Ann. Hum. Genet. (2001), **65**, 271–283
Printed in Great Britain

Evidence for population sub-structuring in São Tomé e Príncipe as inferred from Y-chromosome STR analysis

M. J. TROVOADA[1,2], C. ALVES[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/nº, 4200 Porto, Portugal*
[3] *Faculdade de Ciências da Universidade do Porto, Pr. Gomes Teixeira, 4050 Porto, Portugal*

(Received 4.10.00. Accepted 1.2.01)

SUMMARY

Seven Y-chromosome STR loci, DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393 have been analysed in population samples of Angolares, Forros and Tongas, three ethnic groups from the African archipelago of São Tomé e Príncipe (Gulf of Guinea). Complete typings were obtained for 103 chromosomes, which belonged to 79 different haplotypes. The mean heterozygosity per locus in the overall São Tomean sample was 0.566, with the highest value found among Forros and the lowest among Angolares. Angolares also showed the lowest level of haplotype diversity. On average, the mean pairwise difference between two random haplotypes from Angolares, Forros and Tongas was 4.69, 6.74 and 6.23 repeats, respectively. The genetic distances were found to be statistically significant between Angolares and Forros or Tongas. In accordance, AMOVA revealed that the percentage of variation attributable to differences among groups was only significant when we distinguished between Angolares and non-Angolares. Globally, these results indicate that, with respect to the pool of male lineages of São Tomé e Príncipe, some genetic sub-structuring does exist, basically determined by the Angolares ethnic group.

INTRODUCTION

Small human populations, whose origins trace back to very recent times, can be used as helpful models in microevolution studies. With the added benefit of written records, it may be possible to recover the main demographic events which underlie a population history, and to assess how different evolutionary factors have interacted in modelling the genetic structure of present-day populations.

In this field, analysis of Y-chromosome polymorphisms has been widely proven to be a useful investigative tool (Kayser *et al.* 1997*b*; Perez-Lezaun *et al.* 1997; Hammer *et al.* 1998; Ruiz-Linares *et al.* 1999; Jobling & Tyler-Smith, 2000). The number and kind of informative Y-chromosome loci has progressively increased over the last few years (Jobling & Tyler-Smith, 1995; Hammer *et al.* 1997; Underhill *et al.* 1997). Among them the STRs of the non-recombining region represent a class characterized by high evolution rates. Their study allows the construction of haplotypes at multiple polymorphic loci, which have been used for tracing the origins of paternal lineages in human populations, either at large or restricted geographical scales (Bravi *et al.* 1997; Malaspina *et al.* 1998; Hurles *et al.* 1998, 1999; Bosch *et al.* 1999; Santos *et al.* 1999).

In this study we have surveyed the variation at seven Y-chromosome STR loci in three

Correspondence: Maria Jesus Trovoada, Departamento de Antropologia da Universidade de Coimbra, 3049 Coimbra Codex, Portugal. Tel: +351 239829051; Fax: +351 239823491.
E-mail: mjesus@ci.uc.pt

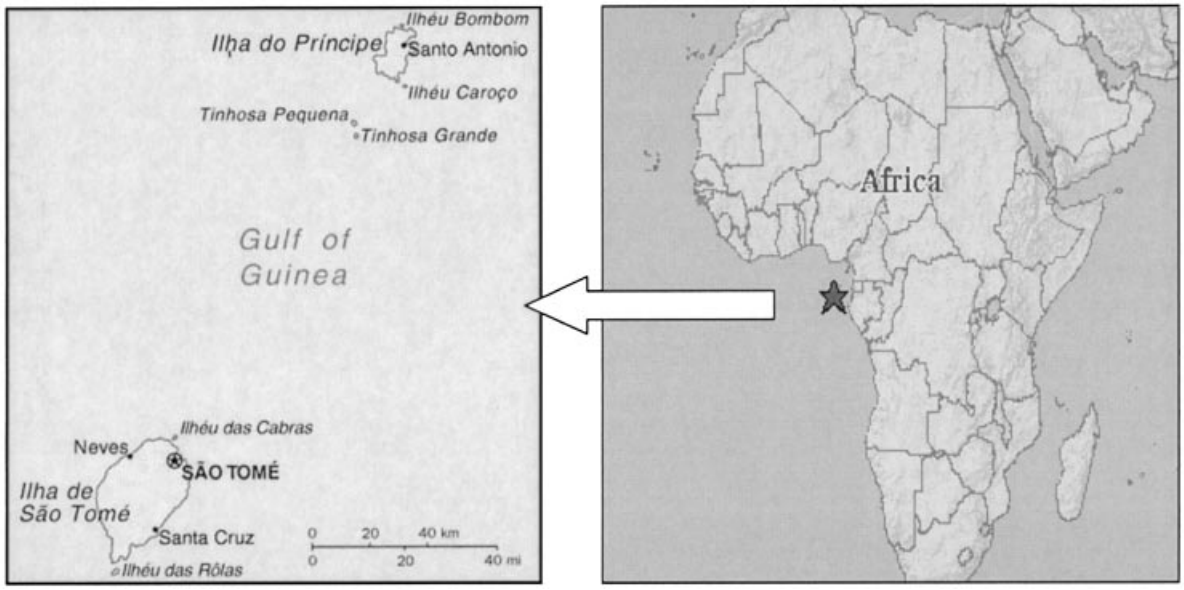

Fig. 1. Map illustrating the geographical location of São Tomé e Príncipe.

population groups from São Tomé e Príncipe. This, the second smallest country in Africa, is an archipelago situated in the Gulf of Guinea, 300 km from Gabon (Fig. 1). It comprises two islands and several small islets with a combined area of barely 1000 sq. km and a total population of approximately 150000 inhabitants.

Most historical sources agree that both islands were not inhabited when first discovered by the Portuguese in the early 1470s. Soon afterwards settlement began with European colonists, mainly Portuguese, and slaves coming originally from the western mainland coast of Africa. The complex process of peopling of São Tomé e Príncipe was strongly influenced by two particular factors: a rather contentious scheme of land occupation, and the rise and fall, along centuries, of different economic activities (Henriques, 2000).

The existence of three population groups, Angolares, Forros and Tongas, is the common perception among present day São Tomé e Príncipe inhabitants and is referred to in the ethnohistorical bibliography of the archipelago (Tenreiro, 1961; Almeida, 1966; Pinto & Carreira, 1979; Ambrósio, 1984). However, ignoring territorial occupation and social differences, the three groups are not easily identifiable, although the supposed distinctiveness of Angolares has additional support from the specificity of their geography, dialect, and even some reported morphological traits (Costa, 1982).

To date, very few studies have focused on the genetic characterization of São Tomé e Príncipe, including a very recent work based on the analysis of mtDNA variation (Mateu *et al.* 1997), which did not aim to investigate its internal population structuring. Here we report data on a characterization of Y-chromosome STRs in population samples representative of the three São Tomean groups. This was performed in order to address the main issue of whether the inferences based on the genetic analysis would be consistent with the traditionally accepted and ethnohistorically documented sub-structuring of the population of São Tomé e Príncipe.

Material and methods

Population samples and data analysed

Blood samples, collected by venipuncture into EDTA tubes, were obtained with informed consent from unrelated males born and living in São Tomé e Príncipe. They comprised 21 Angolares, 38 Forros and 44 Tongas.

A reference data set was obtained from published information for two Portuguese population samples and for Angola, Cabo Verde, Guinea-Bissau and Mozambique, four African countries that were former Portuguese colonies.

Experimental procedures

DNA was extracted from whole blood using a standard phenol-chloroform methodology (Valverde *et al.* 1993). Seven Y-chromosome STR loci were analysed by PCR-based procedures: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393.

A pentaplex system was used to co-amplify DYS19, DYS389I, DYS389II, DYS390 and DYS393 according to the conditions described in Gusmão *et al.* (1999). For DYS391 and DYS392 a duplex amplification was performed using 10 ng genomic DNA in a 12.5 μl reaction volume containing 1.5 mM $MgCl_2$ buffer, 0.5 U *Taq* DNA polymerase, 200 μM of each dNTP and 0.25 μM of each primer. The PCR conditions were: initial denaturation 94 °C −5 min; 35 cycles at 94 °C −1 min, 56 °C −1 min 30 s, 72 °C −2 min; final extension 72 °C −10 min.

PCR products were electrophoresed on 6% polyacrylamide gels. An ABI PRISM 377 DNA sequencer was used for the automated typings. The analysis of the DNA fragments was carried out by the GeneScan 2.1 Analysis software. Typings were performed using sequenced allelic ladders as references.

Allele nomenclature was as proposed by Kayser *et al.* (1997*a*). Taking into account that the PCR fragment DYS389II contains DYS389I plus an additional locus, all DYS389II alleles were modified by subtracting the size of the corresponding DYS389I allele.

Statistical analysis

Haplotype frequencies, haplotype diversity along with its standard deviation, and gene diversity across loci, were calculated for all populations with the Arlequin ver 2.000 software (Schneider *et al.* 2000).

Several measures of population differentiation were computed, but since all gave essentially the same results only those based on FST values were considered. The molecular genetic distances derived from haplotypes were computed considering the sum of squared size differences.

The FST values among populations were used to draw a neighbour-joining tree with the Neighbour routine of the PHYLIP 3.5c software (Felsenstein, 1993) and the tree was further visualized with the Treeview software (Page, 1996).

Computations of pairwise differences between haplotypes were calculated by counting each change of one microsatellite unit as one difference.

RESULTS

Informativeness locus by locus

Table 1 shows the Y-chromosome haplotypes and their absolute frequencies in the three study populations of São Tomé e Príncipe. Overall, the distribution of allele frequencies registered for each of the seven STRs fits well into the patterns reported for African populations (de Knijff *et al.* 1997).

The mean heterozygosity per locus in the overall S. Tomean sample was 0.566, with the lowest value found among Angolares and the highest among Forros (Table 2). The amount of variation detected is very similar to that registered by Pritchard *et al.* (1999) for African populations, 0.620, in a study based upon the analysis of the seven STRs used in this work plus DYS388 (this locus was eliminated, and new statistics were recalculated for the comparisons). However, the mean variance across loci in the number of repeat units detected in São Tomé e Príncipe (0.886) was rather low for an African population compared with the mean variance in repeat score registered by Pritchard *et al.* (1999) for Africans of 1.135 (based upon the 7 STRs).

For each STR, FST values were computed for the three populations under study, and these are presented in Table 3. Except for DYS390, where a significant FST was detected, the remaining values as well as the mean FST value over all the 7 loci (0.006) were very low, denoting the absence

Table 1. *Y-chromosome STR haplotypes found in Angolares, Forros and Tongas and number of matches in the reference populations, in the Caucasian YSTR Database* (http://ystr.charite.de) – [**Ca*] – *and in the African data set of Pritchard et al. 2000* (http://www.stats.ox.ac.uk/) – [**Af*]

	DYS loci																	
Haplotypes	19	389I	389II	390	391	392	393	An	Fo	To	AG	CV	GB	Mo	PC	PN	*Af	*Ca
H1	12	10	18	21	11	11	13	1										
H2	13	9	17	24	10	11	13			1		3				2		5
H3	13	10	17	21	11	13	13			1								
H4	13	10	17	24	10	11	13			1								40
H5	13	10	17	24	11	11	13			1		1						5
H6	13	10	18	21	10	11	13	1										
H7	13	10	18	21	11	11	13		1								1	
H8	13	10	18	24	11	11	12		1									
H9	14	9	16	24	10	13	13			1								8
H10	14	9	16	24	11	13	12		1									2
H11	14	9	16	25	10	14	13		1									3
H12	14	9	17	22	10	11	14	1										5
H13	14	9	17	22	11	11	14	2										
H14	14	9	17	22	11	14	14	1										
H15	14	10	16	24	10	14	13		1							1		8
H16	14	10	16	24	11	12	13		1						1	1		7
H17	14	10	16	24	11	13	13			1		4			4	5		191
H18	14	10	17	21	10	13	14		1									
H19	14	10	18	24	10	11	13			1							5	4
H20	14	11	17	23	9	11	13			1		1						
H21	14	11	17	23	10	11	12			1								
H22	14	11	17	23	10	15	12			1								7
H23	14	11	17	24	11	13	13		1						1	1		4
H24	14	11	17	25	10	14	13		1									
H25	15	8	15	21	10	13	13			1								
H26	15	8	15	23	10	12	13			2								
H27	15	9	15	21	11	11	13			1								
H28	15	9	16	21	10	11	13			1								1
H29	15	9	17	21	10	11	14	1			1	1						1
H30	15	9	18	23	12	11	14		1									
H31	15	9	19	21	10	11	13		1									
H32	15	10	16	21	10	11	14	1										
H33	15	10	16	22	10	11	14			1								1
H34	15	10	16	24	10	11	14			1								1
H35	15	10	16	24	10	13	13		1						2	2		22
H36	15	10	17	21	10	10	13	1										
H37	15	10	17	21	10	11	12		1									1
H38	15	10	17	21	10	11	13		1		5			1			2	
H39	15	10	17	21	10	11	14		1	3	3	5	1	2			1	1
H40	15	10	17	21	10	11	15		1	3			1				3	
H41	15	10	17	21	10	11	17		1									
H42	15	10	17	21	11	11	13			1	2			1			1	
H43	15	10	17	21	12	11	13		1									
H44	15	10	17	22	10	11	15	1										
H45	15	10	17	22	12	11	15	1										
H46	15	10	18	21	10	11	13	1	1	2	7		2	2			4	1
H47	15	10	18	21	10	11	14			1				1			2	1
H48	15	10	18	21	11	11	13	4	3	3	2						6	1
H49	15	10	18	21	11	14	13	1										
H50	15	10	19	21	10	11	13	1	1									
H51	15	10	19	21	11	11	13	1										
H52	15	11	17	21	10	11	13		1	1	1			1			1	1
H53	15	11	17	21	10	11	14			1			2					1
H54	15	11	17	21	10	11	15		1									
H55	15	11	18	25	10	11	13			1							1	
H56	16	9	17	21	10	11	14		1									1
H57	16	9	17	22	11	11	13			1								

Table 1. (*cont.*)

	DYS loci																	
Haplotypes	19	389I	389II	390	391	392	393	An	Fo	To	AG	CV	GB	Mo	PC	PN	*Af	*Ca
H58	16	10	16	21	10	11	15			1	1							
H59	16	10	16	24	10	11	13		1									18
H60	16	10	16	26	11	12	13		1									
H61	16	10	17	21	10	11	13			1			1				1	
H62	16	10	17	21	11	11	14			1	1						1	
H63	16	10	18	20	11	12	13			1								
H64	16	10	18	21	11	11	13	1	1		1							
H65	16	10	18	25	11	11	14		1									3
H66	16	11	15	23	10	11	13			1								
H67	16	11	18	21	10	11	13		1									
H68	17	10	16	21	10	11	15	1										
H69	17	10	16	21	11	11	12			1								
H70	17	10	17	21	10	11	14		1	1			1	3			1	
H71	17	10	17	21	10	12	14			1								
H72	17	10	17	21	11	10	15			1								
H73	17	10	18	21	11	11	13			1								
H74	17	10	18	23	10	14	13			1								
H75	17	11	17	21	10	11	14		2								1	
H76	17	11	17	21	10	11	15		1		1						5	
H77	17	11	17	21	10	12	14		1									
H78	17	11	17	21	11	11	14		1									
H79	18	11	17	21	10	11	14		1									

Abbreviations: An: Angolares; Fo: Forros; To: Tongas; AG: Angola; CV: Cabo Verde; GB: Guinea Bissau; Mo: Mozambique; PC: Central Portugal; PN: North Portugal; Af: African; Ca: Caucasian.

Table 2. *Diversity parameters in São Tomé e Príncipe and reference populations*

Population	Sample size	No. different haplotypes	Haplotype diversity	Gene diversity*	Mean variance in repeat number*
1. Angolares [a]	21	17	0.967 ± 0.030	0.504 ± 0.298	0.554
2. Forros [a]	38	35	0.994 ± 0.008	0.586 ± 0.332	1.043
3. Tongas [a]	44	37	0.991 ± 0.008	0.573 ± 0.324	0.863
S. Tomé (1–3)	**103**	**79**	**0.987 ± 0.006**	**0.566 ± 0.316**	**0.886**
4. Angola [b]	50	32	0.967 ± 0.014	0.455 ± 0.266	0.637
5. Cabo Verde [b]	47	34	0.978 ± 0.011	0.635 ± 0.354	0.925
6. Guinea-Bissau [b]	33	27	0.977 ± 0.018	0.518 ± 0.299	0.497
7. Mozambique [b]	37	30	0.987 ± 0.010	0.568 ± 0.323	1.058
African (1–7)	**270**	**167**	**0.989 ± 0.002**	**0.566 ± 0.314**	**0.866**
African [c]	**229**	**154**	**0.995 ± 0.001**	**0.620 ± 0.340**	**1.135**
8. Portugal C [d]	50	43	0.992 ± 0.007	0.595 ± 0.334	0.685
9. Portugal N [e]	55	39	0.980 ± 0.009	0.517 ± 0.295	0.527
Portugal (8 + 9)	**105**	**77**	**0.990 ± 0.004**	**0.556 ± 0.312**	**0.602**
European [c]	**46**	**37**	**0.985 ± 0.010**	**0.503 ± 0.289**	**0.505**

[a]: this study; [b]: Corte Real *et al.* (2000); [c]: Pritchard *et al.* (1999); [d]: Carvalho *et al.* (2000); [e]: González-Neira *et al.* (2000). *Estimates across loci.

Table 3. *FST values per Y-chromosome locus and across loci*

Population group	DYS loci							Across loci
	19	389I	389II	390	391	392	393	
São Tomean	−0.008	0.026	−0.002	0.042*	0.017	−0.006	−0.020	0.006
African	0.017*	0.016	0.007	0.099*	0.017	0.057*	0.005	0.029*

The São Tomean group includes Angolares, Forros and Tongas; the African one additionally comprises Angola, Cabo Verde, Guinea Bissau and Mozambique. *Significant values at the 5% level.

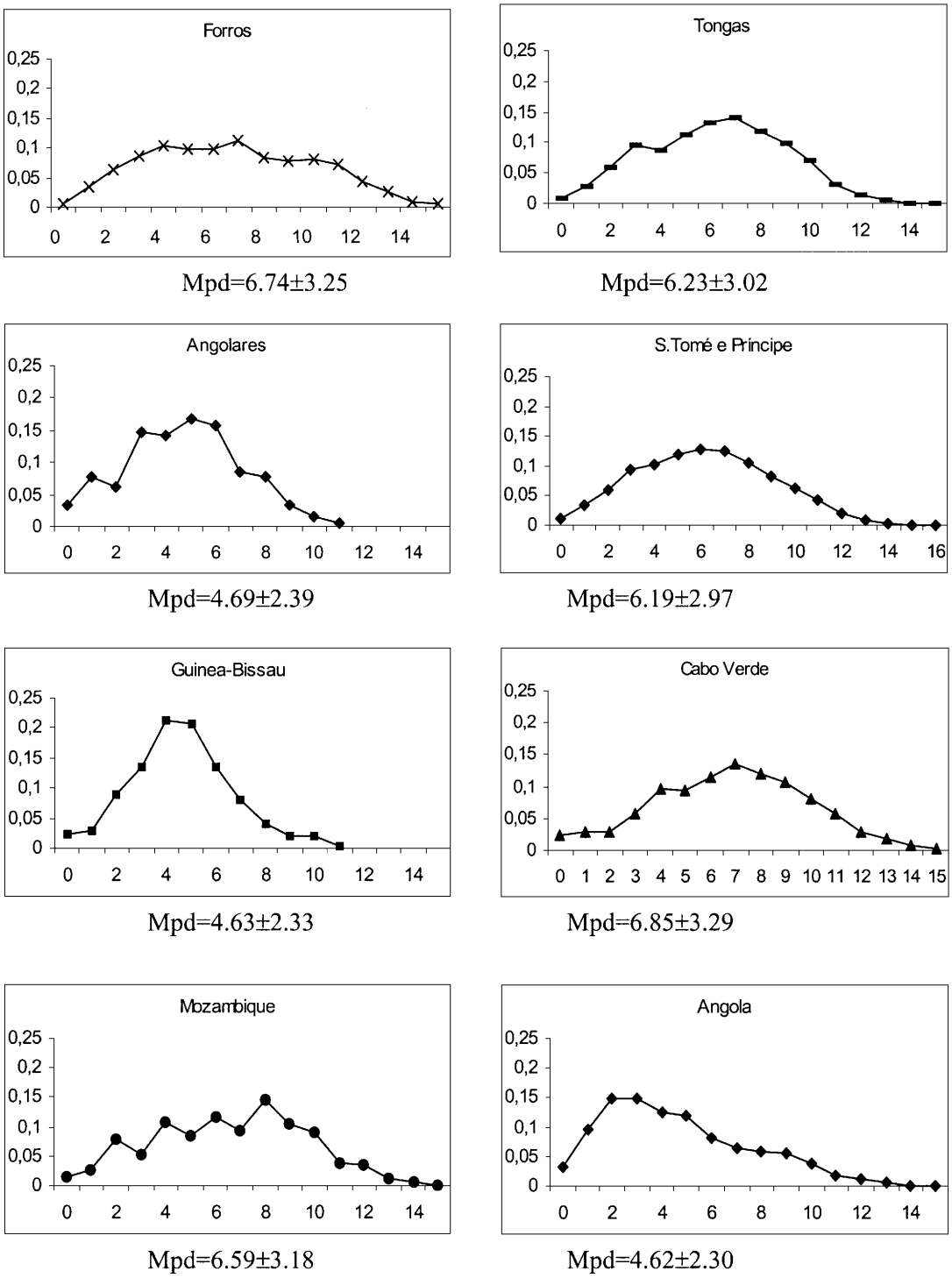

Fig. 2. Distribution of pairwise STR differences and mean pairwise differences (Mpd) within several African populations.

of marked heterogeneity among the São Tomean groups. When the four other African populations (Angola, Cabo Verde, Guinea-Bissau and Mozambique), all former Portuguese colonies, were included in the analysis, the overall FST value obtained was again very low (0.029) and, in spite of being statistically significant, does not reveal high levels of differentiation among the

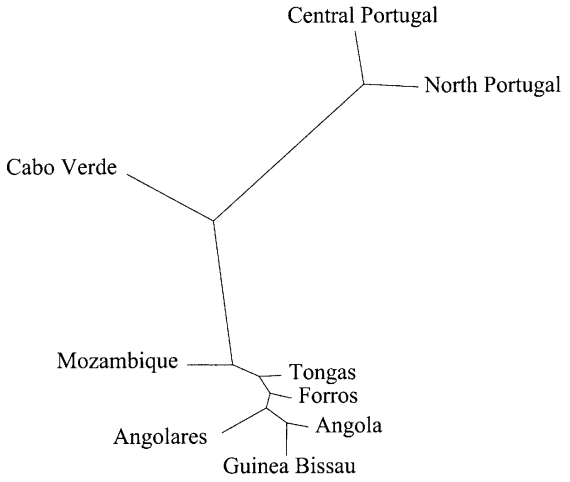

Fig. 3. Neighbour-joining tree of the São Tomean groups and other African and Portuguese populations.

African samples considered. This result is consistent with that obtained by Jorde *et al.* (2000) in a study that has shown that African populations are the least differentiated for Y-chromosome STRs.

STR combined haplotype analysis

In the three São Tomean samples analysed, comprising a total of 103 unrelated males, 79 different combined haplotypes were found (Table 2). The three groups shared two haplotypes, one of them occurring always more than once in each population, which denotes the affinities among the groups. Forros and Tongas shared 6 different haplotypes, Angolares and Forros shared 4, and Angolares and Tongas also 4. The remaining haplotypes were unique to only one of the three populations (Table 1).

Most haplotypes occurred once in each of the three São Tomeans groups, a feature that is reflected in the high values of haplotype diversity observed (Table 2), with Angolares showing the lowest value. Table 2 also shows the equivalent values computed for the 4 reference African populations, as well as for 2 samples from Portugal. Clearly Angolares, together with Angola, lie on the low extreme of the range of values observed.

In order to investigate the pattern of compound haplotype sharing, besides the reference samples already mentioned, two additional sources were consulted; the YSTR reference database (http://ystr.charite.de) that logged 3589 haplotypes from 20 Caucasian populations (mainly European), and the raw data from a worldwide sample of Pritchard *et al.* (1999), made available from http://www.stats.ox.ac.uk/~pritch/data/, that contains 229 African entries. Despite the bias introduced by the strong disproportion between the available data for Caucasian or African male lineages, we pursued the exercise with that in mind. The matches observed for all the haplotypes are indicated in the last two columns of Table 1.

Among the São Tomeans, the Tongas globally share more haplotypes with other populations: 43% of their 37 different haplotypes were shared by Caucasian men, and 49% were present in other African populations. In contrast, Angolares share the lowest number of haplotypes: only 4 out of their 17 haplotypes were found in Caucasians and in Africans, but the fact of being the smallest sample analysed must be taken into account.

Interestingly, of the shared haplotypes, H4 and H17, present in Tongas, matched 40 and 191 times respectively the extended Caucasian database; whereas H35 and H59, present in Forros, matched 22 and 18 times. Two of these haplotypes are also relatively common in North and Central Portugal and it is likely their presence in Tongas or Forros represents Caucasian introgression into the African populations. However it must stressed that it is difficult to assign Y-chromosome STR haplotypes to African or Caucasian populations accurately. As can be seen from Table 1, there are many overlapping haplotypes between Caucasian and African male pools and the pattern of haplotype sharing extends considerably beyond racial or geographical boundaries.

Mismatch distribution of haplotype differences

Pairwise differences across the 7 STR haplotypes were computed for the three São Tomean groups under study and also for the African reference populations (see Fig. 2). The average differences between two haplotypes in Angolares,

Table 4. *Percentage of variation at different levels of population hierarchy produced by AMOVA*

	Variation (%)			
Comparison	To vs An+Fo	Fo vs An+To	An vs To+Fo	An+To+Fo
Within populations	99.98	99.84	97.78	99.36
Between populations	2.67	1.93	−1.02	0.64
Between groups	−2.65	−1.77	3.24*	—

An: Angolares; Fo: Forros; To: Tongas. *Significant values at the 5% level.

Forros and Tongas were 4.69, 6.74 and 6.23 repeats respectively, which means that any random haplotype pair within Angolares is far more similar than within the other São Tomean groups. Among the other African populations, Angola and Guinea-Bissau were also characterized by high levels of relatedness among haplotypes, whilst Cabo Verde was the population where, on average, haplotypes differed the most.

Genetic differentiation among populations

Pairwise genetic distances were computed for the three São Tomean groups and the reference populations. When excluding Cabo Verde, the average distance between African and non-African populations was 7.166, while among Africans it was 0.524. For Cabo Verde the average distances from Africans and Portuguese were very similar: 1.825 and 1.705 respectively. This finding is consistent with the intensive admixture between Europeans and Africans from which the population of Cabo Verde has arisen.

Among the São Tomean groups, the genetic distances were found to be statistically significant between Angolares and Forros or Tongas, whereas between the last two groups no significant genetic heterogeneity was detected (f values of pairwise FSTs were 0.036, 0.018 and 0.472, respectively). Tongas, Forros, Mozambique and Angola did not show significant differences, suggesting a very low level of genetic differentiation among them. The remaining pairwise distances were statistically significant except between North and Central Portugal where the genetic distance was, as expected, extremely low.

A neighbour-joining tree was constructed based upon those distances and is depicted in Fig. 3. The topology of the tree unequivocally differentiates two major groups of populations: African and European. Again, the mixed African and European ancestry of Cabo Verde is reflected in its intermediate position in the tree. Among the African cluster, the positions of Angola, Guinea-Bissau and Angolares in the extremity of the tree seem quite consistent with the fact that all three populations were characterized both by low number of mean pairwise differences and relatively low levels of haplotype diversity.

AMOVA among the São Tomean groups

In order to evaluate more accurately the genetic structure of São Tomé e Príncipe, we have applied AMOVA in 4 different ways: treating Angolares, Forros and Tongas as a single group, or taking one group and comparing it with the other two. The results in Table 4 show that the percentage of variation attributable to differences within populations is extremely high compared to the differences between populations. Between groups, the percentage of variation assumed negative values, except when we distinguish between Angolares and non-Angolares, where a significant fraction of variation due to differences among groups was detected. In order to evaluate the individual locus contribution to this result, we performed a sort of jackknifing, by repeatedly running AMOVA, and each time excluding one or more loci. Non-significant FCT values were only obtained when DYS389I, DYS390 and DYS391 were simultaneously excluded, which leads us to conclude that although the three STRs are the main partial contributors to the overall FCT value, only their combined effect can explain the genetic differentiation observed between Angolares and the other two São Tomean groups.

DISCUSSION

In this study we have analysed a set of 7 Y-chromosome STR loci in three population samples from São Tomé e Príncipe aiming, on a restricted geographical scale, to evaluate the level of genetic sub-structuring within this African archipelago, and in a broader context, to deepen the characterization and understanding of the pattern of Y-chromosome STR variation in African populations.

Population sub-structuring in São Tomé e Príncipe

The pattern of Y-chromosome STR variation in Tongas and Forros was found to be very similar and no signs of genetic heterogeneity were detected among them. Forros are considered to be the descendents of 'the sons of the land': ancient slaves who were set free along the centuries, or even African free-men who also had a fundamental role in the peopling of the islands (Tenreiro, 1961; Henriques, 2000).

The first Africans brought to São Tomé e Príncipe, mainly from Guinea, Benin and Manicongo, soon were joined by slaves from throughout the sub-Saharan mainland. During the initial period of the settlement, miscegenation, especially between African women and European colonists, was instigated in order to increase the population size more rapidly. Several sources report that mulattoes represented a very significant fraction of the population during the first period of colonization, when sugar cultivation and the warehousing of slaves for the transcontinental trade were the main activities in the islands.

The development of the cacao and coffee economy along with the abolition of the slave trade during the 19th century, had profound social and demographic impacts. The scarcity of workers on the plantations within the archipelago led to the recruitment of outside labour and another wave of Africans began to enter São Tomé e Príncipe. With the contract labourers movement a new social class was established in the archipelago: that to which Tongas belong, and one that represented an economically and socially disfavoured stratum (Tenreiro, 1961; Almeida, 1966). Their original homelands were widespread throughout all the African former Portuguese colonies: Angola, Cabo Verde, Guinea-Bissau and Mozambique. Thus, basically Tongas had the same ancestry as Forros. Indeed in a sense, these groups can be considered as two sub-samples of the mainland African parental stock, which began to enter the archipelago with a time displacement of a few centuries.

The genetic data is in perfect agreement with the historical record. For Y-chromosome STR variation, Tongas and Forros share essentially the same genetic profile with each other and with Angola, Mozambique and even Guinea-Bissau. The strong genetic affinities between these populations are graphically well expressed in the tree of Fig. 3. The more marked differences between the São Tomean sample and Cabo Verde do not contradict their common recent African ancestry; they only reflect the different trajectories of the settlement process in the two archipelagos. In summary, the social hierarchization between Forros and Tongas does not have any genetic parallel, at least in the male lineages of their genetic pool.

The origin of Angolares is still uncertain. According to an old legend, they could be the descendents of slaves saved from a shipwreck that occurred in the 16th century. Another more likely hypothesis states that Angolares are descendants of fugitive slaves from the sugar-mills that had to take refuge in the most isolated southwestern regions of São Tomé. The available historical sources do not give reliable clues to the origin of Angolares, but their presence in São Tomé as a distinctive cultural and social group living in relative isolation, is documented as early as the 1840s (Costa, 1982; Ambrósio, 1984; Henriques, 2000). Nowadays, Angolares live on the southwestern coast of São Tomé, where they subsist with fishing as their main livelihood. In spite of their relative integration in the social tissue and economic structure of the island, they still retain a strong group identity and show distinctive hierarchies according to their social,

cultural and occupational activities. Indeed, signs of only slight genetic microdifferentiation were found among Angolares compared to the other groups. Angolares also showed the most reduced levels of Y-chromosome variability, when assessed by various parameters: average gene diversity, haplotype diversity and mean number of pairwise differences. This reduced Y-chromosome diversity could indicate a weak founding effect in the male lineages of Angolares. It could also have arisen by drift or endogamy due to a small effective population size or perhaps a combination of all these factors. Again the results are in good agreement with the ethnohistorical data which indicate that Angolares comprise a relatively small closed group that lived in comparative isolation and with scarce contacts with other São Tomé e Príncipe inhabitants.

Our results have also shown that at least 4 compound haplotypes present in São Tomé e Príncipe most likely represent Portuguese male lineages; this demonstrates that traces of the admixture process that occurred during the archipelago settlement are still alive in the present-day population. Further studies extended to an enlarged battery of genetic markers, namely autosomal polymorphisms and mtDNA, will be most relevant for obtaining reliable admixture estimates and reconstructing the pattern of sex mediated gene flow.

Mismatch distribution of Y-chromosome STRs

In contrast to single nucleotide polymorphisms, where there are long standing robust models which account for the the ways in which past demographic events affect the amount and distribution of the molecular diversity in populations (Rogers *et al.* 1992; Harpending *et al.* 1998; Schneider & Excoffier, 1999), the population dynamics of microsatellite variation has only recently become the subject of intensive examination (Kimmel *et al.* 1998; Gonser *et al.* 2000). Also, it has not yet been possible to develop models to predict the distribution of expected pairwise differences in the evolutionary context of Y-chromosome STRs. Our results are suggestive that mismatch distributions applied to Y-chromosome STRs can be, at least, informative summary statistics. The overall sample of São Tomé e Príncipe and that from Cabo Verde presented very similar bell-shaped distribution curves with high modes. Both populations had their recent origins from processes of population admixture involving heterogeneous African parental stocks and minor, for São Tomé e Príncipe, or substantial, for Cabo Verde, European contributions. Thus, it seems that admixture imprints the distribution of pairwise STR differences in a way similar to that predicted by population expansions from single nucleotide diversity data.

The distributions observed for Guinea-Bissau, Mozambique and Angola were slightly different, either between each other or between the above-referred patterns. Regarding Mozambique the raggedness of the distribution was evident. For single nucleotide diversity, ragged distributions are predicted for stationary populations while regular bell-shaped ones are expected in populations that have experienced expansions in a time-period reflected in the positioning of the modal class in the distribution wave (Harpending, 1994). In Africa, and for mtDNA variation, the former pattern was observed in hunter-gatherers, whilst the latter was systematically presented by farmers or pastoralists (Watson *et al.* 1996). It will be important to find out whether similar effects were also left in the paternally inherited counterpart and particularly among microsatellite variation.

Final remarks

Our data are in clear accordance with those of Pritchard *et al.* (1999), showing that for Y-chromosome STRs there does not seem to exist a considerable excess of variability in African populations. The mean heterozygosity per locus and the mean variance in the number of repeats in our overall sample were even lower than the values registered by Pritchard *et al.* for Africans. For European samples the tendency was the opposite, towards higher values than those obtained by the same authors.

The very different sample schemes used in the two studies must naturally be taken into account. Despite this, it seems clear that they point to a smooth differential among levels of Y-chromosome diversity between Africans and at least Europeans. This conclusion was further reinforced when the enlarged data set for Iberian populations of González-Neira *et al*. (2000) led to estimates, for the same STRs, of mean gene diversity across loci of 0.552 and average variance in the number of repeats per locus of 0.573. Therefore, with respect to the amount of Y-chromosome STR diversity, the major differences between Africans and Europeans seem to lie on the mean variance in the number of repeats, which tend to be higher in Africans. The observation of many overlapping haplotypes between Africans and Caucasians does not suggest the existence of a very sharp macro-geographical structure. This adds support to the relative lack of large-scale geographical pattern in Y-chromosome STRs that has so far been reported (Deka *et al*. 1996; de Knijff *et al*. 1997; Kayser *et al*. 1997*b*; Jorde *et al*. 2000). It might be partially responsible for the very low overall GST values that Y-chromosome STR diversity produces at the continental level (Africa, Asia, Europe), when compared to other kinds of genetic polymorphisms, including Y-chromosome single nucleotide polymorphisms for which, in clear contrast, a strong tendency for geographical clustering has been reported (Hammer *et al*. 1998; Karafet *et al*. 1999)

As invoked elsewhere (Perez-Lezaun *et al*. 1999), a combination of high mutation rates in STRs, leading frequently to convergence in allele sizes, and low effective size of Y-chromosomes, can probably erase long-term population history, making these kind of polymorphisms potentially more useful for reconstructing relatively recent historical episodes. That was the underlying conjecture of this study.

We thank His Excellence the Minister of Health of São Tomé e Príncipe, Dr. António Lima, for giving the permission and encouragement for our research; and the Health Directors who provided the facilities for sample collection. We also thank the laboratory technicians for the excellent technical assistance. Special thanks to the blood donors and the community leaders who made this study possible. This work was supported by Programa PRAXIS XXI through project PRAXIS/2/2.1/BIA/196/94.

REFERENCES

Almeida, A. (1966). *Das etnonimias da Guiné Portuguesa, do arquipélago de Cabo Verde e das ilhas de São Tomé e Príncipe*. Instituto Superior de Ciências Sociais e Política Ultramarina.

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

Bosch, E., Calafell, F., Santos, F. R., Perez-Lezaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., Bertranpetit, J. (1999). Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am. J. Hum. Genet.* **65**, 1623–1638.

Bravi, C. M., Sans, M., Bailliet, G., Martinez-Marignac, V. L., Portas, M., Barreto, I., Bonilla, C., Bianchi, N. O. (1997). Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. *Hum. Biol.* **69**, 641–652.

Carvalho, M., Anjos, M. J., Andrade, L., Coxinho, C., Corte-Real, F., Gamero, J. J., Vieira, D. N., Vide, M. C. (2000). Y chromosome polymorphisms: a comparison between Azores and continental Portuguese samples. *Progress in Forensic Genetics* **8**, 302–304.

Corte-Real, F., Carvalho, M., Andrade, L., Anjos, M. J., Pestoni, C., Lareu, M. V., Carracedo, A., Vieira, D. N., Vide, M. C. (2000). Chromosome Y STRs analysis and evolutionary aspects for Portuguese spoken countries. *Progress in Forensic Genetics* **8**, 272–274.

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

Deka, R., Jin, L., Shriver, M. D., Yu L. M., Saha, N., Barrantes, R., Chakraborty, R., Ferrel, R. E. (1996). Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res.* **6**, 1177–1184.

de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Roewer, L. (1997). Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int. J. Legal Med.* **110**, 134–149.

Felsenstein, J. (1993). Phylip (Phylogeny inference package). version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Gonser, R., Donnelly, P., Nicholson, G., Di Rienzo, A. (2000). Microsatellite mutations and inferences about human demography. *Genetics* **154**, 1793–1807.

González-Neira, A., Gusmão, L., Brión, M., Lareu, M. V., Amorim, A., Carracedo, A. (2000). Distribution of Y-chromosome STR defined haplotypes in Iberia. *Forensic Sci. Int.* **110**, 117–126.

Gusmão, L., Gonzáles-Neira, A., Pestoni, C., Brión, M., Lareu, M. V., Carracedo, A. (1999). Robustness of the Y STRs DYS19, DYS389I/II, DYS390, DYS391, DYS392 and DYS393: optimization of a PCR pentaplex. *Forensic Sci. Int.* **106**, 163–172.

Hammer, M. F., Spurdle, A. B., Karafet, T., Bonner, M. R., Wood, E. T., Novelletto, A., Malaspina, P., Mitchell, R. J., Horai, S., Jenkins, T., Zegura, S. L. (1997). The geographic distribution of human Y chromosome variation. *Genetics* **145**, 787–805.

Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., Griffiths, R. C., Templeton, A. R., Zegura, S. L. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**, 427–441.

Harpending, H. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**, 591–600.

Harpending, H. C., Batzer, M. A., Gurven, M., Jorde, L. B., Rogers, A. R., Sherry, S. T. (1998). Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA*. **95**, 1961–1967.

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

Hurles, M. E., Irven, C., Nicholson, J., Taylor, P. G., Santos, F. R., Loughlin, J., Jobling, M. A., Sykes, B. C. (1998). European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *Am. J. Hum. Genet.* **63**, 1793–1806.

Hurles, M. E., Veitia, R., Arroyo, E., Armenteros, M., Bertranpetit, J., Perez-Lezaun, A., Bosch, E., Shlumukova, M., Cambon-Thomsen, A., Mcelreavey, K., Lopez De Munain, A., Rohl, A., Wilson, I. J., Singh, L., Pandya, A., Santos, F. R., Tyler-Smith, C., Jobling, M. A. (1999). Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am. J. Hum. Genet.* **65**, 1437–1448.

Jobling, M. A. & Tyler-Smith, C. (1995). Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* **11**, 449–456.

Jobling, M. A. & Tyler-Smith, C. (2000). New uses for new haplotypes. The human Y chromosome, disease and selection. *Trends Genet.* **16**, 356–362.

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T., Batzer, M. A. (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988.

Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, A. R., D., Klitz. W., Harihara, S., De Knijff, P., Wiebe, V., Griffiths, R. C., Templeton, A. R., Hammer, M. F. (1999). Ancestral Asian source(s). of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**, 817–831.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Roewer, L. (1997*a*). Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* **110**, 125–133.

Kayser, M., De Knijff, P., Dialtjes, P., Krawczak, M., Nagy, M., Zerjal, T., Pandya, A. (1997*b*). Applications of microsatellite based Y chromosome haplotyping. *Electrophoresis* **18**, 1602–1607.

Kimmel, M., Chakraborty, R., King, J. P., Bamshad, M., Watkins, W. S., Jorde, L. B. (1998). Signatures of population expansion in microsatellite repeat data. *Genetics* **148**, 1921–1930.

Malaspina, P., Cuciani, F., Ciminelli, B. M., Terrenato, L., Santolamazza, P., Alonso, A., Banyko, J., Brdicka, R., García, O., Gaudiano, C., Guanti, G., Kidd, K. K., Lavinha, J., Avila, A., Mandich, P., Moral, P., Qamar, R., Mehdi, S. Q., Ragusa, A., Stefanescu, G., Caraghin, M., Tyler-Smith, C., Scozzari, R., Novelletto, A. (1998). Network analyses of Y-chromosomal types in Europe, Northern Africa, and Western Asia reveal specific patterns of geographic distribution. *Am. J. Hum. Genet.* **63**, 847–860.

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.

Page, R. D. M. (1996). Treeview: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Perez-Lezaun, A., Calafell, F., Seielstad, M., Mateu, E., Comas, D., Bosch, E., Bertranpetit, J. (1997). Population genetics of Y-chromosome short tandem repeats in humans. *J. Mol. Evol.* **45**, 265–270.

Perez-Lezaun, A., Calafell, F., Comas, D., Mateu, E., Bosch, E., Martinez-Arias, R., Clarimon, J., Fiori, G., Luiselli, D., Facchini, F., Pettener, D., Bertranpetit, J. (1999). Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* **65**, 208–219.

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.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., Feldman M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**, 1791–1798.

Rogers, A. R. & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569.

Ruiz-Linares, A., Ortiz-Barrientosm, D, Figueroam, M., Mesam, N., Muneram, J. G., Bedoya, G., Velez, I. D., Garcia, L. F., Perez-Lezaun, A., Bertranpetit, J., Feldman, M. W., Goldstein, D. B. (1999). Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proc. Nat. Acad. Sci. USA* **96**, 6312–6317.

Santos, F. R., Pandya, A., Tyler-Smith, C., Pena, S. D., Schanfield, M., Leonard, W. R., Osipova, L., Crawford, M. H., Mitchell, R. J. (1999). The central Siberian origin for native American Y chromosomes. *Am. J. Hum. Genet.* **64**, 619–628.

Schneider, S. & Excoffier, L. (1999). Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*. **152**, 1079–1089.

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.

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

Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L., Oefner, P. J. (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005.

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.

Watson, E., Bauer, K., Aman, R., Weiss, G., Von Haeseler, A., Pääbo, S. (1996). mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.* **59**, 437–444.