana do Brasil,
Av. Farroupilha, 8001, 92425-900
Canoas, RS, Brazil

A. S. Bassini · L. F. de Souza
Instituto de Medicina Social e de Criminologia de São Paulo,
Rua Barra Funda, 824, 01152-000
São Paulo, SP, Brazil

E. C. S. Nascimento · E. S. de Oliveira
Departamento de Polícia Técnica,
Laboratório Central de Polícia Técnica,
Av. Centenário S/N, 41100-390
Salvador, BA, Brazil

V. R. D. E. Santos · E. C. L. de Oliveira
Laboratório de DNA Forense,
Instituto de Criminalística de Minas Gerais,
Rua Juiz de Fora, 400, 30180-060
Belo Horizonte, MG, Brazil

genetic indices were calculated. The overall power of discrimination was greater than 0.99999999999999999996 and the combined power of exclusion was greater than 0.9999998 in all Brazilian populations. Comparative analysis between populations from different Brazilian macroregions as well as between Brazil and Caucasian, African Americans, and Hispanic US populations are presented.

Keywords Short tandem repeat loci · Next generation multiplex · Allele frequencies · Brazil · Population data

Population

Brazil is a continental country that was colonized by the Portuguese. The current population is about 190 million inhabitants, according to the last National Survey Inquiries in 2010 [1], and is divided into five macroregions: North (N), Northeast (NE), Central West, Southeast, and South (S). The admixture process occurred in different manners among the macroregions of Brazil. In NE, the African component is high and the Native American contribution is low; in the N, the contribution of Native Americans is pronounced, whereas in the S, the Amerindian and African influence is reduced in comparison with all the other macroregions. The Southeast and Central West macroregions can be considered intermediate [1]. Samples were obtained from 835 unrelated individuals, after informed consent was obtained and made anonymous before they were analyzed. The samples were collected reflecting the Brazilian population distribution: North, 88 individuals; Northeast, 155 samples; Central–West, 91 individuals; Southeast, 244 samples; and South, 257 individuals (Fig. 1).

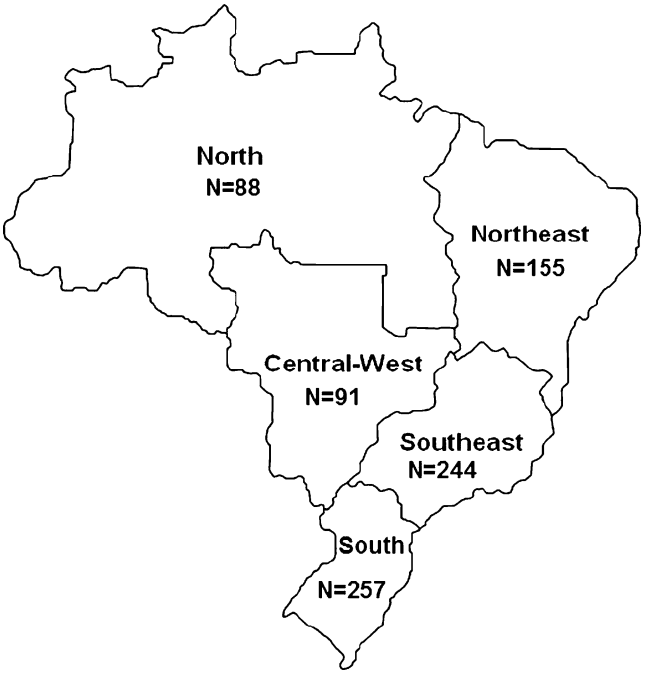

Fig. 1 Map of Brazil showing all five macroregions (North, Northeast, Central West, Southeast, and South) and the amount of individuals analyzed (*N*)

DNA extraction

Genomic DNA (gDNA) was extracted from whole blood samples using QIAamp blood kit (Qiagen) following the manufacturer's instructions. Alternatively, gDNA was extracted using Chelex method [2].

PCR amplification and typing

The extracted DNA was amplified using the AmpFlSTR® next generation multiplex (NGM)™ polymerase chain reaction (PCR) Amplification Kit (Applied Biosystems, Foster City, CA, USA) in the GeneAmp 9700 PCR system (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's recommendations. The analysis of the amplified PCR product was performed on the ABI 3130 or ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, USA). The data analysis and allele identification were performed using GeneMapper® ID (version 3.2) and GeneMapper® ID-X (version 1.2) analysis software.

Quality control

All laboratories participated on GEP/ISFG Collaborative Study and/or Quality Control AICEF/GITAD.

E. A. Auler-Bittencourt ·
R. C. da Rocha Hirschfeld-Campolongo · M. L. A. P. O. e Sousa
Instituto de Criminalistica de São Paulo,
Rua Moncorvo Filho, 410, 05507-060 São Paulo, SP, Brazil

H. P. de Figueiredo · B. B. P. de Figueiredo
Instituto de Perícias Científicas de Mato Grosso do Sul,
Rua da Paz 185, 79002-190 Campo Grande, MS, Brazil

R. Silva
Laboratório de Metabolismo Macromolecular Firmino Torres de Castro, Instituto de Biofísica Carlos Chagas Filho/UFRJ,
21941-092 Rio de Janeiro, RJ, Brazil

R. S. Moura-Neto (✉)
Laboratório de Biologia Molecular Forense,
Instituto de Biologia/UFRJ,
21941-092 Rio de Janeiro, RJ, Brazil
e-mail: rodrigomouraneto@ufrj.br

Table 1 Matrix of Reynolds' genetic distances based on the allelic frequencies of five population Brazilian groups and US African Americans, Caucasian, and Hispanic

Brazil	–								
Brazil—CW	0.0034	–							
Brazil—NE	0.0027	0.0083	–						
Brazil—N	0.0063	0.0103	0.0111	–					
Brazil—SE	0.0009	0.0054	0.0036	0.0085	–				
Brazil—S	0.0011	0.0042	0.0057	0.0068	0.0031	–			
African American	0.0138	0.0217	0.0089	0.0238	0.0123	0.0194	–		
Caucasian	0.0051	0.0090	0.0090	0.0124	0.0058	0.0047	0.0240	–	
Hispanic	0.0025	0.0056	0.0069	0.0080	0.0037	0.0029	0.0187	0.0074	–

CW Central west, *NE* northeast, *N* north, *SE* southeast, *S* south

Analysis of data

Calculations of allele frequencies, power of discrimination, and power of exclusion were carried out using PowerStats version 1.2 [3]. Observed and expected heterozygosity and *p* values of the Hardy–Weinberg equilibrium tests and coancestry coefficient (Fst) values were assessed using GDA Package version 1.0 [4]. Genetic distances (Fst, Table 1) between the Brazilian data and US Caucasian, Hispanic, and African American [5] population samples were calculated from allelic frequencies using PowerMarker software [6]. TreeView software was used for graphic representation [7]. Multidimensional analysis was done using SPSS V.15 [8].

Results

Allele frequency distribution, statistical, and forensic parameters, regarding the 15 NGM short tandem repeat (STR) loci, were determined among 835 Brazilians individuals and the results are shown in Supplemental Table S1. All loci were highly polymorphic in all five populations. The observed heterozygosity varies between 0.716 for D22S1045 and 0.905 for D18S51. The power of discrimination was lowest for D22S1045, 0.890, and highest for D1S1656, 0.978. The power of exclusion ranges from 0.454 for D22S1045 to 0.805 for D18S51. The overall PD was greater than 0.99999999999999999996 and the overall PE was greater than 0.9999998 in all Brazilian populations. Coancestry (Fst) values for the 15 NGM STR loci were estimated using the five Brazilian population groups reported here (Table 2). The overall Fst of 0.0018 is well below the both values of 0.01 and 0.03 recommended by the National Research Council [9]. No evidence of deviations from Hardy–Weinberg equilibrium was found, when using a Bonferroni correction [10] ($p<0.05/15=0.0033$), or the usual threshold of statistical significance at 0.05. There were no departures detected for the two syntenic loci vWA and D12S319. The same unlinked observation was on the work of Phillips et al. [11] and Budowle et al. [5] and supports that for forensic identity testing the "product rule" approach is justified for the loci vWA and D12S391, although the use of Fst assumes that departures from independence may occur. As pointed out by Budowle et al. [5], the independence between these syntenic loci, at the population level-based statistics, is likely due to the result of the relatively high mutation rates of STR loci. However, for family analyses, vWA and D12S391 may be linked; these loci are located on chromosome 12 at positions 6,093 and 12,340 kb, respectively [5, 12]. Although the genetic distance is about 12 cm between these two syntenic loci, physical map distances may not necessarily correlate to the recombination rate due to the presence of a recombinational hotspot. Three-generation family analyses will be necessary to assess the

Table 2 Fst values based on five Brazilian population groups

Locus	Fst
D10S1248	0.00152
vWA	0.00081
D16S539	0.00317
D2S1338	0.00112
D8S1179	−0.00001
D21S11	−0.00018
D18S51	0.00074
D22S1045	0.00630
D19S433	0.00034
TH01	0.00538
FGA	0.00108
D2S441	0.00392
D3S1358	0.00274
D1S1656	0.00058
D12S391	0.00063
Avg	0.00180

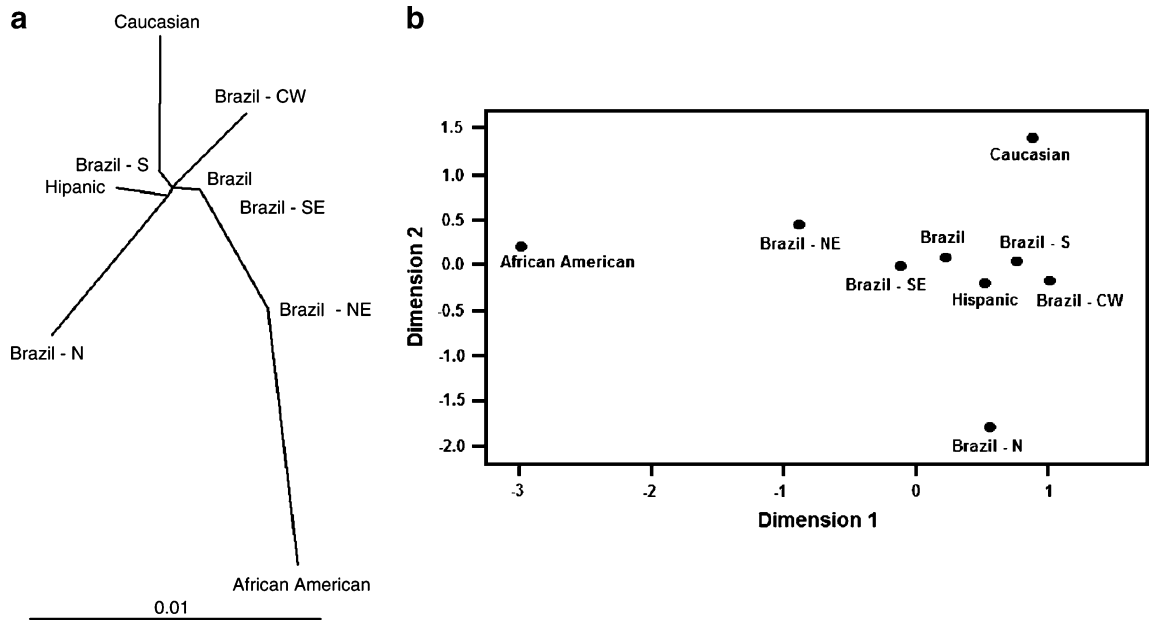

Fig. 2 **a** Neighbor-joining tree of five Brazilian population groups and Brazil all together and US African Americans, Caucasian, and Hipanic groups using Reynolds as a measure of genetic distance. **b** Multidimensional analysis using Reynolds' matrix among populations (coefficient of stress, 0.09462)

degree of linkage effect for these two loci (as reported by Budowle et al. [5]). On the other hand, O'Connors et al. [12] had stated that, for any forensic analyses, the genotype probabilities for D12S391 and vWA should not be multiplied to determine the match probability of an autosomal STR profile. They recommended that the observed diplotype combination of alleles should be used for profile probability calculations, since the D12S391 and vWA loci are not independent.

Others remarks

Based on gene frequencies for 15 NGM STRs, to which data are available for US population samples [5], pairwise genetic distances were calculated between populations using the Reynolds' formulas implemented in PowerMarker software [6]. The analysis (Fig. 2a) showed a clear separation between Caucasian, African Americans, and Hispanic. All the Brazilian samples lie more closely to the Hispanic group than to the African American populations or to the Caucasian. Nevertheless, although the three Brazilian populations from Central West, Southeast, and South are grouped together, the remaining populations from Northeast and North are quite distant from this group. Among the Brazilian samples, the populations from Central West, Southeast, and South are closer to the European cluster. The Northeast group is less distant to the African cluster than the remaining groups. The group showing the highest differentiation was that from the North, most probably due to a higher Native American contribution. Our study, by comprising individuals from all federative units of Brazil, provides additional information on the genetic variation of the Brazilian population. However, since the NJ tree imposes a bifurcating model onto a distance matrix, which may be inadequate for closely related populations, we performed a multidimentional analysis (Fig. 2b), which resembles the same tree topology.

Acknowledgments We would like to thank Bruce Budowle for critical reading of the manuscript.

References

1. Instituto Brasileiro de Geografia Estatística (IBGE): www.ibge.gov.br
2. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10(4):506–513
3. PowerStats v1.2 Software: www.promega.com/geneticidtools/powerstats/
4. Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data at http://alley.eeb.uconn.edu/gda
5. Budowle B, Ge J, Chakraborty R, Eisenberg AJ, Green R, Mulero J, Lagace R, Hennessy L (2010) Population genetic analysis of the NGM STR loci. Int J Legal Med doi:10.1007/s00414-010-0516-7
6. Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21(9):2128–2129
7. Page RDM (2001) TreeView: taxonomy.zoology.gla.ac.uk/rod/rod.html

8. Statistical Package for the Social Sciences–SPSS v15 (2005): www.spss.com
9. National Research Council II Report (1996) The evaluation of forensic evidence. National Academy, Washington, DC
10. Weir BS (1990) "Multiple tests" in genetic data analysis. Sinauer, Sunderland, pp 109–110
11. Phillips C, Fernandez-Formoso L, Garcia-Magariños M, Porras L, Tvedebrink T, Amigo J, Fondevila M, Gomez-Tato A, Alvarez-Dios J, Freire-Aradas A, Gomez-Carballa A, Mosquera-Miguel A, Carracedo A, Lareu MV (2011) Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. Forensic Sci Int Genet 5(3):155–169. doi:10.1016/j.fsigen.2010.02.003
12. O'Connor KL, Hill CR, Valloone PM, Butler JM (2010) Linkage disequilibrium analysis of D12S391 and vWA in US population and paternity samples. Forensic Sci Int Genet. doi:10.1016/j.fsigen.2010.09.003