Pattern of mtDNA Variation in Three Populations from São Tomé e Príncipe

M. J. Trovoada[1], L. Pereira[2], L. Gusmão[2], A. Abade[1], A. Amorim[2,3] and M. J. Prata[2,3,*]

[1]*Departamento de Antropologia da Universidade de Coimbra, 3049 Coimbra Codex, PORTUGAL*

[2]*Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias s/n°, 4200-465 Porto, PORTUGAL*

[3]*Faculdade de Ciências da Universidade do Porto, Pr. Gomes Teixeira, 4099-002 Porto, PORTUGAL*

Summary

We have analysed the matrilineal genetic composition of three self-reported ethnic groups from São Tomé e Príncipe (Gulf of Guinea), an African archipelago whose settlement begun in the late fifteenth century. Sequence data from the hypervariable segments I (HVS-I) and II (HVS-II) were obtained for 30 Angolares, 35 Forros and 38 Tongas. The repertory of mtDNA lineages in São Tomé e Príncipe denoted a fully African maternal pool, primarily arisen from a Central/Southwestern substratum. The absence of any lineages of putative European descent means that the European impact at the mitochondrial pool was virtually nil. Angolares showed a clear reduction of mtDNA diversity and a slight genetic differentiation relative to Tongas or Forros, whereas the latter two groups did not present any signs of genetic boundaries between each other. The data obtained here reinforce the depiction of genetic substructuring in São Tomé e Príncipe previously derived from Y-chromosome STRs. In addition, the crossing of mtDNA and Y-STR information led to the inference that the female mediated gene flow within the archipelago was less restricted than the male, a pattern that could be framed in the cultural traditions and socio-historical interactions among the groups.

Keywords : mtDNA, haplogroups distribution, São Tomé e Príncipe

Introduction

São Tomé e Príncipe is a small archipelago located on the Equator in the Gulf of Guinea, approximately 300 Km from the nearest African seashore (Fig. 1). Presently, over 150000 inhabitants live in the two major islands (São Tomé and Príncipe) and several small islets that constitute the archipelago. All were probably uninhabited when first visited by Portuguese navigators in the early 1470s. Soon afterwards settlement began, illustrating the important reconfiguration of the peopling of Africa that took place from the beginning of the Discoveries period (Newman, 1995).

*Correspondence: Maria João Prata, Instituto de Patologia e Imunologia Molecular, da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias s/n, 4200-465 Porto - Portugal. Tel: +351 22 5570700. Fax: +351 22 5570799. E-mail: mprata@ipatimup.pt

The settlement of São Tomé e Príncipe involved two main phases that parallel major shifts in the social, demographic and economic tissues. The first, initiated soon after discovery, involved a huge movement of people that included the European colonists - almost exclusively males, belonging to very disparate social strata and mainly from Portugal - and slaves brought from the African Western mainland, who largely overwhelmed the former. The second phase began in the middle of the nineteenth century and was triggered by the development of the coffee and cocoa economy boom. By that time, slavery was already officially abolished and the archipelago faced strong social conflicts. In order to meet the new labour needs, the local supply was largely supplemented with thousands of immigrant labourers mainly from Angola, Mozambique and Cabo Verde, which were then, as was São Tomé e Príncipe, Portuguese colonies.

Figure 1 Map of Africa showing São Tomé e Príncipe and location of the samples compared in this study. Population codes are defined in Table 1.

Among present day São Tomé e Príncipe inhabitants a common perception is the existence of three distinct population groups, Angolares, Forros and Tongas, which are also referred to in the ethnohistorical bibliography from the archipelago (Tenreiro, 1961; Ambrósio, 1984; Henriques, 2000). Forros are considered to be the descendents of an African *élite*, whose initial establishment dates back to the early settlement period. Tongas are the descendents of the economically and socially disfavoured contract labourers that entered the archipelago from the mid nineteenth century. The origin of the Angolares is less clear, notwithstanding the fact that their identification as a distinct community has strong support from their geographical confinement to the south-eastern tip of São Tomé, from their cultural and livelihood traditions, and even from their own language, one of the most important badges of population identity. In a previous study (Trovoada *et al*. 2001) we analysed Y-chromosome STRs in order to investigate the male-mediated relationships among these three ethnic groups, and the genetic structure of São Tomé e Príncipe. We detected signs of a slight genetic microdifferentiation of Angolares comparative to the other two groups. However, the demographic history of a population cannot be fully inferred from the pattern of variation of a single kind of marker or genomic segment. Although some recent works focusing on São Tomé e Príncipe have dealt with a number of autosomal loci (Tomás *et al*. 2002), or even mtDNA variation (Mateu *et al*. 1997), affording valuable insights for understanding the genetic history of the archipelago, they were not designed to investigate the genetic relationships among Angolares, Forros and Tongas.

Here we present sequence data from mtDNA detected by surveying hypervariable region I (HVS-I) and II (HVS-II) in individuals from the three self-reported São Tomean groups.

This paper mainly addresses the issue of whether mtDNA variation *per se*, or the combined inferences from Y-chromosome STRs and mtDNA, would be consistent with the traditionally accepted and ethnohistorically documented sub-structuring of the population of São Tomé e Príncipe.

Material and Methods

Sampling and DNA Extraction

Blood samples were collected from unrelated volunteers from the three self-reported São Tomé e

Príncipe population groups. A total of 103 samples were analysed, consisting of 30 Angolares, 35 Forros and 38 Tongas.

DNA was extracted from whole blood by standard phenol chloroform (Valverde *et al.* 1993) or Chelex (Lareu *et al.* 1994) methodologies.

mtDNA Amplification and Sequencing

The primers, PCR and sequencing strategy were as described in Pereira *et al.* (2001). PCR products were purified using Microspin^{TM} S-300 HR columns (Pharmacia) according to the manufacturer's specifications. The sequence reactions were carried out using the kit BigDye^{TM} Terminator Cycle Sequencing Ready Reaction (AB Applied Biosystems) and then a protocol based on $MgCl_2$/ethanol precipitation was used for purification of samples. They were run in an automatic sequencer, ABI Prism 3100 Genetic Analyzer, and analysis was carried out with the DNA Sequencing Analysis 3.7 software (AB Applied Biosystems).

Genetic Analysis and Population Comparisons

The nucleotide positions considered for the analysis were from 16051 to 16390 for HVS-I and from 73 to 340 for HVS-II (Anderson *et al.* 1981). For comparison with other populations from Africa, we only considered HVS-I from nucleotide positions 16090-16365, because this is the stretch for which most information is available.

The haplogroup classification was according to Watson *et al.* (1997), Rando *et al.* (1998), Macaulay *et al.* (1999), Richards & Macaulay (2000), Bandelt *et al.* (2001), Pereira *et al.* (2001) and Salas *et al.* (2002).

To evaluate the accuracy of our mtDNA data set, we have applied the approach described in Bandelt *et al.* (2002). After filtering for "speedy transitions," a network of sequences was constructed with the program NETWORK 3.1 (http://www.fluxus-engineering.com) using the median-joining algorithm, with the tolerance threshold ε set to infinity.

Molecular indices, analysis of variance (AMOVA), mismatch distributions and tests of the standard neutral model (via Fu's F_s statistic) were calculated with the ARLEQUIN 2.0 software (Schneider *et al.* 2000).

Relative frequencies of sub-Saharan haplogroups from Angolares, Forros, Tongas and other African samples were used to perform Principal Component (PC) analysis. In order to obtain an accurate definition we restricted the analysis to Western/Central/Eastern African populations and, furthermore, when multiple samples were available from the same country, namely for Senegal, Nigeria and Kenya, they were pooled in a single sample. The African populations used for comparisons are displayed in Table 1 together with the corresponding code, place of origin, sample size and bibliographic reference.

Results

The HVS-I and HVS-II sequences obtained, and their distribution among Angolares, Forros and Tongas, are presented in the Appendix.

Checking for Phantom Mutations

In order to evaluate if any systematic artifact introduced in the course of the sequencing process could have produced phantom mutations compromising the accuracy of our data set, we applied the filtering approach described in Bandelt *et al.* (2002). The network obtained for the HVS-I weighty variation presented some secondary reticulation that was basically due to the presence of alternative transversions in sites 114, 265 and 286. The two different transversions in those three sites have already been described in other African data sets (Graven *et al.* 1995; Pereira *et al.* 2001; Salas *et al.* 2002), and so they are not private in our sequences. When these three non-binary sites were treated as binary we obtained the tree displayed in Fig. 2. It shows that after filtering out the speedy transitions the most frequent haplotype corresponds to the Cambridge reference sequence. The São Tomean weighty network still gives one triangle, indicating ambiguity in the mutational change involving site 114, and one rectangle involving sites 165/186, but the tree is quite harmonious and does not present signals of artificial variation at any site. Therefore, the observed pattern seems to indicate that phantom mutations are not affecting our mtDNA data set.

Table 1 Code, place of origin, sample size and bibliographic references for the African populations considered in the study

Code	Place	Sample size	Reference
North Africa			
SAH	Western Sahara	25	Rando *et al*. (1998)
MA	Mauritania	30	Rando *et al*. (1998)
MO	Morocco	32	Rando *et al*. (1998)
BM	Morocco (Berber)	60	Rando *et al*. (1998)
EGY	Egypt	68	Krings *et al*. (1999)
MZB	Algeria (Mozabite)	85	Côrte-Real *et al*. (1996) Macaulay *et al*. (1999)
West Africa			
HA+KA	Niger (Hausa and Kanuri)	20+14	Watson *et al*. (1997)
FUL	Nigeria (Fulbe)	60	Watson *et al*. (1997)
SON+TU	Nigeria (Songhai and Tuareg)	10+23	Watson *et al*. (1997)
YOR	Nigeria (Yoruba)	21+14	Watson *et al*. (1997); Vigilant *et al*. (1991)
SEN	Senegal	50	Rando *et al*. (1998)
SER	Senegal (Serer)	23	Rando *et al*. (1998)
WO	Senegal (Wolof)	48	Rando *et al*. (1998)
MAN	Senegal (Mandenka)	119	Graven *et al*. (1995)
CVNW	Cabo Verde NW	108	Brehm *et al*. (2002)
CVSE	Cabo Verde SE	184	Brehm *et al*. (2002)
Central Africa			
MBU	Democratic Republic of Congo (Mbuti)	20	Vigilant *et al*. (1991)
BIA	Central African Republic (Biaka)	17	Vigilant *et al*. (1991)
BIO	Equatorial Guinea (Bubi)	45	Mateu *et al*. (1997)
STM	São Tomé e Príncipe	50	Mateu *et al*. (1997)
ANG	São Tomé e Príncipe (Angolares)	30	This study
FOR	São Tomé e Príncipe (Forros)	35	This study
TON	São Tomé e Príncipe (Tongas)	38	This study
ST	São Tomé e Príncipe	103	This study
East/Southeast Africa			
TK	Kenya (Turkana)	36	Watson *et al*. (1997)
KIK	Kenya (Kikuyu)	25	Watson *et al*. (1997)
SO	Somalia	27	Watson *et al*. (1997)
NUB	Nubia	80	Krings *et al*. (1999)
SUD	Southern Sudan	76	Krings *et al*. (1999)
MOZ	Mozambique	109	Pereira *et al*. (2001)
South Africa			
KNG1	Botswana (!Kung)	25	Vigilant *et al*. (1991)
KNG2	South Africa (!Kung)	43	Chen *et al*. (2000)
KWE	South Africa (Khwe)	31	Chen *et al*. (2000)

Descriptive Parameters of HVS-I and HVS-II Sequences

Table 2 shows several diversity measures calculated for HVS-I and/or HVS-II in Angolares, Forros and Tongas and in the overall São Tomean sample. When ranking the three São Tomean groups, a rather clear distinction stood out between Angolares on one side, and Tongas and Forros, on the other, with Angolares showing reduced diversity levels particularly concerning mean heretozygosity and proportion of different haplotypes. Yet, within the Angolares the mean number of pairwise differences between haplotypes tended to be slightly higher than that within Forros or Tongas, which means that the reduction in diversity within Angolares did not corresponded to the narrowing of the molecular relatedness between mtDNA lineages. This feature was clearly evident in the pattern of mismatch distribution for HVS-I in Angolares, which is shown in Fig. 3, as well as the distribution in Tongas, Forros and in the overall São Tomean sample. Angolares presented the highest number of sequences with zero differences, but also had

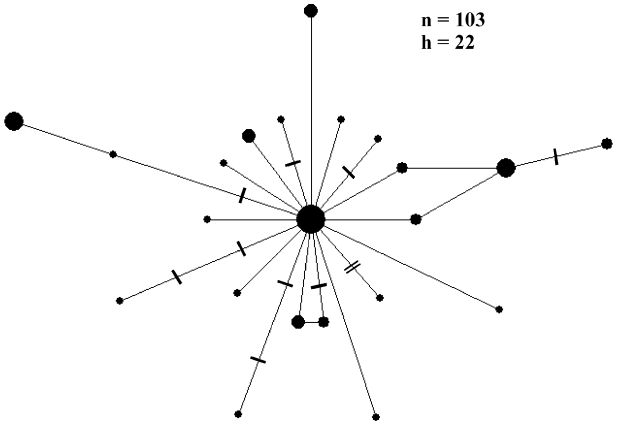

Figure 2 Network representing the weighty variation within 16051-16365 for the São Tomé e Príncipe mtDNA data. Every unit-length link signifies one weighty transition except where "–" indicates a transversion or "="a deletion. The size of the nodes is proportional to the number of sampled haplotypes. n = sample size; h = number of haplotypes relative to weighty mutations.

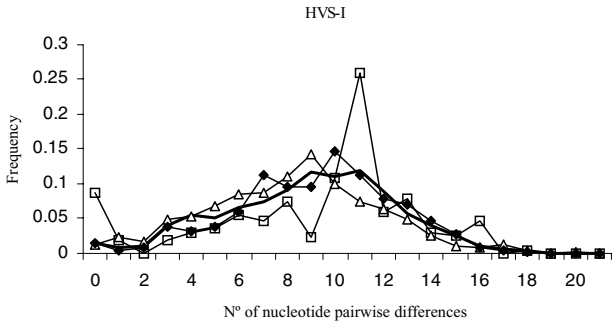

Figure 3 Mismatch distributions for HVS- I within Angolares (□), Forros (Δ), Tongas (♦) and the overall São Tomean sample (—).

a mode of the highest number of differences; the raggedness coefficient (RC) of its distribution was high (0.085) and differed statistically ($P < 0.001$) from the expectation under an expansion model. Although questionable as an appropriate null hypothesis for such a complex settlement history, it is interesting that using this model no significant differences were found neither in Forros and Tongas, which had much smoother mismatch distributions (RC: 0.007 and 0.011, respectively), nor in the overall São Tomean sample (RC: 0.005), which showed a bell-shaped pattern quite similar to that previously found for a different sample from São Tomé e Príncipe by Mateu *et al.* (1997).

Haplogroups Profiles

Among the 103 São Tomeans analysed, 71 different sequences were found which belonged to 32 distinct haplogroups (see Appendix and Fig. 4), all specific to sub-Saharan regions. Therefore no single instance of a putative European sequence was detected.

Within the overall sample, four haplogroups (not considering sub-clusters), L1b, L1c, L2a1 and L3e1*, attained considerably high frequencies (10% was the low boundary established) representing 56% of the sequences.

	Population	*K*(K/N)[a]	*S*(S/l)[b]	*H*(SE)[c]	*M*[d]
HVS-I	Angolares	16 (53.3)	39 (11.4)	0.912± 0.037	9.11
	Forros	29 (82.9)	44 (12.9)	0.988± 0.010	8.08
	Tongas	28 (73.7)	56 (16.5)	0.984± 0.009	9.09
	S. Tomé	*62 (60.2)*	*68 (20.0)*	*0.985± 0.004*	*8.87*
HVS-II	Angolares	17 (56.7)	22 (8.2)	0.913± 0.038	7.63
	Forros	27 (77.1)	29 (10.8)	0.985± 0.010	5.99
	Tongas	22 (57.9)	24 (9.0)	0.967± 0.012	6.27
	S. Tomé	*52 (50.5)*	*32 (11.9)*	*0.979± 0.005*	*6.73*
HVS-I + HVS-II	Angolares	18 (60.0)	60 (9.9)	0.919± 0.038	16.62
	Forros	30 (85.7)	73 (12.0)	0.992± 0.008	14.07
	Tongas	29 (76.3)	80 (13.2)	0.986± 0.009	15.36
	S. Tomé	*70 (68.0)*	*100 (16.5)*	*0.988± 0.004*	*15.57*

Table 2 Diversity indices for HVS-I and HVS-II within Angolares, Forros, Tongas and the overall São Tomean sample

[a] K = number of different sequences and percentage of sample size (N) in brackets
[b] S = number of segregating sites and percentage of all sites in brackets. l = sequence length
[c] H = sequence diversity ± standard error
[d] M = average number of pairwise differences

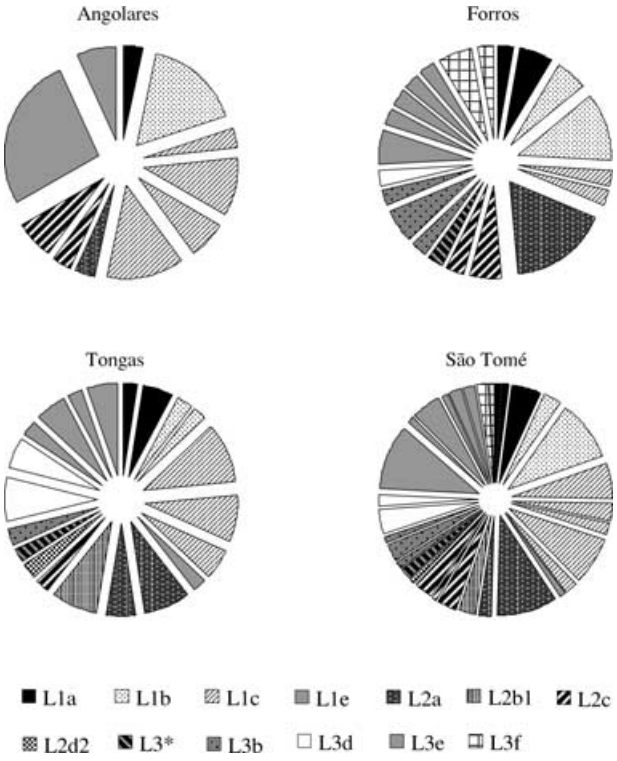

Figure 4 mtDNA haplogroup profiles in Angolares, Forros, Tongas and in overall São Tomé. Subclusters within major cluster are only schematically represented by breaks in specific colour' patterns.

L1b is a cluster particularly concentrated in Western Africa, although with some overflow into Central and North Africa (Salas *et al*. 2002).

The phylogeography of L1c is still very incipient. The majority of L1c are from Central Africa, with a few in the west and the southeast. Likely due to drift effects - the haplogroup reaches particularly high frequency among Biaka, the western group of Pygmies from Republic Central African, and among the Bubi from Bioko. Recently, Salas *et al*. (2002) noticed that more than one-third of L1c haplotypes in their database belonged to African Americans, and only a few showed matches with continental Africans. These findings, together with the geographical distribution of L1c, led the authors to suggest that the origin of the cluster could probably be somewhere in Central Africa, towards the Atlantic west coast.

L2a1 is a sub-cluster of L2a, the most frequent and widespread mtDNA cluster in Africa. The sub-cluster is well represented in southeastern Africans but its place of origin appears to have been West Africa (Salas *et al*. 2002).

The fourth most frequent haplogroup found in São Tomé was L3e1*, the oldest and more diverse clade among L3e types. Albeit practically omnipresent in sub-Saharan Africa, it is supposed to have had a West/Central African origin, having spread to southern regions with the Bantu expansion (Bandelt *et al*. 2001).

Seven haplogroups, jointly accounting for 33% of the sample, had frequencies between 3% and 7%: L1a, L2b, L2c, L3*, L3b, L3d and L3e2. All but the first are either widespread or typical of Western Africa; L1a tends to be more concentrated in East Africa, the most likely region for its origin.

The set of rare haplogroups, those appearing only once or twice in the São Tomé sample, is quite diverse and includes, for instance, lineages specific to the Atlantic Western coast (L3e4), from Eastern Africa (L1e) or from Southern/Eastern Africa (L3e1a, which is most common in the Bantu or Khoisan, Bandelt *et al*. 2001).

Now bringing into focus the three São Tomean groups, each of the internal haplogroup profiles presented some singularities comparative to the overall sample. Among Forros and Tongas, apart from minor frequency differences, the basic profile as characterised by the presence of many different haplogroups is very similar. In contrast, the Angolares show a reduced number of haplogroups and a particularly high frequency of L3e1* and L1c lineages, representing 60% of their mtDNA pool.

Genetic Structure Analysis

Analysis of population pairwise FSTs revealed statistically significant differences between Angolares and Forros (FST = 0.037, $P = 0.04$) or Angolares and Tongas (FST = 0.030, $P = 0.04$); while Tongas and Forros did not showed statistical differences (FST = 0.01, $P = 0.14$).

Consistently, when AMOVA was applied without any hierarchical grouping, the overall FST value for the São Tomean sample was low (0.026), but significantly greater than zero ($P = 0.02$), indicating that a considerable proportion of the mtDNA variation resulted from interpopulation differences.

When applying AMOVA taking one ethnic group and comparing it with the other two, the fraction of variance attributable to differences between groups was always non-statistically significant, but attained the highest value (2.44%, $P = 0.33$) when Angolares were distinguished from the non-Angolares. Then the proportion of variance between populations took its minimum value (1.01% compared with 3.78% for Tongas *vs* Angolares + Forros, or 3.23% for Forros *vs* Angolares + Tongas, no value statistically different from zero).

Comparison with Other African Populations

The African populations listen in Table 1 were used for comparison with the data set now obtained. In this comparative exercise only HVS-I between positions 16090-16365 was considered due to the limitations of the comparative data available.

Molecular Diversity

Several molecular diversity measures were computed and are shown in Table 3. Gene diversity values in Forros, Tongas and in the two overall São Tomean samples are very similar to those found in mainland populations from the cluster of Central/Western populations in which São Tomé e Príncipe geographically belongs. Distinctly, the lower diversity of Angolares clearly resembles the situation in Western African islands such as Cabo Verde NW or Bioko. Likewise, Fu's Fs statistic in Angolares was found to assume a small negative value (statistically not significant), visibly falling in the range of values typical either of islander populations (again like Cabo Verde NW or Bioko) or from the Pygmies and Khoisan-speaking populations from Southern Africa. The remaining African populations, including Forros and Tongas, are characterized by large negative values for Fu's Fs, which has been consensually interpreted as a signal of population demographic expansion.

Principal Component Analysis

Fig. 5 shows a two-dimensional plot that partially summarises the results obtained with the PC analysis. In the first two PCs, which accounted for 54% of the variance, three population clusters were clearly distinguishable: one that occupied the lower-left quarter of the plan and that, apart from the extremely individualised Biaka, contained the island population of Bioko, and the Angolares from São Tomé e Príncipe; positioned inside the opposite lower quarter of the plan, the cluster of eastern African populations; and in the middle of the upper half-plan, the cluster comprising the remaining populations from the Central/Western Atlantic fringe (including the São Tomean sample studied by Mateu *et al.*, the Forros and Tongas). The second PC reflected a rather consistent cline from NW towards SE Africa (Biaka excepted). L1c is the haplogroup that mainly contributed to the PC1 or PC2 (responsible for 31% and 23% of the variance, respectively). This haplogroup is distinctively very frequent in the Biaka, and is responsible for moving the Bioko and Angolares, where it also reaches considerably high frequencies, in the Pygmies direction. For PC1, L3* is the second haplogroup with highest impact, and is determinant in the positioning of most of the Eastern African populations. L1a, in the negative range of values, and L2c, in the positive one, are the two other haplogroups that mostly accounted for PC2. L1a represented 23.5% of the two single lineages found in the Biaka and was also very frequent in eastern Africa; L2c was particularly frequent in Senegal and almost absent in East Africa.

In this population/haplogroup framework, the near overlapping between Cabo Verde is notorious, particularly for the SE fraction and Senegal. The position of the Forros and Tongas is somewhat withdrawn from this cluster, in an intermediate position between the most Northern Atlantic populations and the Eastern ones. The Angolares are the most deviant São Tomean group, showing tight affinities with Bioko and being, among the populations considered, that most closely approaching the Biaka.

Discussion

General Pattern of mtDNA Variation in São Tomé e Príncipe

The diversity indices obtained here for the overall São Tomean sample are in full accordance with previous data derived for the archipelago by Mateu *et al.* (1997). The observed high level of mtDNA diversity is in

Table 3 Diversity and neutrality measures for HVS-I in African populations

Population	N[a]	K(K/N)[b]	S(S/l)[c]	H(SE)[d]	M[e]	Fu's Fs
North Africa						
SAH	25	20 (80.0)	29 (10.5)	0.973±0.022	5.11	−12.41*
MA	30	22 (73.3)	28 (10.1)	0.970±0.018	5.83	−11.48*
MO	32	29 (90.6)	44 (15.9)	0.988±0.014	5.84	−25.01*
BM	60	38 (63.3)	47 (17.0)	0963±0.015	4.44	−25.74*
EGY	68	59 (86.8)	66 (23.9)	0.993±0.005	6.82	−25.07*
MZB	85[f]	29 (34.1)	35 (12.7)	0.942±0.010	4.73	−11.14*
West Africa						
HA+KA	34	30 (88.3)	41 (14.9)	0.991±0.010	6.19	−24.70*
FUL	60	38 (63.3)	43 (15.6)	0.972±0.010	6.82	−23.15*
SON+TU	33	29 (87.9)	41 (14.9)	0.992±0.009	7.26	−21.30*
YOR	34[f]	32 (94.1)	44 (15.9)	0.996±0.008	7.31	−25.01*
SEN	50	42 (84.0)	41 (14.9)	0.987±0.008	6.24	−25.22*
SER	23	21 (91.3)	40 (14.5)	0.992±0.015	8.09	−11.98*
WO	43	39 (90.7)	42 (15.2)	0.991±0.006	7.50	−24.97*
MAN	110[f]	46 (41.8)	47 (17.0)	0.963±0.008	6.23	−24.53*
CVNW	108	28 (25.9)	32 (11.6)	0.908±0.013	5.84	−5.58
CVSE	184	101 (54.9)	69 (25.0)	0.984±0.003	6.29	−24.85*
Central Africa						
MBU	13[f]	5 (38.5)	19 (6.8)	0.756±0.097	7.13	3.76
BIA	17	8 (47.1)	20 (7.2)	0.890±0.043	7.81	1.67
BIO	45	16 (35.6)	30 (10.9)	0.910±0.020	7.09	−0.55
STM	50	32 (64.0)	46 (16.7)	0.973±0.011	7.86	−14.52*
ANG	30	16 (53.5)	36 (13.0)	0.913±0.037	8.74	−1.34
FOR	35	29 (82.9)	41 (14.9)	0.988±0.010	7.52	−18.66*
TON	38	28 (73.7)	52 (18.8)	0.984±0.009	8.55	−12.61*
ST	103	61 (59.2)	63 (22.8)	0.985±0.004	8.36	−24.68*
East Africa						
TK	36	32 (88.9)	54 (19.6)	0.991±0.010	9.66	−20.75*
SO	27	24 (88.9)	41 (14.9)	0.992±0.013	6.90	−16.25*
KIK	25	23 (92.0)	45 (16.3)	0.993±0.013	7.96	−14.55*
NUB	80	50 (62.5)	64 (23.2)	0.974±0.008	7.88	−24.83*
SUD	76	63 (82.9)	73 (26.4)	0.993±0.004	8.33	−24.77*
MOZ	109	49 (45.0)	57 (20.6)	0.960±0.008	7.78	−23.62*
South Africa						
KNG1	24[f]	9 (37.5)	16 (5.8)	0.830±0.053	2.97	−1.27
KNG2	43	12 (27.9)	31 (11.2)	0.812±0.045	7.30	1.84
KWE	31	10(32.30)	34 (12.3)	0.884±0.028	8.75	3.00

[a] N = sample size
[b] K = number of different sequences and percentage of sample size (N) in brackets
[c] S = number of segregating sites and percentage of all sites in brackets. l = sequence length
[d] H = sequence diversity ± standard error
[e] M = average number of pairwise differences
[f] some sequences were not considered for this analysis since there were many positions not scored.
* $P < 0.05$.

the range of values generally found in Africa. Moreover, the broad mtDNA haplogroup profile detected in São Tomé e Príncipe, as reflected in the plot derived from the PC analysis, indicates that the maternal pool of São Tomé e Príncipe primarily arose from a Central/Southwestern African substratum, shaded with signs that denote the archipelago intricate settlement history.

It is noteworthy that despite the insularity, São Tomé e Príncipe maintained a substantial mtDNA diversity, suggesting that the present-day repertoire of maternal lineages is not the result of long-term internal evolution

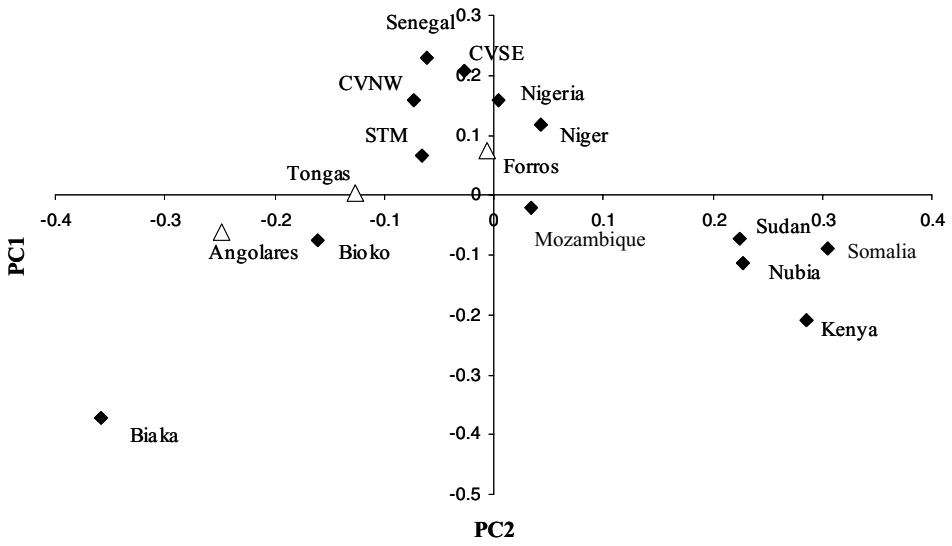

Figure 5 Plot of the two principal components of haplogroup frequency profiles for African samples. CVSE - Cabo Verde SE; CVNW - Cabo Verde NW; STM – São Tomé (Mateu *et al.* 1997).

and rather can be explained by the influx of lineages originating in considerably disperse regions, which were carried by the different people who during the last five centuries settled in the archipelago. Among them were the European colonists, mainly from Portugal, which, particularly during the first centuries, played a major role in the peopling of the archipelago and in the development of the political, social and economic structures. Notwithstanding, the European impact seems to have been virtually nil in the mtDNA pool. Without exception, the detected mtDNA haplotypes belong to clusters within a geographic distribution restricted to Sub-Saharan Africa, testifying to a full matrilineal African background. This finding contrasts with that find using other genomic regions. In a previous study (Trovoada *et al.* 2001) we detected some Y-chromosome haplotypes (defined by microsatellite loci) of likely European ancestry. Recently, Tomás *et al.* (2002), using autosomal markers, estimated the European genetic contribution in São Tomé e Príncipe as 10.7%, although the value dropped to 6.5% when individuals with recent European ancestry were removed from the analysis.

On the whole, these results indicate that the European component of the present São Tomé e Príncipe population was essentially mediated by males, signalling therefore the sex-biased mating pattern that has prevailed in the archipelago and that was common in colonial-based systems.

Genetic Relationships Among Angolares, Forros and Tongas

As pointed out by several indices, namely the proportion of different haplotypes, gene diversity and Fu's Fs, Angolares have a clear reduction in mtDNA diversity in comparison with Tongas and Forros. The latter two groups basically share the same pattern of mtDNA variation, and do not manifest any signs of genetic heterogeneity. In contrast, when Angolares were compared to them both, they displayed a slight, but statistically significant, genetic differentiation. These relationships are also perceptible in the orthogonal projection of the two first components from the PC analysis, where Angolares are positioned considerably distant to Tongas or Forros.

The results here obtained with mtDNA are in full agreement with those previously derived from Y-STRs haplotypes (Trovoada *et al.* 2001), and reinforce the depiction of genetic affinities between the three São Tomean ethnic groups as previously outlined.

The absence of genetic heterogeneity between Forros and Tongas for the maternal and paternal pools is easily explainable. Forros are considered to be "the sons of the land," the more ancient African inhabitants of the archipelago, supposedly descended from the freedman slaves who throughout the centuries

strengthened an African *elite* that also had a fundamental role in the development of the social and economic organization of the archipelago (Tenreiro, 1961; Henriques, 2000). The Tongas, as they are now known, are descendent from the African contract labourers who, from the middle of the 19th century onwards, entered the archipelago in mass. If initially they represented an economically and socially disfavoured stratum, nowadays they are fully integrated into the SãoTomean society, and their perception as a group is a social construction that is progressively disappearing. As inferred from the mtDNA and Y-chromosome STRs, Tongas and Forros can be considered a homogeneous entity.

The observed reduced diversity in the maternal pool of Angolares can be attributable to genetic drift since, until rather recently, Angolares lived in relative isolation with scarce contacts with other São Tomé e Príncipe inhabitants. The origin of this ethnic group in São Tomé e Príncipe is still uncertain. The oldest ethnohistorical evidence, referring to Angolares as a distinctive cultural and social group living in relative isolation, suggests that they could be the descendents of slaves saved from a shipwreck that occurred as early as the middle of the 16th century (Costa, 1982; Ambrósio, 1984; Henriques, 2000). However, since this episode does not have any documented support it is far from being accepted as a historical fact. What was frequently, and very early, officially reported was the existence of many fugitive slaves from the sugar-mills, who escaped colonial control by taking refuge in the most inaccessible forest of the Southeastern region of São Tomé (Seibert, 1998). The archipelago always faced strong social conflicts, including many slave uprisings that greatly destabilized the plantation economy. Such rebellions resulted in a growing number of self-liberated Africans in the Southeastern tip of São Tomé (Caldeira, 1999). There this community, likely the foundation core of Angolares, set up a *de facto* African type of land ownership and the situation was soon tacitly recognised by the Portuguese. The Angolares were left alone and lived in relative isolation for a long time. Nowadays, in spite of tending to be absorbed into the social and economic structure of the island, Angolares still retain a strong group identity that stems from their distinctive social and cultural traditions, occupational activities, geographical confinement and even language. In fact, Angolares speak a more clearly perceived different creole among the three spoken in the archipelago: Sãotomense (in most of São Tomé island), Principense (in Príncipe, now nearly extinct) and Angolar, spoken in the southern tip of São Tomé. The substratum of the three creoles was largely Kwa and Western Bantu languages, impregnated with ancient and modern Portuguese. Angolar shares 70% lexical similarity with Sãotomense and 67% with Principense, and the non-shared lexicon fraction is largely of Bantu origin (http://www.ethnologue.com).

If the linguistic evidence seems to favour the second scenario described for the origin of Angolares (Seibert, 1998), also the phylogeographical analysis of their mtDNA lineages points in the same direction. Within the group the most highly represented haplogroup was L1c, a haplogroup with a geographical distribution that has led to the suggestion that the origin of the cluster could be in the still uncharacterised areas of Angola and the Congo delta (Salas *et al*. 2002). These regions are well documented to have been crucial in Portuguese slave trade activities (Pinto & Carreira, 1979; Russel-Wood, 1995; Henriques, 2000). The representation of L1c, now detected in the São Tomean sample, seems to reinforce the proposed scenario for the origin of L1c, and thus the very same area seems to have been a major source of Angolares ancestors. Yet, the spectrum of the Angolares' L1c lineages is quite diverse, which indicates either some scattering of the source area, or that this haplogroup was quite diverse in the source population.

The second most predominant haplogroup was L3e1*, with the representative lineage fully shared by eight individuals, denoting a clear founder and/or drift effect. Again the hypothetical place of origin is an area phylogeographically very incipiently studied, that coincides with the broad recruitment belt of the slaves brought to São Tomé.

The other less frequent lineages found in Angolares are either widespread in Africa or typically from the central/western fringe, except for one L1a1a lineage which has a predominantly East/Southeaster African distribution, and might indeed represent a more recent introduction within Angolares. It seems then that

later inputs did not leave major footprints in the Angolares' current mtDNA pool. Moreover, the repertory of mtDNA types is quite diverse (as reflected for instance in the mean number of pairwise differences between sequences) and taken together these findings seem to be rather compatible with the Angolares being the descendents of fugitives slaves, who with time consolidated a community who lived in relative isolation from other São Tomé e Príncipe inhabitants.

Pattern of Sex Mediated Gene Flow Among Groups

Although the overall FST retrieved with AMOVA - considering Angolares, Forros and Tongas as a single group - was statistically significant (overall FST $= 0.026$, $P = 0.02$), when the analysis was executed with hierarchical grouping no significant between-group differences were detected, suggesting therefore no strong restrictions in inter-group female gene flow. At least two structural population features might have accounted for that. First is the strong tendency towards a polygamous mating system in São Tomé e Príncipe. Across the archipelago, polygyny is a common mode of family organisation (Caldeira, 1999) and corresponds to the retention of the traditional system from the mainland African populations that contributed to the peopling of the islands. As in most polygynic societies, and in São Tomé e Príncipe, the practice is associated with individual economic prosperity (Caldeira, 1999). The high level of polygyny could have contributed to a comparatively high effective population size of females, explaining therefore the absence of a clear-cut structure detected in the female pool.

Secondly, paralleling the former effect but specifically concerning the Angolares, is their long held practice of looking for women outside the community. Most likely the first fugitive slaves were male. From the refugee area, after having achieved a minimal group organization, Angolares often led assaults on the sugar-plantations and kidnapped African women, a practice that was common until the end of the 18th century, indicating therefore that female deficit was a structural problem among the group (Caldeira, 1999). It was also reported that in São Tomé many women voluntarily returned to the plantations. So, the fairly well-documented reduced number of females among Angolares was compensated for by a renewed influx of women and that reshuffling of female lineages might have also have been responsible for the lack of a strong substructuring at the mtDNA pool within the archipelago.

In our previous work based upon Y-STRs markers, a significant proportion of diversity could be attributed to differences between groups when Angolares were compared with Tongas plus Forros. Therefore, the combined information from the maternal and paternal lineages, suggests that a sex-biased gene flow between Angolares and the other São Tomé e Príncipe inhabitants must indeed have occurred, with the migration rate apparently lowered for the male component.

In conclusion, the analysis of mtDNA variation reinforces the depiction of genetic relationships among Angolares, Tongas and Forros already outlined with the study of Y-STRs markers, and also points to the slight genetic differentiation of Angolares relative to Tongas or Forros, with the latter two groups not showing genetic heterogeneity among themselves.

For a small archipelago with such a recent population history, this is really quite a remarkable finding. In São Tomé e Príncipe, a complex net of social and historical forces must have interacted so that a self-reported community like the Angolares, with presumed distinct origin from other São Tomé e Príncipe inhabitants, has retained and still preserves signs of genetic distinctiveness, despite sharing such restricted territory with other people.

Acknowledgments

We thank the blood donors and the community leaders who made this study possible. We are grateful to Dr António Lima, Dra Julieta Espírito Santo, Dr. Carlos Vera Cruz, Dr. Leonel Pontes, the health directors and the laboratory technicians who provided the facilities for sample collection. This work was partially supported by IPATIMUP through Programa Operational Ciência, Tecnologia e Inovação (POCTI), Quadro Comunitário de Apoio III.

Appendix Control region sequences in Angolares, Forros and Tongas

Angolares	Forros	Tongas	HVS-I	HVS-II	Haplogroup
	1		148 172 187 188$^{C/G}$ 189 223 230 311 320	93 152 189 204 207 236 247 263 315.1	L1a
		1	129 148 168 172 187 188$^{C/G}$ 189 209 213 223 230*	93 185 189 236 247 263 309.1 315.1	L1a1
	1		93 129 148 168 172 187 188$^{C/G}$ 189 223 230 278 293 311 320	93 95$^{A/C}$ 185 189 236 247 263 315.1	L1a1a
		2	129 148 168 172 187 188$^{C/G}$ 189 223 230 278 293 311 320	93 95$^{A/C}$ 185 189 236 247 263 309.1 315.1 320	L1a1a
1	1		129 148 168 172 187 188$^{C/G}$ 189 223 230 311 320	93 185 189 236 247 263 315.1	L1a1
	2	1	126 187 189 223 264 270 278 311	73 152 182 185$^{G/T}$ 195 247 263 315.1 357	L1b
3	1		111 126 187 189 223 239 270 278 293 311	73 146 152 182 185$^{G/T}$ 189 247 263 315.1 357	L1b1
1			111 187 189 223 239 270 278 293 311	73 146 152 182 185$^{G/T}$ 189 247 263 315.1 357	L1b1
		1	114$^{C/A}$ 126 189 223 264 270 274 278 293 311	73 152 182 185$^{G/T}$ 195 247 263 309.1 315.1 357	L1b1
1			114$^{C/G}$ 126 187 189 223 264 270 278 293 311	73 152 182 185$^{G/T}$ 195 247 263 315.1 357	L1b1
	1		126 187 189 223 264 270 278 293 311	73 152 182 185$^{G/T}$ 189 195 247 263 309.1 315.1 357	L1b1
	2		126 187 189 223 264 270 278 293 311	73 152 182 185$^{G/T}$ 189 195 247^{DelG} 263 309.1 315.1 357	L1b1
		1	129 163 187 189 209 223 278 293 294 298 311 360	73 152 182 186$^{C/A}$ 189$^{A/C}$ 194 198 247^{DelG} 263 309.1 315.1 316	L1c1
		1	129 163 187 189 223 278 293 294 304 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 247 263 315.1 316	L1c1
		1	129 163 187 189 223 278 293 294 304 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 315.1 316	L1c1
1			129 187 189 223 278 293 294 311 360	73 152 186$^{C/A}$ 189 195 247 250$^{T/G}$*	L1c1
	1		93 129 187 189 223 263 278 293 294 311 360 368	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 247 263 297 315.1 316	L1c1
		1	93 129 187 189 223 263 278 293 294 311 360 368	73 182 186$^{C/A}$ 189$^{A/C}$ 195 247 263 297 315.1 316	L1c1
2			129 187 189 223 274 278 293 294 311 360	73 93 95$^{A/C}$ 152 182 186$^{C/A}$ 189$^{A/C}$ 195 236 247 263 297 315.1 316	L1c1a
1			129 187 189 223 274 278 293 294 311 360	73 95$^{A/C}$ 152 182 186$^{C/A}$ 189$^{A/C}$ 195 236 247 263 297 315.1 316	L1c1a
1			129 187 189 214 234 249 258 274 278 293 294 311	73 151 152 186$^{C/A}$ 189$^{A/C}$ 195 247 263 297 315.1 316	L1c1a1
1			51 129 187 189 214 234 249 258 274 278 293 294 311	73 186$^{C/A}$ 189$^{A/C}$ 195 247 263 297 315.1 316	L1c1a1
1			129 163 187 189 265$^{A/C}$ 278 286$^{C/G}$ 294 311 320 360	73 151 152 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 297 315.1 316	L1c2
		1	129 187 189 214 223 265$^{A/C}$ 278 286$^{C/A}$ 291 294 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 297 315.1 316	L1c2
		2	129 187 189 223 265$^{A/C}$ 278 286$^{C/G}$ 294 311 343$^{A/T}$ 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 297 315.1 316	L1c2
3			129 187 189 223 265$^{A/C}$ 286$^{C/A}$ 292 294 311 360	73 152 182 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 297 309.1 315.1 316	L1c2
	1		172 187 189 223 265$^{A/C}$ 278 286$^{C/G}$ 294 311 360	73 151 152 182 186$^{C/A}$ 189$^{A/C}$ 195 198 247 263 297 309.1 315.1 316	L1c2
		2	129 183$^{A/C}$ 189 215 223 278 294 311 360	73 152 182 186$^{C/A}$ 189$^{A/C}$ 247 263 309.1 315.1 316	L1c3
		1	(42) 129 166 187 189 223 254 278 311	73 146 152 182 195 198 247 263 315.1	L1e
	1		111 223 278 294 309 311 390	73 152 195 315.1	L2a1
	1		129 223 278 294 309 390	73 143 146 152 195 263 315.1	L2a1
1			189 192 223 239$^{C/A}$ 278 279$^{C/A}$ 294 309*	73 195 263 315.1	L2a1
	1		189 192 223 278 294 309 390	73 143 146 152 195 263 309.1 315.1	L2a1
		2	189 192 223 278 294 309 390	73 146 152 195 263 309.1 315.1	L2a1
	1		193 213 223 239 278 294 309 390	73 146 152 195 263 309 309.1 315.1	L2a1
		1	223 278 294 309 390	73 143 146 152 195 198 263 309.1 315.1	L2a1
	2		223 278 294 309 390	73 146 152 195 263 309.1 315.1	L2a1
		2	223 278 286 294 309 390	73 146 152 195 263 309.1 315.1	L2a1a
		2	114$^{C/A}$ 129 213 223 278 355 362 390	73 150 152 182 195 198 204 249^{DelA} 263 309.1 315.1	L2b1

Appendix (*Continued*)

Angolares	Forros	Tongas	HVS-I	HVS-II	Haplogroup
		1	114$^{C/A}$ 213 223 278 362 390	73 150 152 182 195 198 204 263 315.1	L2b1
	2		81 93 175 223 278 320 390	73 146 150 182 195 198 263 309.1 315.1 325	L2c
1			81 93 175 223 278 320 390	73 93 146 150 182 195 198 263 309.1 315.1 325	L2c
	1		114 223 278 318 390	73 93 146 150 152 182 195 198 263 309.1 315.1 325	L2c1
1			223 264 278 390	73 146 150 152 182 195 198 263 309.1 315.1 325	L2c2
		1	84 93 220 223 264 278 311 390	73 93 146 150 152 182 195 198 263 315.1 325	L2c2
1			93 223 264 278 390	73 93 146 150 152 182 195 198 263 309.1 315.1 325	L2c2
		1	111$^{C/A}$ 145 184 223 239 278 292 355 390	73 146 150 152 182 263 315.1	L2d2
	1	1	223 290 355	73 150 152 235 263 309.1 315.1	L3*
	1		223 278 362	73 263 315.1	L3b
		1	124 183$^{A/C}$ 189 214 223 278	73 263 309.1 315.1	L3b1
	2		124 223 278 362	73 263 315.1	L3b1
	1		124 223 278 311 362	73 257 263 309.1 315.1	L3b2
	1		124 223 311	73 101$^{G/C}$ 150 152 263 309.1 315.1	L3d
		3	124 223 311	73 152 200 263 309.1 315.1	L3d
		1	124 209 223 319 362	73 152 263 315.1 (385)	L3d1
		1	124 223 300 319	73 152 263 315.1	L3d1
		1	104 129 183$^{A/C}$ 189 223 260 327	73 150 263 309.1 315.1	L3e1*
8			172 223 327 (399)	73 150 189 200 207 263 309.1 315.1	L3e1*
	1		207 223 327	73 150 183 189 200 263 309.1 315.1	L3e1*
	1		223 325^{DelT} 327	73 150 185 189 209 263 309.1 315.1	L3e1*
	1		185 209 223 327	73 150 152 189 195 200 263 309.1 315.1	L3e1a
1			223 258$^{A/T}$ 320	73 150 189 195 315.1	L3e2
	1		223 320	73 150 195 198 263 309.1 315.1	L3e2
1		2	223 320	73 150 195 198 263 315.1	L3e2
	1		126 172 182$^{A/C}$ 183$^{A/C}$ 189 223 320	73 150 195 263 315.1	L3e2b
	1	1	223 265$^{A/T}$	73 150 195 263 315.1	L3e3
		1	51 223 264	73 150 263 309.1 315.1	L3e4
		1	51 223 264 299	73 150 263 309.1 315.1	L3e4
	1		129 209 223 311	73 189 200 263 315.1	L3f
	1		209 223 311	73 150 189 200 263 309.1 315.1	L3f
	1		209 215 223 256 292 311	73 185 189 195 263 309.1 315.1	L3f1

The transversions and the deletions are specified by the appropriate changed or deleted bases respectively. ".1" indicates the insertion of a C. The additional positions outside the range considered for the analysis, are indicated inside brackets. * Incomplete sequences.

References

Ambrósio, A. (1984) *Subsídios para a História de São Tomé e Príncipe*. Livros do Horizonte.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–65.

Bandelt, H.-J., Alves-Silva, J., Guimarães, P. E., Santos, M. S., Brehm, A., Pereira, L., Coppa, A., Larruga, J. M., Rengo, C., Scozzari, R., Torroni, A., Prata, M. J., Amorim, A., Prado, V. F. & Pena, S. D. (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann Hum Genet* **65**, 549–63.

Bandelt, H.-J., Quintana-Murci, L., Salas, A. & Macaulay, V. (2002) The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet* **71**, 1150–1160.

Brehm, A., Pereira, L., Bandelt, H. J., Prata, M. J. & Amorim, A. (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann Hum Genet* **66**, 49–60.

Caldeira, A. M. (1999) *Mulheres, sexualidade e casamento em S.Tomé e Príncipe (Séculos XV-XVIII)*. Lisboa. Edições Cosmos.

Chen, Y.-S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. (2000) mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am J Hum Genet* **66**, 1362–83.

Côrte-Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B. C. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331–50.

Costa, F. F. (1982) A ilha de S.Tomé. Um reino de escravos na linha do Equador. *História* **50**, 66–78.

Henriques, I. C. (2000) *São Tomé e Príncipe. A invenção de uma sociedade*. Vega e Autor.

Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. & Excoffier, L. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* **12**, 334–345.

Krings, M., Salem, A. E., Bauer, K., Geisert, H., Malek, A. K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., Sajantila, A., Pääbo, S. & Stoneking, M. (1999) mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet* **64**, 1166–76.

Lareu, M. V., Phillips, C. P., Carracedo, A., Lincoln, P. J., Syndercombe Court, D. & Thomson, J. A. (1994) Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic Sci Int* **66**, 41–52.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B. & Torroni, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–49.

Mateu, E., Comas, D., Calafell, F., Perez-Lezaun, A., Abade, A. & Bertranpetit, J. (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* **61**, 507–18.

Newman, J. L. (1995) *The peopling of Africa: a geographic interpretation*. Yale University Press.

Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J. & Amorim, A. (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* **65**, 439–58.

Pinto, F. L. V. & Carreira, A. (1979) Portuguese participation in the slave trade: opposing forces, trends of opinion within Portuguese society: effects on Portugal's socio-economic development. In: *The African slave trade from the fifteenth to the nineteenth century. The general history of Africa. Studies and documents 2*. pp. 119–147. Unesco.

Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H. J. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* **62**, 531–50.

Richards, M. & Macaulay, V. (2000) *Genetic data and the colonization of Europe: genealogies and founders*: In: Renfrew, C. & Boyle, K. Archaeogenetics: DNA and the population Prehistory of Europe. Cambridge: McDonald Institute for Archaeological Research, pp. 139–151.

Russell-Wood, A. J. R. (1995) *The Portuguese Empire, 1415–1808. A world on the move*. The Johns Hopkins University Press.

Salas, A., Richards, M., De la Fe, T., Lareu, M. V., Sobrino, B., Sánchez-Diz, P., Macaulay, V. & Carracedo, A. (2002) The making of the African mtDNA landscape. *Am J Hum Genet* **71**, 1082–111.

Schneider, S., Roessi, D. & Excoffier, L. (2000) *Arlequin ver. 2000: A software for population genetics data analysis*. Genetics and Biometry laboratory, University of Geneva. Switzerland.

Seibert, G. (1998) *A Questão da Origem dos Angolares de São Tomé*. Brief Papers n°5/98, CEsA, Lisboa.

Tenreiro, F. (1961) *A ilha de São Tomé*. Memórias da Junta de Investigações do Ultramar. Lisboa.

Tomás, G., Seco, L., Seixas, S., Faustino, P., Lavinha, J. & Rocha, J. (2002) The peopling of São Tomé (Gulf of Guinea): origins of slave settlers and admixture with the Portuguese. *Hum Biol* **74**, 397–411.

Trovoada, M. J., Alves, C., Gusmão, L., Abade, A., Amorim, A. & Prata, M. J. (2001) Evidence for population substructuring in São Tomé e Príncipe as inferred from Y-chromosome STR analysis. *Ann Hum Genet* **65**, 271–83.

Valverde, E., Cabrero, C. & Cao, R. (1993) Population genetics of three VNTR polymorphisms in two different Spanish populations. *Int J Legal Med* **151**, 251–256.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) African populations and the evolution of human mitochondrial DNA. *Science*, **253**, 1503–7.

Watson, E., Forster, P., Richards, M. & Bandelt, H. J. (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* **61**, 691–704.

Received: 30 April 2003
Accepted: 21 August 2003

