

Allele frequencies of 6 STR loci (TH01, TPOX, CSF1PO, D13S317, D16S539 and D7S820) in São Tomé e Príncipe (West Africa)

Licínio Manco, Isabel Silva, Pedro Arede, M. Jesus Trovoada, Augusto Abade, Manuela Alvarez

Centro de Investigação em Antropologia e Saúde

Departamento de Ciências da Vida, Universidade de Coimbra, Portugal

lmanco@antrop.uc.pt

Resumo As frequências alélicas de seis marcadores autossómicos do tipo *short tandem repeat* (STR) (TH01, TPOX, CSF1PO, D13S317, D7S820 e D16S539) foram estimadas numa amostra populacional (N=53-85) de São Tomé e Príncipe (África Ocidental). Não foram encontrados desvios significativos ao equilíbrio de Hardy-Weinberg em qualquer dos *loci*. A distribuição de frequências e valores de heterozigotia são semelhantes aos encontrados noutras populações Africanas sub-Saarianas.

Palavras-chave São Tomé e Príncipe; short tandem repeats (STRs); genética populacional.

Abstract Allele frequencies of six autosomal short tandem repeat (STR) markers (TH01, TPOX, CSF1PO, D13S317, D7S820 and D16S539) were estimated in a population sample (N=53-85) of unrelated individuals from São Tomé e Príncipe (West Africa). No deviations from Hardy-Weinberg equilibrium were observed for all loci. Allele frequency distribution and heterozygosity values were similar with those from other sub-Saharan African populations.

Key words: São Tomé e Príncipe; short tandem repeats (STRs); population genetics.

Introduction

Microsatellites or short tandem repeats (STRs) are abundant and widespread in the human genome and can be easily analyzed by automated typing procedures. They are highly discriminating and useful genetic markers in several human fields including linkage analysis (Cashman *et al.*, 1995) and identity testing (Butler, 2004). Microsatellites are also valuable tools in anthropological research, scoring genetic variability in human populations (Rosenberg *et al.*, 2002).

São Tomé e Príncipe, a former Portuguese colony, is a small archipelago located in the Gulf of Guinea (West Africa) inhabited by populations descendent from African slaves brought from the continental mainland. Presently, over 180000 inhabitants live in the two major islands of the archipelago, distributed by three main population groups: Angolares, Forros and Tongas (Trovoada *et al.*, 2007). Several autosomal microsatellite data have been already reported for the Saotomean population (Gusmão *et al.*, 2001; Pereira *et al.*, 1999; 2000).

In this study, the distribution pattern of six autosomal loci (TH01, TPOX, CSF1PO, D13S317, D7S820 and D16S539) currently used in commercial kits for identification purposes was analyzed in a population sample from São Tomé e Príncipe providing additionally information on population genetic diversity. The genetic relationship between the sampled population and other African populations from the mainland continent was also examined.

Material and methods

Unrelated individuals (N=53-85) from the São Tomé e Príncipe archipelago were typed for TH01 (N=85), TPOX (N=80), CSF1PO (N=70), D13S317 (N=57), D7S820 (N=53) and D16S539 (N=57) polymorphic STRs. After informed consent, genomic DNA was extracted from blood samples using the chelex method.

Two PCR multiplex were performed, one for TH01, TPOX, CSF1PO *loci* and the second for D13S317, D7S820 and D16S539 *loci*, using primers labeled with Cy5. Primer sequences were available in <http://www.cstl.nist.gov/biotech/strbase>. PCR amplification was carried out using the Multiplex Qiagen Kit (Qiagen) under the following conditions: initial activation step at 95°C for 15 min, followed by 30 cycles of 95°C – 45 s; 60°C – 45 s and 72°C – 1 min. The PCR products were analyzed in ReproGel High Resolution (GE Healthcare) using the automatic sequencer ALFexpress II (Amersham Pharmacia Biotech). Allele sizes were determined automatically with the software ALFwin Fragment Analyzer 1.00 by comparison with ladders and using control samples provided in the GenePrint STR Multiplex System (for TH01, TPOX, CSF1PO *loci*) and GenePrint SilverSTR III System (for D13S317, D7S820 and D16S539 *loci*) (Promega Corporation).

Allele frequencies, heterozygosity values and exact P values for accordance with Hardy-Weinberg equilibrium (Guo and Thompson, 1992) and for statistical significance on genetic differentiation between populations (Raymond and Rousset, 1995) were calculated using the software package Arlequin, vs. 3.01 (Excoffier *et al.* 2005; <http://cmpg.unibe.ch/software/arlequin3/>). Allele frequencies from Cabinda (Beleza *et al.*, 2004), Mozambique (Alves *et al.*, 2004), Equatorial Guinea (Alves *et al.*, 2005) and Guinea-Bissau (Gonçalves *et al.*, 2002) were used for genetic differentiation tests with São Tomé e Príncipe population sample, applying the Bonferroni correction for multiple tests (Hochberg, 1988). A neighbor-joining tree was constructed using the PHYLIP3.5c software package (Felsenstein, 1993) based on Reynolds Genetic distance (Reynolds *et al.*, 1983) with a bootstrap of 10000 resamples, using population data from different Portuguese regions, Central Portugal (Manco *et al.* 2007-2008), North Portugal (Pinheiro *et al.*, 2005), Madeira (Fernandes *et al.*, 2002), Azores (Velosa *et al.*, 2002), and from other African populations, Cabinda (Beleza *et al.*, 2004), Mozambique (Alves *et al.*, 2004), Equatorial Guinea (Alves *et al.*, 2005) and Guinea-Bissau (Gonçalves *et al.*, 2002).

Results and discussion

Allele frequencies for the 6 STR loci in the population sample of São Tomé e Príncipe are displayed in Table 1. No deviations from Hardy-Weinberg expectations were observed for all *loci*. Population allele distribution and heterozygosity values were similar to other geographic neighboring populations confirming the typical sub-Saharan African genetic profile of the Saotomean population.

When comparing allele frequency data from São Tomé e Príncipe to other 4 African continental mainland populations from Cabinda (Beleza *et al.*, 2004), Mozambique (Alves *et al.*, 2004), Equatorial Guinea (Alves *et al.*, 2005) and Republic of Guinea-Bissau (Gonçalves *et al.*, 2002) several significant differences were found (exact $P < 0.05$) but applying the Bonferroni correction (SISA) the significance level lowered to 0.0127 and significant values were only found with Mozambique at *loci* TH01 (exact $P = 0.007 \pm 0.001$) and TPOX (exact $P = 0.000 \pm 0.000$), and with Guinea-Bissau at *loci* TH01 (exact $P = 0.000$), which seems to reflect the colonization history of São Tomé e Príncipe that received most part of slaves from the West African coast neighboring the archipelago.

Table 1. Allele frequencies and other statistical parameters for the 6 STRs in a population sample from São Tomé e Príncipe.

Allele	TH01 (N=85)	TPOX (N=80)	CSF1PO (N=70)	D7S820 (N=57)	D13S317 (N=53)	D16S539 (N=57)
5						0.009
6	0.053	0.056				
7	0.359	0.050	0.071			
8	0.359	0.300	0.050	0.254	0.009	0.018
9	0.141	0.300	0.029	0.123		0.263
9.3	0.070					
10	0.018	0.075	0.336	0.342	0.028	0.088
11		0.181	0.164	0.193	0.377	0.307
12		0.038	0.279	0.079	0.387	0.211
13			0.071		0.142	0.079
14				0.009	0.047	0.026
15					0.009	
<i>p</i>	0.626	0.180	0.912	0.636	0.855	0.149
Ho	0.741	0.763	0.843	0.737	0.698	0.825
He	0.719	0.781	0.775	0.766	0.698	0.784

N: number of samples; *p*: Hardy-Weinberg equilibrium, exact test; Ho: observed heterozygosity; He: expected heterozygosity;

A neighbor-joining tree was constructed based upon the 6 STR allele frequencies in our sampled population and from populations from Central Portugal (Manco *et al.*, 2007/2008), North Portugal (Pinheiro *et al.*, 2005), Madeira (Fernandes *et al.*, 2002), Azores (Velosa *et al.*, 2002), Cabinda (Beleza *et al.*, 2004), Mozambique (Alves *et al.*, 2004), Equatorial Guinea (Alves *et al.*, 2005) and Guinea-Bissau (Gonçalves *et al.*, 2002) (Figure 1).

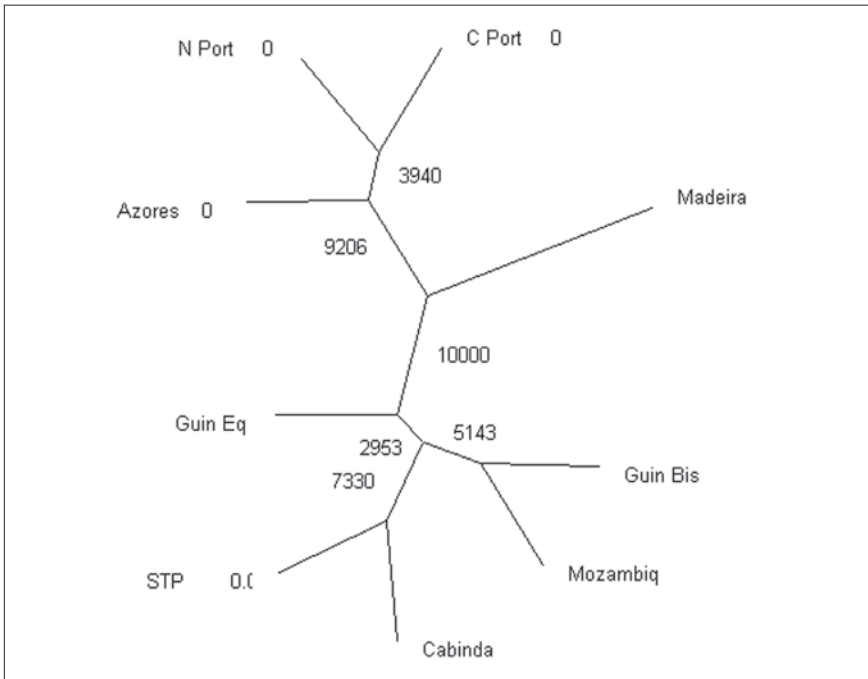


Figure 1. Neighbor-joining tree based on allele frequencies of the 6 STR loci (TH01, TPOX, CSF1PO, D13S317, D7S820 and D16S539) in the studied population sample from São Tomé e Príncipe (STP), and in different Portuguese regions (C Port – Central Portugal, N Port – North Portugal, Madeira, Azores) and other African populations (Cabinda, Mozamb – Mozambique, Guin Eq – Equatorial Guinea and Guin Bis – Guinea-Bissau) with previous reported data.

The topology of the tree unequivocally differentiates African and European populations. Among the African cluster, São Tomé e Príncipe is closer to Cabinda and Equatorial Guinea consistent with the low significant differences at STR loci. The population from Central Portugal shows close genetic proximity with the neighbor mainland North Portuguese population than to Madeira and Azores archipelagos in accordance to the regional distances.

Acknowledgements

This work was performed in part with the support of Centro de Investigação em Antropologia e Saúde – Universidade de Coimbra. Special thanks to the volunteer donors from São Tomé e Príncipe and the community leaders.

Bibliographic references

- Alves, C.; Gusmão, L.; López-Parra, A. M.; Soledad-Mesa, M.; Amorim, A.; Arroyo-Pardo E. 2005. STR allelic frequencies for an African population sample (Equatorial Guinea) using AmpFISTR Identifiler and Powerplex 16 kits. *Forensic Science International*, 148: 239-242.
- Alves, C.; Gusmão, L.; Damasceno, A.; Soares, B.; Amorim, A. 2004. Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 System) and a report on genotypic variations. *Forensic Science International*, 139:201-205.
- Beleza, S.; Alves, C.; Reis, F.; Amorim, A.; Carracedo, A.; Gusmão, L. 2004. 17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola). *Forensic Science International*, 141: 193-196.
- Butler, J. M. 2004. Short tandem repeat analysis for human identity testing. *Current Protocols on Human Genetics*, 14: 14.8.
- Cashman, S. M.; Patino, A.; Martinez, A.; Garcia-Delgado, M.; Miedzybrodzka, Z.; Schwarz, M.; Shrimpton, A.; Ferec, C.; Raguene, O.; Macek, M. Jr.; Morral, N.; De Arce, M. 1995. Identical intragenic microsatellite haplotype found in cystic fibrosis chromosomes bearing mutation G551D in Irish, English, Scottish, Breton and Czech patients. *Human Heredity*, 45: 6-12.
- Hochberg Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75: 800–802.
- Excoffier, L.; Laval, G.; Schneider, S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1: 47-50.
- Felsenstein, J. 1989. Phylip-phylogeny inference package (version 3.2). *Cladistics*, 5: 164-166.
- Fernandes, A. T.; Brehm, A.; Alves, C.; Gusmão, L.; Amorim, A. 2002. Genetic profile of the Madeira Archipelago population using the new PowerPlex16 System kit. *Forensic Science International*, 125: 281-283.

- Gonçalves, R.; Jesus, J.; Fernandes, A. T.; Brehm, A. 2002. Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex 16 System kit. *Forensic Science International*, 129: 78-80.
- Gusmão, L.; Prata, M. J.; Miranda, C.; Trovoada, M. J.; Amorim, A. 2001. STR data from S. Tomé e Príncipe (Gulf of Guinea, West Africa). *Forensic Science International*, 116: 53-54.
- Guo, S. W.; Thompson, E. A. 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics*, 48: 361-372.
- Manco L., Silva I., Arede, P., Abade, A., Alvarez, M. 2007/2008. Estudo populacional de 6 STRs (TH01, TPOX, CSF1PO, D7S820, D13S317 e D16S539) na região Centro de Portugal. *Antropologia Portuguesa*, 24/25: 141-148.
- Pereira, L.; Gusmão, L.; Prata, M. J.; Mota, P.; Trovoada, M. J.; Amorim, A. 1999. Detection of additional structural variation at the FES/FPS system and population data from S. Tomé e Príncipe and North Portugal. *International Journal of Legal Medicine*, 112: 204-206.
- Pereira, L.; Gusmão, L.; Prata, M. J.; Mota, P.; Trovoada, M. J.; Amorim, A. 2000. Variation in STR Loci of the human myelin basic protein gene: north Portugal and São Tomé e Príncipe. *Human Biology*, 72: 481-487.
- Pinheiro, M. F.; Cainé L.; Pontes, L.; Abrantes, D.; Lima, G.; Pereira, M. J.; Rezende, P. 2005. Allele frequencies of sixteen STRs in the population of Northern Portugal. *Forensic Science International*, 148: 221-223.
- Raymond, M.; Rousset, F. 1995. An exact test for population differentiation. *Evolution*, 49: 1280-1283.
- Reynolds, J., Weir, B. S., Cockerham, C. C. 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105: 767-779.
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., Feldman, M. W. 2002. Genetic structure of human populations. *Science*, 298: 2381-2385.
- Trovoada, M. J., Tavares, L., Gusmão, L., Alves, C., Abade, A., Amorim, A., Prata, M. J. 2007. Dissecting the genetic history of São Tomé e Príncipe: a new window from Y-chromosome biallelic markers. *Annals of Human Genetics*, 71: 77-85.
- Velosa, R. G.; Fernandes, A. T.; Brehm, A. 2002. Genetic profile of the Azores Archipelago population using the new PowerPlex 16 system kit. *Forensic Science International*, 129: 68-71.

Artigo recebido a 7 de Julho de 2010 e aceite a 15 de Fevereiro de 2011.

