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

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 Principe (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 Principe 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 Principe 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 Principe 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 locus TH01 (exact P=0.007±0.001) and TPOX (exact P=0.000±0.000), and with Guinea-Bissau at loci TH01 (exact P=0.000), which seems to reflect the colonization history of São Tomé e Principe 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.

<table>
<thead>
<tr>
<th>Allele</th>
<th>TH01 (N=85)</th>
<th>TPOX (N=80)</th>
<th>CSF1PO (N=70)</th>
<th>D7S820 (N=57)</th>
<th>D13S317 (N=53)</th>
<th>D16S539 (N=57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.009</td>
</tr>
<tr>
<td>6</td>
<td>0.053</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.359</td>
<td>0.050</td>
<td>0.071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.359</td>
<td>0.300</td>
<td>0.050</td>
<td>0.254</td>
<td>0.009</td>
<td>0.018</td>
</tr>
<tr>
<td>9</td>
<td>0.141</td>
<td>0.300</td>
<td>0.029</td>
<td>0.123</td>
<td></td>
<td>0.263</td>
</tr>
<tr>
<td>9.3</td>
<td>0.070</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.018</td>
<td>0.075</td>
<td>0.336</td>
<td>0.342</td>
<td>0.028</td>
<td>0.088</td>
</tr>
<tr>
<td>11</td>
<td>0.181</td>
<td>0.164</td>
<td>0.193</td>
<td>0.377</td>
<td>0.307</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.038</td>
<td>0.279</td>
<td>0.079</td>
<td>0.387</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>0.071</td>
<td></td>
<td>0.142</td>
<td>0.079</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>0.009</td>
<td>0.047</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.009</td>
</tr>
<tr>
<td>(\chi^2)</td>
<td>0.626</td>
<td>0.180</td>
<td>0.912</td>
<td>0.636</td>
<td>0.855</td>
<td>0.149</td>
</tr>
<tr>
<td>Ho</td>
<td>0.741</td>
<td>0.763</td>
<td>0.843</td>
<td>0.737</td>
<td>0.698</td>
<td>0.825</td>
</tr>
<tr>
<td>He</td>
<td>0.719</td>
<td>0.781</td>
<td>0.775</td>
<td>0.766</td>
<td>0.698</td>
<td>0.784</td>
</tr>
</tbody>
</table>

N: number of samples; \(\chi^2\): 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


Allele frequencies of 6 STR loci


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