



Distribution of forensic marker allelic frequencies in Pernambuco, Northeastern Brazil

S.M. Santos^{1,2}, C.A. Souza², K.C.N. Rabelo^{1,2}, P.R.E. Souza³,
R.R. Moura^{1,4}, T.C. Oliveira^{1,2} and S. Crovella^{1,4}

¹Laboratório de Imunopatologia Keizo Asami,
Universidade Federal de Pernambuco, Recife, PE, Brasil

²Laboratório de Perícia e Pesquisa em Genética Forense da Gerência Geral de
Polícia Científica da Secretaria de Defesa Social de Pernambuco, Recife, PE,
Brasil

³Laboratório Genoma, Universidade Federal Rural de Pernambuco,
Recife, PE, Brasil

⁴Departamento de Genética, Universidade Federal de Pernambuco,
Recife, PE, Brasil

Corresponding author: S.M. Santos
E-mail: sandra.santos@sds.pe.gov.br

Genet. Mol. Res. 14 (2): 4303-4310 (2015)

Received March 27, 2014

Accepted August 8, 2014

Published April 30, 2015

DOI <http://dx.doi.org/10.4238/2015.April.30.2>

ABSTRACT. Pernambuco is one of the 27 federal units of Brazil, ranking seventh in the number of inhabitants. We examined the allele frequencies of 13 short tandem repeat loci (*CFSIPO*, *D3S1358*, *D5S818*, *D7S820*, *D8S1179*, *D13S317*, *D16S539*, *D18S51*, *D21S11*, *FGA*, *TH01*, *vWA*, and *TPOX*), the minimum recommended by the Federal Bureau of Investigation and commonly used in forensic genetics laboratories in Brazil, in a sample of 609 unrelated individuals from all geographic regions of Pernambuco. The allele frequencies ranged from 5 to 47.2%. No significant differences for any loci analyzed were observed compared with other publications in other various regions of Brazil. Most of the

markers observed were in Hardy-Weinberg equilibrium. The occurrence of the allele 47.2 (locus *FGA*) and alleles 35.1 and 39 (locus *D21S11*), also described in a single study of the Brazilian population, was observed. The other forensic parameters analyzed (matching probability, power of discrimination, polymorphic information content, paternity exclusion, complement factor I, observed heterozygosity, expected heterozygosity) indicated that the studied markers are very informative for human forensic identification purposes in the Pernambuco population.

Key words: Autosomal STRs; Human identification; Population data

INTRODUCTION

Studies estimating the genetic admixture in established populations indicate that the most striking migratory movement in recent centuries has been the colonization of the Americas (Bonilla et al., 2004; Bedoya et al., 2006; Martinez-Marignac et al., 2007; Tang et al., 2007). One method of increasing the understanding of the genetic makeup of a population is to estimate the contribution of parental populations based on the distribution of allele frequencies of genetic markers (Salzano, 2004).

The most frequently used genetic markers used in forensic laboratories and paternity tests include short tandem repeats (STRs), the Y chromosome, and mitochondrial DNA (Martins et al., 2010; Ziętkiewicz et al., 2012). Those with high discriminatory power, higher frequency of heterozygotes, low mutation rate, high degree of informativeness, smaller size, and lower training stutters, are considered to be the most important for human identification (Hares, 2012).

The distribution of allele frequencies of DNA markers differs between biogeographically separated populations and those from a different origin (Gusmão et al., 2001). The Brazilian populations, including Pernambuco, are considered to be the most heterogeneous in the world (Aguilar et al., 2012). Several studies have been conducted to examine the real contribution of Native Americans, Europeans, and Africans, but the only consensus is that the dynamics of crossovers in Brazil are both unique and complex (Pena et al., 2009).

Material evidence based on DNA analysis is valuable in the process of criminal investigation and identification (Gershaw et al., 2011). Because of the importance and applicability of STR typing (Choi and Seo, 2009) targeting human identification, in November 1997, the Federal Bureau of Investigation (FBI) established 13 loci that would be used as the Combined DNA Index System (CODIS), which were tested and validated for forensic applications (Budowle et al., 1998; Butler, 2005; Michelin et al., 2008).

These markers have been widely used worldwide; even after more than a decade, the markers continue to play an important role in the field of forensics, particularly in the construction of criminal databases (FBI, 1997; Krzyżanińska et al., 2006; Li et al., 2008). When a biological sample is derived from a single source, these 13 markers satisfactorily meet the needs of any forensic laboratory (Ge et al., 2012).

A database of genetic profiles, for forensic purposes or for locating missing persons, has been adopted in many countries, including Brazil; this justified the importance of STRs in the forensic community (Corte-Real, 2004; Walsh, 2004). DNA databases have become central to criminal investigations and for understanding local allele frequencies (Gill et al., 2006;

Jacques and Minervino, 2007; Martins et al., 2010).

In 2012 and 2013, Brazil took an important step in combating crime by publishing Law No. 12.654/12, which establishes DNA as a form of criminal identification (Brazil, 2012), and Decree No. 7.950/2013, which establishes the National Bank of Genetic Profiles and Integrated Bank Network of Genetic Profiles (Brazil, 2013). In this study, we examined the allele frequencies of 13 STR loci. The results will serve to support the actions of forensic genetics in Northeastern Brazil.

MATERIAL AND METHODS

DNA samples

Because of the forensic application of our research, sampling was determined in proportion to the number of inhabitants residing in each of the 8 areas of Pernambuco, with territorial division proposed by the Pact for Life Government Program aimed at preventing and reducing violence. There are 8 territories and 26 integrated security areas, linking overt police actions with police actions in the state (Pernambuco, 2007). Samples were collected mainly in the host cities of each region, resulting in very diverse samples. Buccal mucosa samples derived from 609 adult subjects, unrelated, all born and residing in the State of Pernambuco, were analyzed.

The samples were collected using a mouth swab after obtaining informed consent from each subject. The sample was composed of individuals living throughout the territory, respecting the proportionality of the number of inhabitants, according to the census of the Brazilian Institute of Geography and Statistics (IBGE, 2010): territory 1- capital (N = 115); territory 2 - metropolitan region (N = 150); territory 3 - south forest (N = 37); territory 4 - north forest (N = 60); territory 5 - wild 1 (N = 65); territory 6 - wild 2 (N = 56); territory 7 - backcountry 1 (N = 61); territory 8 - backcountry 2 (N = 65).

This study was approved by the Ethics Committee on Human Research of the Centro de Ciências da Saúde, Universidade Federal de Pernambuco, CEP/CCS/UFPE. The Certificate Presentation to Ethical Consideration (CAAE) was 14961813.2.0000.5208. The territorial division according to the model adopted by the Government of the State of Pernambuco: "Pact for Life" program is displayed in Figure 1.

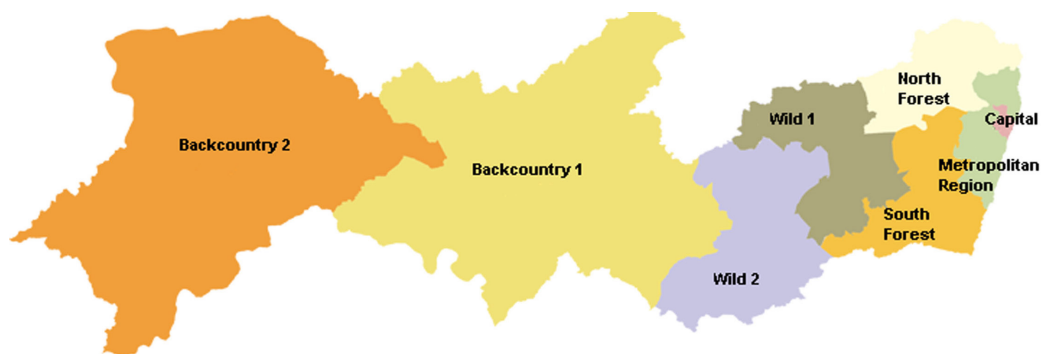


Figure 1. Map of the State of Pernambuco, divided into 8 territories. Adapted from: <http://www.sds.pe.gov.br/>.

DNA extraction

Biological samples were extracted using the DNA IQ™ System kit (Promega, Madison, WI, USA) as recommended by the manufacturer and adopted at the “Laboratório de Perícia e Pesquisa em Genética Forense” (Recife).

STR genotyping

To amplify the STR 13 regions (loci) (*D21S11*, *TH01*, *D3S1358*, *FGA*, *TPOX*, *D8S1179*, *D16S539*, *D7S820*, *D13S317*, *D5S818*, *VWA*, *D18S51*, and *CSF1PO*), as well as amelogenin to identify gender, we used the AMPFLSTR Identifier® Plus® PCR Kit (Life Technologies, Carlsbad, CA, USA) and PowerPlex®16HS (Promega) commercial kits following the manufacturer protocols. Amplicon separation and detection were performed using the ABI3500 automated sequencer (Applied Biosystems; Foster City, CA, USA) with the GS-600 LIZ standard and POP 4 polymer (Applied Biosystems) size. Genotyping was performed using the GeneMapper IDX software (Life Technologies).

Statistical analysis

Allele frequencies, Hardy-Weinberg equilibrium, expected heterozygosity (H_E), and observed heterozygosity (H_O) were calculated using the Arlequin software version 3.5.1.2 (Excoffier, 2010). Other statistical parameters of forensic and paternity importance included matching probability (MP), power of discrimination (PD), polymorphism information content (PIC), power of exclusion (PE), and typical and paternity index (TPI), and were obtained using the PowerStats, version 12 (Tereba, 1999).

RESULTS AND DISCUSSION

Deviations from Hardy-Weinberg equilibrium were observed for the markers *FGA*, *D5S818*, *vWA*, and *D18S51*, but these deviations disappeared after Bonferroni correction (level of significance: 0.0038). The presence of rare alleles or the characteristic substructure of Brazilian populations may account for this finding (Soloum et al., 2013). For all markers, H_E values were very similar to H_O values.

Of the 13 loci studied, *D18S51* showed a PIC value of 0.87, suggesting high informational power. The *TPOX* locus showed a peak of 0.65, indicating that this marker was the least informative. The *D18S51* and *FGA* loci showed higher discrimination powers of 0.973 and 0.976, respectively. Our results for the Brazilian population (Grattapaglia et al., 2001; Fridman et al., 2008; Aguiar et al., 2012) agreed with those for the populations of Pará - North Region (Ribeiro-Rodrigues et al., 2008), the Mato Grosso do Sul - Midwest Region (Silva et al., 2004), Minas Gerais - Southeast Region (del Castillo et al., 2009), and Rio Grande do Sul - South Region (Leite et al., 2003) for most markers and indicators assessed.

In a previous study examining STR allele frequencies in the population of Pernambuco, the *D5S818* and *D3S1338* loci were not evaluated (Dellalibera et al., 2004). The present study confirmed that both are important in forensic laboratories, with higher H_E , high informative power based on the PIC values, and high PD observed (Table 1).

Table 1. Allele frequency distribution of 13 STR loci observed in the present study.

Allele	D21S11	TH01	D3S1358	FGA	TPOX	D8S1179	D16S539	D7S820	D13S317	D5S818	vWA	D18S51	CSF1PO
5		0.000821			0.032841				0.000821				
6		0.206897			0.006568			0.007389	0.000821	0.018062	0.001642	0.000000	0.013957
7		0.241379			0.477833		0.014778	0.171593	0.097701	0.022989		0.000000	0.021346
8		0.166667			0.106732	0.006568	0.176515	0.128079	0.061576	0.046798		0.000000	0.017241
9		0.17734											
9.3		0.202791			0.073071	0.064039	0.091133	0.273399	0.055829	0.055829		0.000011	0.295567
10		0.003284										0.000000	
10.2					0.252874	0.070608	0.302956	0.239737	0.310345	0.311987	0.005747	0.000218	0.280788
11	0.000821	0.000821	0.007389		0.044335	0.119869	0.2578	0.147783	0.325944	0.353859		0.998305	0.300493
12	0.000821		0.004105		0.004926	0.272578	0.128079	0.027094	0.10509	0.178161	0.004105	0.000011	0.056665
13.2													
14			0.092775	0.000821	0.000821	0.258621	0.028736	0.004926	0.04023	0.00821	0.082102	0.021598	0.00027
15			0.29064	0.000821		0.157636			0.000001	0.000821	0.132184	0.030295	0.002463
16			0.29803	0.000821		0.039409				0.000821	0.285714	0.021118	
17			0.172414	0.003284		0.003284				0.000821	0.261905	0.016827	
17.2										0.000821			
18			0.119869	0.013957		0.013957					0.155993	0.006212	
18.2				0.002463		0.002463							
19			0.013957	0.051724		0.051724					0.050082	0.002427	
20	0.000821		0.000821	0.114122		0.114122	0.000821				0.016420	0.001189	
21				0.140394		0.140394					0.003284	0.000357	
21.2				0.002463		0.002463							
22				0.162562		0.162562						0.000006	
22.2				0.004926		0.004926							
23				0.145320		0.145320						0.000006	
23.2				0.14532		0.14532						0.000006	
24			0.000821	0.000821		0.000821						0.000011	
24.2	0.005747			0.167488		0.167488							
25	0.001642			0.10509		0.10509							
25.2	0.000821												
26.2	0.000821												
27	0.035304												
27.2				0.015588		0.015588						0.000006	
28	0.170772			0.005747		0.005747							
29	0.17734			0.005747		0.005747						0.000011	
29.2	0.001642												
30	0.231527												
30.2	0.028736												
31	0.07235												
31.2	0.116585			0.000821		0.000821							
32	0.010673			0.000821		0.000821							
32.2	0.085386												
33.2	0.030378												
34	0.003284												
34.2	0.004926												
35	0.010673												
35.1	0.000821												
36	0.005747												
37	0.001642												
39	0.000821												
43.2				0.000821		0.000821							
47.2				0.000821		0.000821							

In this study, the occurrence of rare alleles, described in a single study for the Brazilian population (Aguiar et al., 2012), was also observed. These included allele 47.2 (locus *FGA*) and alleles 35.1 and 39 (locus *D21S11*). Allele frequencies and other statistical parameters of forensic interest for the 13 loci suggested by CODIS are shown in Tables 1 and 2, respectively.

Our study examining the allele frequencies of the 13 STR loci recommended by CODIS was more robust than previous reports for the state of Pernambuco. The results presented here will be applied for routine statistical calculations of the Laboratório de Perícia e Pesquisa em Genética Forense and will serve as basis for criminal studies performed in Pernambuco.

Table 2. Forensic parameters in Pernambuco, Brazil.

STRs	MP	PD	PIC	PE	TPI	H_o	H_e	P (HWE)
<i>D21S11</i>	0.073	0.927	0.78	0.614	2.6	0.85878	0.85878	0.70609
<i>TH01</i>	0.112	0.888	0.7	0.497	1.94	0.79146	0.79922	0.57657
<i>D3S1358</i>	0.091	0.909	0.74	0.617	2.62	0.79146	0.77437	0.19426
<i>FGA</i>	0.033	0.967	0.86	0.719	3.63	0.867	0.87552	0.00567
<i>TPOX</i>	0.138	0.862	0.65	0.376	1.49	0.66502	0.68845	0.07229
<i>D8S1179</i>	0.061	0.939	0.78	0.595	2.47	0.79967	0.80953	0.42405
<i>D16S539</i>	0.08	0.92	0.75	0.54	2.15	0.76683	0.78549	0.09766
<i>D7S820</i>	0.074	0.926	0.77	0.614	2.61	0.81117	0.79993	0.16464
<i>D13S317</i>	0.087	0.913	0.74	0.52	2.05	0.75698	0.76896	0.15552
<i>D5S818</i>	0.114	0.886	0.7	0.521	2.06	0.76026	0.74008	0.03337
<i>vWA</i>	0.071	0.929	0.77	0.546	2.18	0.77011	0.79904	0.00002
<i>D18S51</i>	0.027	0.973	0.87	0.699	3.38	0.85386	0.883	0.00001
<i>CSF1PO</i>	0.116	0.884	0.69	0.503	1.97	0.74056	0.73981	0.14177

MP = matching probability; PD = power of discrimination; PIC = polymorphism information content; PE = power of exclusion; TPI = typical paternity index; H_e = expected heterozygosity; H_o = observed heterozygosity; P (HWE) = uncorrected P values from Hardy-Weinberg equilibrium test.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Secretaria de Defesa Social de Pernambuco (SDS/PE/ Brasil). R. Moura is recipient of a post-graduate scholarship from Fundação de Amparo à Pesquisa do Estado de Pernambuco (FACEPE, IBPG 0926-2.02/11).

REFERENCES

- Aguiar VRC, Wolfgramm Ede V, Malta FS, Bosque AG, et al. (2012). 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. *Forensic Sci. Int. Genet.* 6: 504-509.
- Bedoya G, Montoya P, Garcia J, Soto I, et al. (2006). Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc. Natl. Acad. Sci. U. S. A.* 103: 7234-7239.
- Bonilla C, Parra EJ, Pfaff CL, Dios S, et al. (2004). Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann. Hum. Genet.* 68: 139-153.
- Brazil (2012). Lei nº 12.654, de 28 de Maio de 2012. Casa Civil. Subchefia para Assuntos Jurídicos. Publicada no DOU de 29.5.2012. Available at [<http://www4.planalto.gov.br/legislacao/legislacao-1/leis-ordinarias>]. Accessed February 3, 2014.
- Brazil (2013). Decreto 7.950, de 12 de Março de 2013. Casa Civil. Subchefia para Assuntos Jurídicos. Publicado no DOU de 13.3.2013. Available at [<http://www4.planalto.gov.br/legislacao/legislacao-1/decretos>]. Accessed February 10, 2014.
- Budowle B, Moretti TR, Niezgoda SJ and Brown BL (1998). CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools. In: Proceedings of the Second European Symposium on Human Identification, Promega Corporation, Madison, 73-88.

- Butler JM (2005). *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*. Elsevier Academic Press, New York.
- Choi JY and Seo TS (2009). An integrated microdevice for high-performance short tandem repeat genotyping. *Biotechnol. J.* 4: 1530-1541.
- Corte-Real F (2004). Forensic DNA databases. *Forensic Sci. Int.* 146 (Suppl): S143-S144.
- del Castillo DM, Perone C, de Queiroz AR, Mourao PH, et al. (2009). Populational genetic data for 15 STR markers in the Brazilian population of Minas Gerais. *Leg. Med.* 11: 45-47.
- Dellalibera E, Havro ML, Souza M, Kajihara K, et al. (2004). Genetic analysis of 13 STR loci in the population from the State of Pernambuco, northeast Brazil. *Forensic Sci. Int.* 146: 57-59.
- Excoffier L (2010). Arlequin Ver. 3.5.1.2. Available at [<http://cmpq.unibe.ch/software/arlequin3>]. Accessed November 12, 2013.
- FBI/Codis (1997). Available at [<http://www.fbi.gov/about-us/lab/codis/ndis-statistics>]. Accessed January, 16, 2014.
- Fridman C, dos Santos PC, Kohler P, Garcia CF, et al. (2008). Brazilian population profile of 15 STR markers. *Forensic Sci. Int. Genet.* 2: e1-e4.
- Ge J, Eisenberg A and Budowle B (2012). Developing criteria and data to determine best options for expanding the core CODIS loci. *Investig. Genet.* 3: 1.
- Gershaw CJ, Schweighardt AJ, Rourke LC and Wallace MM (2011). Forensic utilization of familial searches in DNA databases. *Forensic Sci. Int. Genet.* 5: 16-20.
- Gill P, Fereday L, Morling N and Schneider PM (2006). The evolution of DNA databases-recommendations for new European STR loci. *Forensic Sci. Int.* 156: 242-244.
- Grattapaglia D, Schmidt AB, Costa e Silva, Stringher C, et al. (2001). Brazilian population database for the 13 STR loci of the AmpFISTR Profiler Plus and Cofiler multiplex kits. *Forensic Sci. Int.* 118: 91-94.
- Gusmão L, Alves C and Amorim A (2001). Molecular characteristics of four human Y-specific microsatellites (DYS434, DYD437, DYS438, DYS439) for population and forensic studies. *Ann. Hum. Genet.* 65: 285-291.
- Hares DR (2012). Expanding the CODIS core loci in the United States. *Forensic Sci. Int. Genet.* 6: e52-e54.
- IBGE (2010). Censo Demográfico 2010. Available at [<http://censo2010.ibge.gov.br/resultados>]. Accessed January 23, 2014.
- Jacques GS and Minervino AC (2007). Aspectos Éticos e Legais dos Bancos de Dados de Perfis Genéticos. *Perícia Federal, Brasília*, 26: 17-20. Available at [www.apcf.org.br]. Accessed February 04, 2014.
- Krzyżanińska A, Kowalczyk E, Markowska J and Dobosz T (2006). STR loci D2S1338 and D19S433 in a population sample from the Lower Silesia region. *Arch. Med. Sadowej. Kryminol.* 56: 236-238.
- Leite FP, Menegassi FJ, Schwengber SP, Raimann PE, et al. (2003). STR data for 09 autosomal STR markers from Rio Grande do Sul (southern Brazil). *Forensic Sci. Int.* 132: 223-224.
- Li CT, Guo H, Zhao ZM and Li L (2008). Analysis of genetics and genomics of short tandem repeat loci commonly used in Kinship Testing. *Fa. Yi. Xue Za Zhi.* 24: 214-220.
- Martinez-Marignac VL, Valladares A, Cameron E, Chan A, et al. (2007). Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum. Genet.* 120: 807-819.
- Martins JA, Costa JC, Paneto GG, Figueiredo RF, et al. (2010). Genetic profile characterization of 10 X-STRs in four populations of the southeastern region of Brazil. *Int. J. Legal Med.* 124: 427-432.
- Michelin K, Pacheco AC, Auler-Bittencourt EA, Lima MJM, et al (2008). Banco de Perfis Genéticos no Combate aos Crimes Sexuais. *Perícia Federal, Brasília*. 26: 13-16. Available at [www.apcf.org.br]. Accessed: February 02, 2014.
- Pena S D J, Bastos-Rodrigues L, Pimenta JR and Bydlowski SP (2009). DNA Tests Probe the Genomic Ancestry of Brazilians. *Braz. J. Med. Biol. Res.* 42: 870-876.
- Pernambuco (2007). Pacto pela Vida. Available at [<http://www.pactopelavida.pe.gov.br>]. Accessed January 29, 2014.
- Ribeiro Rodrigues EM, Leite FP, Hutz MH, Palha TJ, et al. (2008). A multiplex PCR for 11 X chromosome STR markers and population data from a Brazilian Amazon Region. *Forensic Sci. Int. Genet.* 2: 154-158.
- Salzano MF (2004). Interethnic variability and admixture in Latin América-social implications. *Rev. Biol. Trop.* 52: 405-415.
- Silva DA, Crouse CA, Chakraborty R, Goes AC, et al. (2004). Statistical analyses of 14 short tandem repeat loci in Brazilian populations from Rio de Janeiro and Mato Grosso do Sul states for forensic and identity testing purposes. *Forensic Sci. Int.* 139: 173-176.
- Soloum de Neves Manta F, Pereira R, Vianna R, Rodolfo Beuttenmüller de Araújo A, et al. (2013). Revisiting the genetic ancestry of Brazilians using autossomal AIM-INDELS. *PLoS One* 8: 75145.
- Tang H, Choudhry S, Mei R, Morgan M, et al. (2007). Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am. J. Hum. Genet.* 81: 626-633.

- Tereba A (1999). Tools for Analysis of Population Statistics, Profile DNA 2:14-16. Available at [<http://www.promega.com/geneticidtools/powerstats/>]. Accessed January 22, 2014.
- Walsh SJ (2004). Recent advances in forensic genetics. *Expert. Rev. Mol. Diagn.* 4: 31-40.
- Zietkiewicz E, Witt M, Daca P, Zebracka-Gala J, et al. (2012). Current genetic methodologies in the identification of disaster victims and in forensic analysis. *J. Appl. Genet.* 53: 41-60.