

A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea

E. MATEU¹, D. COMAS¹, F. CALAFELL*, A. PÉREZ-LEZAUN¹, A. ABADE²
AND J. BERTRANPETIT^{1,†}

¹*Laboratori d'Antropologia, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain*

²*Departamento de Antropologia, Universidade de Coimbra, Coimbra, Portugal*

(Received 25.7.97. Accepted 17.10.97)

SUMMARY

The hypervariable segment I of the control region of the mtDNA was sequenced in 45 unrelated individuals from Bioko and 50 from São Tomé, two islands in the Gulf of Guinea that have had very different settlement patterns: Bioko was colonized around 10000 BP, while São Tomé was first settled by the Portuguese, who brought African slaves to the island. Two different patterns of sequence variation are evident and are also clearly a consequence of their very different demographic histories. The Bubi present a low genetic diversity and it is likely that the island was colonized by a small number of individuals with small later migration. São Tomeans might be considered a subset of a mainland African population relocated to the island. They present high genetic diversity with a high number of sequences being shared with many continental populations. This study, with knowledge of the population history in island populations, strengthens the genetic approach to unravel past demographic events.

INTRODUCTION

Mitochondrial DNA sequence analysis has proved to be a powerful tool in the study of human population history. It has been applied to very different space and time frames, from the origins of anatomically modern humans (Cann *et al.* 1987; Vigilant *et al.* 1991) to the colonization of continents (Ward *et al.* 1991; Kolman *et al.* 1996; Richards *et al.* 1996; Comas *et al.* 1997), and to specific populations (Santos *et al.* 1994; Bertranpetit *et al.* 1995; Mountain *et al.* 1995; Calafell *et al.* 1996; Comas *et al.* 1996). Several properties of mitochondrial DNA (mtDNA)

make it particularly suitable for evolution studies: absence of recombination, maternal inheritance, and a high mutation rate (Stoneking, 1993).

The impact of demographic processes in mtDNA sequence variation has been thoroughly modelled. In particular, the distribution of nucleotide pairwise differences (also known as mismatch distribution; Rogers & Harpending, 1992; Harpending *et al.* 1993) is particularly sensitive to the past demographic history of a population. In the absence of selection and with mutation rates uniformly distributed across nucleotides, it has been shown (Rogers & Harpending, 1992; Harpending *et al.* 1993) that stationary populations would have irregular mismatch distributions, while a population expansion would generate a bell-shaped distribution, with a mode travelling to the right with time. A few African populations (Pygmies and

* Current address: Department of Genetics, Yale University School of Medicine, New Haven, CT, USA.

† Correspondence: Jaume Bertranpetit, Laboratori d'Antropologia, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Catalonia, Spain. Tel.: +34-3-402 14 61. Fax: +34-3-411 08 87. E-mail: jaumb@porthos.bio.ub.es

!Kung in particular; Harpending *et al.* 1993) present irregular mismatch distributions, while other African populations (Graven *et al.* 1995; Watson *et al.* 1996) and virtually all other populations present bell-shaped mismatch histograms (Harpending *et al.* 1993). It has been shown that, in Europe, mismatch modes decrease from SE to NW, which has been interpreted as the footprint of the colonization of Europe by anatomically modern humans (Calafell *et al.* 1996; Comas *et al.* 1996, 1997; Francalacci *et al.* 1996) recently confirmed by Neandertal mtDNA analysis (Krings *et al.* 1997). However, some of the implicit assumptions in the Rogers & Harpending (1992) model have been reassessed by Aris-Brosou & Excoffier (1996), who found that mutation rate variation across nucleotides (Hasegawa *et al.* 1993; Wakeley, 1993) could also produce bell-shaped mismatch distributions. Therefore, alternative hypotheses for the interpretation of mismatch distributions should be considered carefully before inferring a particular population history.

As mentioned above, some African populations present ragged pairwise difference distributions, while others have bell-shaped distributions. Watson *et al.* (1996) noted that the first are hunter-gatherers while the latter are farmers or pastoralists. Thus, they interpreted that the demographic expansion that produced bell-shaped mismatch distributions in Africa was the spread of the Neolithic. The latest direct counts of mutation events in the mtDNA (Howell *et al.* 1996; Parsons *et al.* 1997) resulted in estimates of the mutation rate that were higher than previously thought; this adds support to a role for a recent demographic event, such as the Neolithic expansion, in generating bell-shaped mismatch distributions in Africa (Watson *et al.* 1996) and maybe elsewhere (Sajantila *et al.* 1995; Pääbo, 1996; von Haeseler *et al.* 1996).

Sub-Saharan African populations share a pattern of mitochondrial DNA variation, with high levels of genetic diversity both within and between populations (Vigilant *et al.* 1991). This has also been found in nuclear genes (Armour *et al.* 1996; Tishkoff *et al.* 1996), and is compatible

with the 'Out of Africa' hypothesis, according to which non-African humans have a common, recent origin in Africa. This genetic pattern could increase the power of genetic population history analysis in African populations: a population bottleneck in an African population might be detected by comparison with the high levels of genetic diversity in other populations, or possible migrations can be traced back due to interpopulation variability.

We have analysed two samples from two African islands: the Bubi from Bioko and individuals from São Tomé. Bioko is a 2000 km² island in the Gulf of Guinea, 30 km off the Cameroon coast, which, together with four smaller islands and the mainland territory of Rio Muni constitute the Republic of Equatorial Guinea. The island was first colonized 10000 years BP (Vara & Bolekia, 1993), at the end of the last glacial period. Around 2000 years BP, farming, and possibly a Bantu language, were introduced to the island (Martín del Molino, 1993). The Bantu-speaking Bubi are the only population native to Bioko, and thought to be the descendants of the original colonizers of the island; the contact with Europeans decimated them to a few thousand at the turn of the 20th century. Nowadays, they number 35000, and share the island with mainland Fang and Fernandinos, the later being descendants of former slaves liberated by the English in the 19th century. São Tomé e Príncipe, a former Portuguese colony, is located on the Equator in the Gulf of Guinea. It consists of two main islands (São Tomé and Príncipe) and a number of islets. Their total area is 964 km², of which São Tomé comprises 865 km². São Tomé island was probably uninhabited when first visited by European navigators in the 1470s. Thereafter, the Portuguese began to settle convicts and exiled Jews from Portugal on the island and established sugar plantations, using slave labour from the African mainland; for some years São Tomé was important in the trade and transshipment of slaves. A recent (1995) population size estimate for São Tomé is around 100000 inhabitants and is mostly of African descent.

Bioko and São Tomé are, thus, two African islands in close proximity, with similar areas and population sizes, but with very different settlement patterns. The Bubi of Bioko are the descendants of one or a few ancient waves of migration from the continent, whereas the São Tomeans represent an admixed population with a recent origin. It is likely that the original settlers of Bioko were a small number of individuals whose descendants in around 500 generations increased to the actual number of Bubis. However, São Tomé may have been peopled by a larger number of imported slaves. These different patterns of settlement may have had genetic consequences. If the number of settlers of Bioko was small enough, we may be able to observe a reduction of genetic diversity in the Bubi when compared to the mainland populations and to São Tomé. Under isolation, gene diversity should be reduced, whereas the complexity of the genetic variation, as measured by coalescence patterns, mean pairwise differences, and nucleotide diversity, should remain comparable to that of mainland populations. In order to discover how two very different population histories have influenced the genetic structure of the population, we have sequenced a fragment of 360 base pairs in the hypervariable segment I of the mtDNA in 45 Bubi and 50 individuals from São Tomé, and we have compared the sequences to a set of 15 African and two European populations.

MATERIAL AND METHODS

Population sampling

A 360-nucleotide sequence in hypervariable segment I (HVSI) of the mtDNA control region was analysed in 45 individuals from Bioko island, Equatorial Guinea. The sample comprised self-described unrelated Bubi individuals from the villages of Moka-Bioko, Moka-Malabo, in southern Bioko, and Rebola, in northern Bioko. The sample from São Tomé island comprises 50 individuals from different places covering the whole island.

Sample collection and DNA extraction

DNA was extracted from hair roots for the Bubi individuals. Hairs with their roots were plucked and stored in a vial with 95% ethanol. One root of each sample was introduced in a 1.5 ml sterile microfuge tube containing 0.5 ml of extraction buffer (10 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 100 mM NaCl, 2% SDS, 39 mM DTT, and 20 mg ml⁻¹ proteinase K), then incubated at 37 °C and shaken at 180–200 rpm for at least 3 hours. After a phenol-chloroform extraction (Sambrook, 1989), DNA was concentrated in Centricon-30 tubes and stored at –20 °C. For the São Tomé sample DNA (supplied by A. Abade, Coimbra) was extracted from fresh blood using standard protocols.

mtDNA amplification

Amplification was performed using approximately 150–250 ng of the DNA sample in a 25 µl reaction volume; the temperature profile for 30 cycles of amplification was 94 °C for 1 min, 58 °C for 1 min and 72 °C for 1 min. The primers used in this reaction, L15996 (5'-CTCCACCATAGC-ACCCAAAGC-3'), and H16401 (5'-TGATTTCCGAGGATGGTG-3'; Vigilant *et al.* 1989) amplified a 446-base pair (bp) segment containing the 360-bp region that was subsequently sequenced.

mtDNA sequencing

Out of the 45 Bubi samples, 13 were sequenced with an automatic DNA sequencer, while the remaining 32 were sequenced manually; the choice of method depended only on sequencer time availability. All the São Tomé samples were sequenced with an automatic DNA sequencer. Automated sequencing was performed according to manufacturer's specifications. The sequencing reaction was performed separately on each strand with the DNA Sequencing Kit™ (Perkin Elmer), Dye Terminator Cycle Sequencing with AmpliTaqR DNA Polymerase. The product of

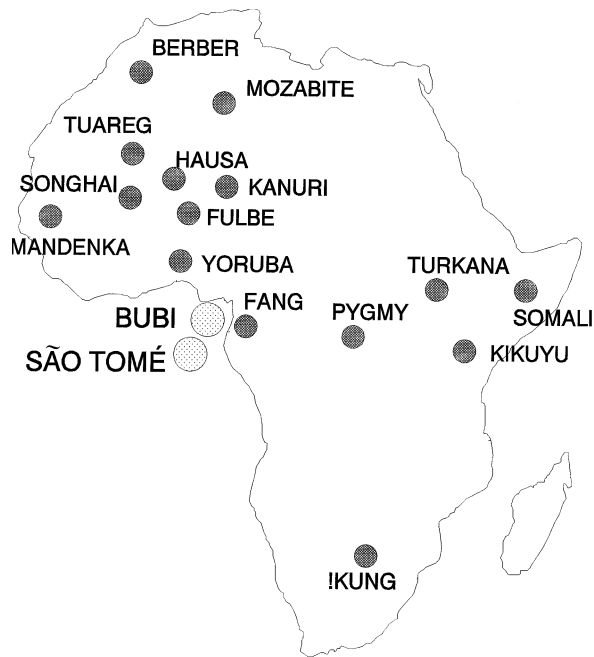


Fig. 1. African populations included in our study.

the sequence reaction was run in an ABI PRISM 377 (Perkin Elmer).

When sequencing manually, the amplification product was purified with GeneClean (BIO 101) and 7 μ l of the purified amplified product were sequenced with Sequenase Version 2.0 (USB) following supplier's recommendations, except for the annealing step, which was performed by boiling the annealing reaction mixture for 3 min in presence of nonidet P-40, followed by a short time in a dry-ice ethanol bath. Both strands were sequenced using the amplification primers. Reaction products were separated by electrophoresis, dried, fixed, and subjected to autoradiography.

Sequences were aligned with the ESEE program (Cabot, 1988), and the segment from positions 16024 to 16383 (Anderson *et al.* 1981) was used for analysis.

Statistical analysis

A set of 15 African population samples, comprising a total of 645 individuals and including the Fang in mainland Equatorial Guinea (Figure 1; Table 1) and two European (a British

sample, $n = 100$, Piercy *et al.* 1993 and a Portuguese sample, $n = 54$, Côrte Real *et al.* 1996) were used as reference.

Nucleotide diversity (Nei & Tajima, 1981) was estimated as $(n/n-1) \sum_{i=1}^l (1-x_i^2)$, where n is sample size, l is sequence length, and x_i is the frequency of a nucleotide (A, C, G or T) at position i . Similarly, sequence diversity was estimated as $(n/n-1) \sum_{i=1}^k (1-p_i^2)$, where p_i is the frequency of each of the k different sequences in the sample. The significance of the difference in sequence diversity between two populations was tested through a permutation procedure: the individuals in both populations were randomly assigned to one of two samples of the same size as the original ones; sequence diversities were computed for the new random samples, and the difference was recorded. This procedure was repeated 10000 times, and the probability of the difference not being significantly smaller than zero was estimated as the fraction of permuted differences that were less extreme than the observed value. Tajima's (1989) D statistic, which is the standardized difference between two different estimates of $\Theta = 2N_e\mu$, was computed. Under a number of assumptions, Tajima's D measures the deviation from mutation-drift equilibrium. Bertorelle & Slatkin (1995), and Aris-Brosou & Excoffier (1996) have examined the effects of population expansion and mutation rate variability on D . We also estimated Θ from the number of segregating sites (Watterson, 1975) through equation $\Theta = P_n / (1 + 1/2 + \dots + (n-1)^{-1})$ in Nei (1987), where P_n is the number of segregating sites divided by 360 and n is the number of the observed sequences.

The phylogeny of mtDNA sequences in the Bubi and in the São Tomeans was approached through a neighbour-joining tree (Saitou & Nei, 1987) based on Kimura's two-parameter model with a transition to transversion ratio set to 15:1. We used PHYLIP (Felsenstein, 1989) to produce these sequence phylogenies. The distribution of branch lengths in a phylogeny can be used to quantify its degree of *starness*, that is, whether most of the sequences attach with long branches to a central point, as opposed to a

Table 1. African population set used for comparison^a

Population	Reference	<i>n</i>	Country	Language Family (Branch)
Bubi	This study	45	Equat. Guinea	NK (Bantu)
São Tomé	This study	50	São Tomé e Príncipe	Portuguese and Creole
Fang	Pinto <i>et al.</i> 1996	11	Equat. Guinea	NK (Bantu)
Yoruba	Vigilant <i>et al.</i> 1991; Watson <i>et al.</i> 1996	35	Nigeria	NK (Central Niger Congo, non Bantu)
Fulbe	Watson <i>et al.</i> 1996	61	Nigeria, Niger, Benin, Cameroon, Burkina Faso	NK (West Atlantic)
Mandenka	Graven <i>et al.</i> 1995	119	Senegal	NK (Mande)
Kanuri	Watson <i>et al.</i> 1996	14	Nigeria, Niger	NS (Saharan)
Songhai	Watson <i>et al.</i> 1996	10	Niger, Mali	NS (Songhai)
Hausa	Watson <i>et al.</i> 1996	20	Nigeria, Niger	AA (Chadic)
Somali	Watson <i>et al.</i> 1996	27	Kenya, Somalia, Ethiopia	AA (Cushitic)
Turkana	Watson <i>et al.</i> 1996	37	Kenya	NS (East Sudanic)
Kikuyu	Watson <i>et al.</i> 1996	25	Kenya	NK (Bantu)
Pygmy	Vigilant <i>et al.</i> 1991	37	CAR, Zaire	NK (Bantu)
!Kung	Vigilant <i>et al.</i> 1991	25	Namibia, Botswana	KH (Northern)
Tuareg	Watson <i>et al.</i> 1996	26	Niger, Mali	AA (Berber)
Mozabite	Côrte-Real <i>et al.</i> 1996	85	Algeria	AA (Berber)
Berber	Pinto <i>et al.</i> 1996	18	Morocco	AA (Berber)

^a N: Sample size. Abbreviations: NK, Niger-Kordofanian; NS, Nilo-Saharan; AA, Afro-Asiatic; KH, Khoisanid.

pattern in which branching events are more regular. The two patterns may reflect different population histories: a star-like tree may correspond to an expanding population, whereas a regular tree could reflect a stationary population (von Haeseler *et al.* 1996). In order to characterize the branch length distribution, we computed its third- and fourth-degree moment (i.e. its skewness and kurtosis).

Genetic distances among populations were estimated as $d = d_{ij} - (d_{ii} + d_{jj})/2$ (Nei, 1987), where d_{ij} is the mean nucleotide pairwise difference between populations i and j , and d_{ii} and d_{jj} are the mean internal pairwise differences of populations i and j , respectively. The standard error of the distances was estimated by resampling nucleotide positions in 1000 bootstrap iterations. A neighbour-joining tree (Saitou & Nei, 1987) was built from the genetic distance matrix, and its robustness was assessed from 1000 bootstrap iterations (Felsenstein, 1985).

RESULTS

Nucleotide diversity

The complete sequence of a 360-bp segment of the mtDNA control region (HVS1, positions 16024 to 16383 according to the numeration by Anderson *et al.* 1981) was determined in a sample of 45 Bubi from Bioko and 50 individuals from São Tomé (Table 2). For the Bubi sample, 18 different sequences were found, with 32 variable nucleotide positions. The São Tomeans presented 32 different sequences, with 53 variable nucleotides. Overall, 48 different sequences and 61 segregating sites were found. Four of the segregating sites presented transversional changes, 36 were transitions between pyrimidines (C to T or *vice versa*), 19 were purine transitions, and in two positions (16114 and 16265), we observed both transitions and transversions. Whereas two thirds of the transitions observed involve pyrimidines, these represent 55.8% of nucleotides in this segment. This might represent part of the uneven distribution of mutation rates across nucleotides (Bertorelle & Slatkin, 1995; Aris-Brosou & Excoffier, 1996).

Table 3. Nucleotide and sequence polymorphism in some African and two European populations

	<i>n</i>	<i>k</i>	<i>S</i>	π	<i>D</i>	<i>h_s</i>	Θ_w	<i>d_{ii}</i>
Bubi (1)	45	18	32	0.0210	0.114 NS	0.928	0.020	7.564
São Tomé (1)	50	32	53	0.0231	-1.034 NS	0.973	0.033	8.300
Fang (2)	11	10	33	0.0256	-0.854 NS	0.982	0.031	9.200
Yoruba (3,4)	35	33	47	0.0208	-1.160 NS	0.997	0.031	7.761
Fulbe (4)	61	39	47	0.0199	-0.963 NS	0.973	0.028	7.156
Mandenka (5)	119	54	58	0.0193	-0.729 NS	0.971	0.030	7.751
Kanuri (4)	14	13	35	0.0207	-1.397 NS	0.989	0.031	7.450
Songhai (4)	10	9	31	0.0257	-0.752 NS	0.978	0.030	9.244
Hausa (4)	20	19	33	0.0171	-1.333 NS	0.995	0.026	6.158
Somali (4)	27	24	46	0.0216	-1.322 NS	0.991	0.033	7.761
Turkana (4)	37	35	63	0.0296	-1.072 NS	0.995	0.042	10.65
Kikuyu (4)	25	23	49	0.0241	-1.276 NS	0.993	0.036	8.670
Pygmy (3)	37	23	34	0.0252	1.051 NS	0.944	0.023	10.27
!Kung (3)	25	12	21	0.0089	-1.038 NS	0.843	0.015	3.187
Tuareg (4)	26	21	41	0.0197	-1.282 NS	0.985	0.030	7.105
Mozabite (6)	85	29	35	0.0133	-1.056 NS	0.942	0.014	4.730
Berber (2)	18	15	24	0.0123	-1.440 NS	0.980	0.019	4.438
Portuguese (6)	54	38	40	0.0099	-1.988 *	0.934	0.024	3.597
British (7)	100	71	67	0.0123	-2.121 *	0.966	0.036	4.453

All parameters have been computed for the 360-bp segment between positions 16,024 and 16,383. *n*, sample size; *k*, number of different sequences; *S*, number of segregating sites; π , nucleotide diversity; *D*, Tajima's (1989) statistic, and its statistical significance (* 0.01 < *p* < 0.05). *h_s*, sequence diversity; Θ_w , estimator of 2*N_e* μ (Watterson, 1975); *d_{ii}*, mean number of nucleotide differences between pairs of individuals. *Sources*: 1, this study; 2, Pinto *et al.* 1996; 3, Vigilant *et al.* 1991; 4, Watson *et al.* 1996; 5, Graven *et al.* 1995; 6, Côrte-Real *et al.* 1996; 7, Piercy *et al.* 1993.

Nucleotide diversity (Nei & Tajima, 1981) was 0.0210 ± 0.0006 in the Bubi and 0.0231 ± 0.0006 in the São Tomeans. Both values are similar to those found in other West African populations (Table 3). Tajima's statistic was *D* = 0.114 in the Bubi and *D* = -1.034 in the São Tomeans, both not significantly different from zero according to Table 3 in Tajima (1989). All African populations sequenced so far present non-significant Tajima's statistics, whereas European populations tend to present Tajima's statistics significantly smaller than zero.

Sequence diversity

Three sequences (MAL 6, MAL 10 and MAL 19; Table 2) accounted for 42% of the individuals in the Bubi sample, while nine sequences were found once. In contrast, we found 32 different sequences in 50 São Tomeans, and the three most frequent sequences represented 26% of the individuals in the sample. Sequence diversity was lower in the Bubi (0.928) than in the São Tomeans (0.973; Table 3). The difference in sequence diversity was significant (*p* = 0.0001) according to the permutation test. The Bubi

have also significantly smaller sequence diversities than the Mandenka (*p* = 0.0001) and the Yoruba (*p* = 0.0001), while sequence diversity in the São Tomeans was similar to that in the Mandenka (*p* = 0.5601) but lower than in the Yoruba (*p* = 0.0060). We also estimated Θ = 2*N_e* μ as suggested by Watterson (1975). Differences in Θ among populations should be a function of effective population size, since mutation rate is presumably constant across populations. After the !Kung, the Bubi present the second lowest Θ value among sub-Saharan African populations (Table 3), whereas the São Tomeans have a Θ in the mid to upper African range.

The Bubi shared five sequences with other African populations, including two with the São Tomeans. Two of them (MAL 12 and SAO 122; Table 2) were found in several West African populations and in the Kenyan Kikuyu. The other three sequences were found in one sample each (i.e. the São Tomeans, the Pygmies and the Fulbe). Eleven of the sequences found in the São Tomeans had previously been described in other African populations. The São Tomeans shared sequences with a variety of West and East

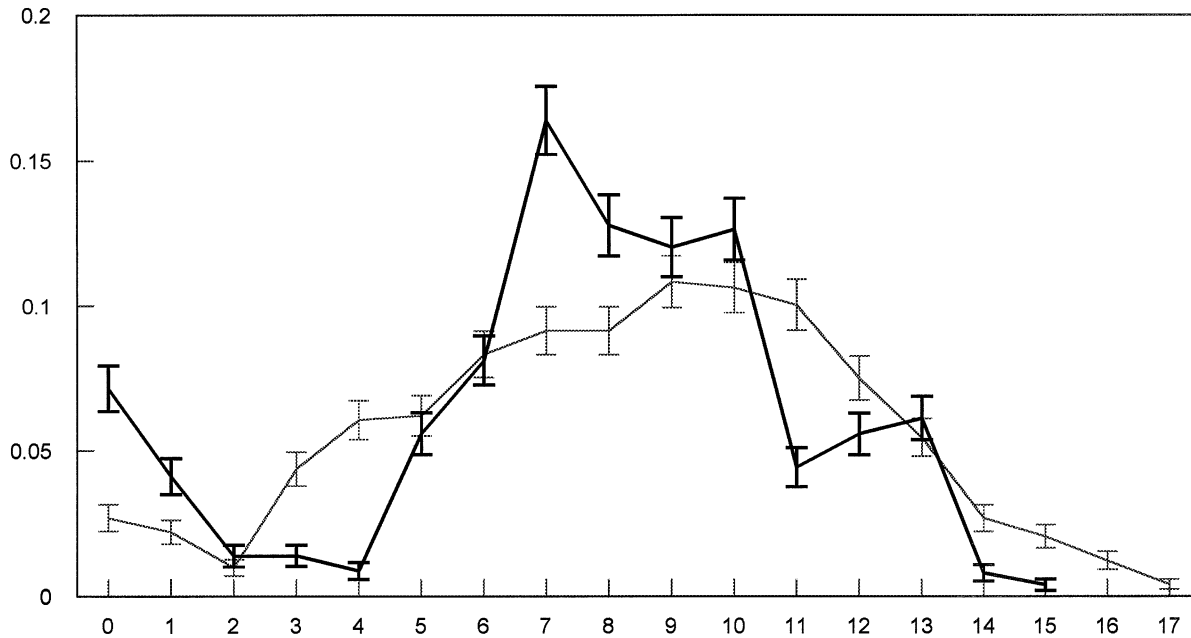


Fig. 2. Nucleotide pairwise difference (mismatch) distribution in the Bubi (black line) and São Tomeans (grey line). Standard errors were estimated by resampling with replacement 1000 times over pairs of individuals.

African populations; they shared five sequences with the Fulbe, four with the Mandenka, three with the Kikuyu and the Hausa, two with the Bubi, Yoruba, Turkana, and Tuareg and one with Kanuri and Mozabite. None of the sequences we found in both samples had previously been described in individuals of European ancestry.

A group of four sequences (SAO 118, SAO 198, SAO 160 and MAL 46; Table 2) bears a characteristic group of mutations that were found to be associated with a 9-bp deletion in the CoII/tRNA^{Lys} intergenic region in African populations (Soodyall *et al.* 1996). If the association described by these authors holds in the Bubi and São Tomeans, then the frequency of the 9-bp deletion would be 8.9% in the Bubi and 8% in the São Tomeans. These frequencies are intermediate between those found in West Africans (0–2%) and in Central Africans (0–30%).

Pairwise difference distribution

The mean number of pairwise nucleotide differences was 7.56 in the Bubi and 8.30 in the São Tomeans; both values fall in the range

observed for other West African populations (Table 3). The nucleotide pairwise difference distributions, also called mismatch distribution (Rogers & Harpending, 1992; Harpending *et al.* 1993) are shown in Figure 2. The Bubi distribution presents several modes, at zero, seven, ten and thirteen differences, and its raggedness coefficient (Harpending *et al.* 1993) is 0.0223. The São Tomean distribution is smoother (raggedness coefficient, 0.0045), bell-shaped, and presents a single mode at nine differences. As shown by the standard errors (Fig. 2), the different shapes of the two distributions cannot be attributed to a random effect. The multiple peaks and troughs in the ragged Bubi distribution are significantly different from the values in the smoother São Tomé distribution.

Phylogenetic tree of sequences

We estimated the phylogeny of Bubi and São Tomean sequences by means of genetic distances and neighbour-joining trees (Fig. 3). While the Bubi phylogeny (Fig. 3a) presented a few, deep-rooting branches, the São Tomé tree reflects a star-like phylogeny (Fig. 3b). It has been shown

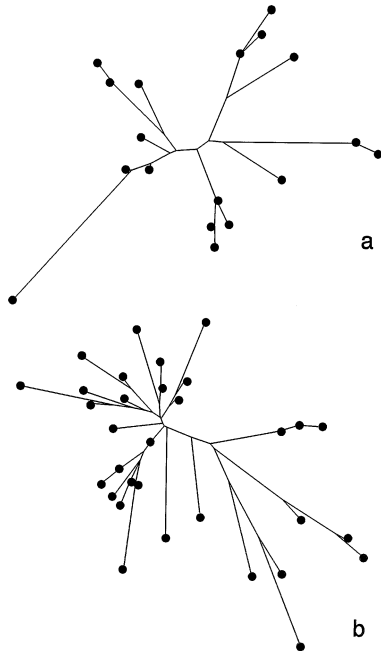


Fig. 3. Phylogenetic trees of Bubi (Fig. 3a) and São Tomeans (Fig. 3b) sequences.

that stable demographic histories result in deep-rooting phylogenies, while population expansion tend to generate star-like phylogenies (Slatkin & Hudson, 1991; Haeseler *et al.* 1996). The difference between the two populations may be measured by parameters of the distribution of branch length in the two trees. Skewness and kurtosis are significant higher in Bubi ($\mu_3 = 2.142 \pm 0.414$; $\mu_4 = 5.360 \pm 0.809$) than in São Tomeans ($\mu_3 = 1.250 \pm 0.304$; $\mu_4 = 0.597 \pm 0.599$), being these parameters an indirect estimator of the degree of *starness* of a phylogeny.

Genetic distances among African populations

Genetic distances were computed between several African populations, plus two European reference populations. The Bubi presented relatively long distances to other African populations. São Tomé presented the shortest genetic distances to the Yoruba (0.11), and to other West African and Sahelian populations (0.2–0.3). A neighbour-joining tree (Fig. 4) based on those distances presented two particularly long, robust branches, which divide the populations in three main groups: (i) The Pygmy and !Kung popula-

tions; the latter join the tree with a long branch; (ii) the Western African and Sahelian populations which are connected to each other by short branches, and (iii) the North African and European populations. The Bubi, as well as the mainland Equatorial Guinea Fang, appear at the base of the Pygmy-!Kung branch, whereas the São Tomeans are found in the tight West African-Sahelian cluster.

DISCUSSION

We have sequenced the mtDNA HVSI region in individuals from two islands in the Gulf of Guinea that have very different settlement patterns: Bioko was first colonized around 10000 BP (Vara & Bolekia, 1993) and might have received few immigrants since then, while São Tomé was first settled by the Portuguese, who brought African slaves to the island.

We have found some similarities in the mtDNA sequences of both populations. The sequence pool of both samples is clearly African, as shown by the tree of populations (Fig. 4) and by the fact that the Bubi and the São Tomeans share identical sequences only with Africans. Thus not a single sequence of the present population has a female European ancestry. As most African populations, and unlike populations from other continents, both Bubi and São Tomeans present a high mean number of pairwise differences. However, two different patterns of sequence variation between the two populations are evident and are also clearly consequences of their very different demographic histories.

The Bubi present a low sequence diversity and Θ estimate; both may be the result of a long-term female effective population size that is lower than that of other African populations. From what is known from historical and archaeological sources (Vara & Bolekia, 1993), it is likely that the island was colonized by a small number of individuals and that later contributions from the continent were not significant in population size terms. Thus, the population growth might have been slow, even during the Neolithic transition, and until recent times. This is re-

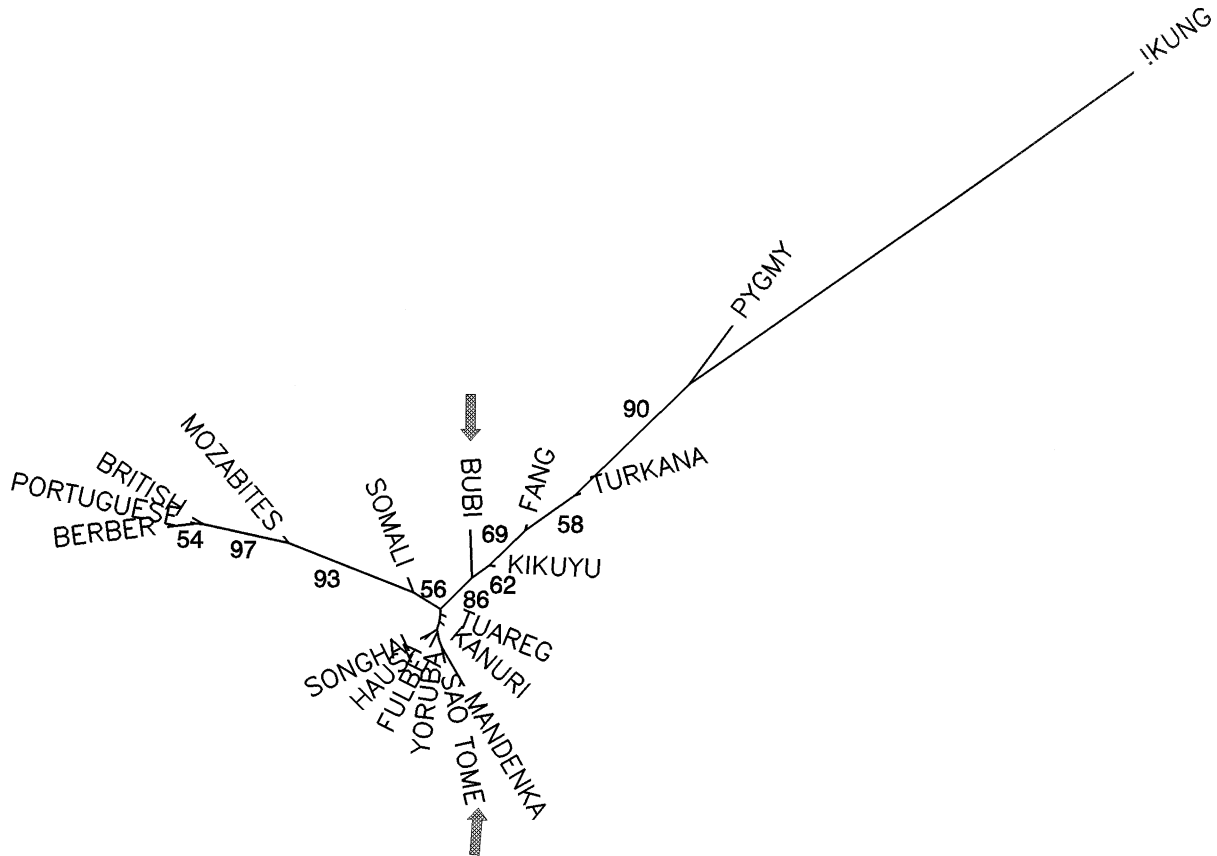


Fig. 4. Hypervariable segment I neighbour-joining tree among 17 African and two European population. Figures in the branches represent the percent fraction of times they were found in 1000 bootstrap iterations.

flected in the sequence phylogeny, with deep-rooting branches, and in the ragged distribution of pairwise differences. The ancient origin of the Bubi is also underscored by their location in the population tree, out of the tight West African cluster and at the base of the branch leading to such ancient African populations as the Pygmies and the !Kung.

Given the origin of the São Tomeans as the descendants of slaves, they might be considered a subset of a mainland African population relocated to the island. This origin is clearly reflected in their mtDNA sequence variation pattern, which is typical of an African population. They present high nucleotide and sequence diversities, and high Θ estimates, which are indications of a relatively high female effective population size. Graven *et al.* (1995) suggested that the Mandenka from Senegal had an effective population size that was noticeably larger than the observed female census size, due

to past or present migration from other populations. The same can be applied to the São Tomeans and their immediate ancestors in the continent. Those were most likely farming Bantu-speakers. The pairwise difference distribution we observed in the São Tomeans conforms to the pattern described by Watson *et al.* (1996) for the mainland farmers: it is smooth and bell-shaped, which corresponds to a starlike sequence phylogeny. The very recent mainland African origin of the São Tomeans resulted in a higher number of sequences being shared with the mainland populations, and, in the population tree, they clearly belong to the tight West African cluster.

We have seen how two very different population histories generated different mtDNA sequence variation patterns in two populations. The accuracy with which sequence variation patterns could be predicted from known population history provides reliability to the opposite

endeavour: inferring unknown or partially known population histories from observed genetic variation. This approach, thus, has a robust genetic foundation with widening possibilities thanks to the growing theoretical framework that helps to understand and explain the extent of human genome variation.

This research was supported by DGICYT (Spain) grant PB92-0722 and PB95-0267-C02-01 and by Generalitat de Catalunya (Catalonia), Grup de Recerca Consolidat 1995SGR00205 and 1996SGR00041 to J. B. E. M. was awarded a PhD fellowship from the University of Barcelona; F. C. was supported by a post-doctoral fellowship from CIRIT (Catalonia), who also granted a PhD fellowship to D. C. A. P.-L. was awarded a PhD fellowship by DGICYT (Spain). Samples from Bioko were kindly supplied by Carmen Maté and Montserrat Colell during her field primate behaviour studies. Samples from São Tomé were collected with funding from STRDA/P/CEN/532/92 portuguese project. We also thank the Servei de Seqüenciació, Serveis Científic Tècnics, University of Barcelona, for their invaluable technical support.

REFERENCES

- ANDERSON, S., BANKIER, A. T., BARRELL, B. G., DE BRULJN, M. H., COULSON, A. R., SANGER, F., SCHREIER, P. H., SMITH, A. J. H., STADEN, R. & YOUNG, G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
- ARIS-BROSOU, S. & EXCOFFIER, L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol. Biol. Evol.* **13**, 494–504.
- ARMOUR, J. A. L., ANTTINEN, T., MAY, C. A., VEGA, E. E., SAJANTILA, A., KIDD, J. R., KIDD, K. K., BERTRANPETIT, J., PÄÄBO, S. & JEFFREYS, A. J. (1996). Minisatellite diversity supports a recent African origin for modern humans. *Nature Genetics*, **13**, 154–160.
- BERTORELLE, G. & SLATKIN, M. (1995). The number of segregating sites in expanding human populations, with implications for estimates of demographic parameters. *Mol. Biol. Evol.* **12**, 887–892.
- BERTRANPETIT, J., SALA, J., CALAFELL, F., UNDERHILL, P., MORAL, P. & COMAS, D. (1995). Human mitochondrial DNA variation and the origin of the Basques. *Ann. Hum. Genet.* **59**, 63–81.
- CABOT, E. L. (1988). *ESEE, the eyeball sequence editor, version 1.06*. Burnaby, B. C., Canada, V5C 2YZ.
- CALAFELL, F., UNDERHILL, P., TOLUN, A., ANGELICHEVA, D. & KALAYDJIEVA, L. (1996). From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann. Hum. Genet.* **60**, 35–49.
- CANN, R. L., STONEKING, M. & WILSON, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.
- COMAS, D., CALAFELL, F., MATEU, E., PÉREZ-LEZAUN, A. & BERTRANPETIT, J. (1996). Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol. Biol. Evol.* **13**, 1067–1077.
- COMAS, D., CALAFELL, F., MATEU, E., PÉREZ-LEZAUN, A., BOSCH, E. & BERTRANPETIT, J. (1997). Mitochondrial DNA variation and the origin of the Europeans. *Hum. Genet.* **99**, 443–449.
- CÔRTE-REAL, H. B. S. M., MACAULAY, V. A., RICHARDS, M., 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**, 331–350.
- FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **35**, 785–791.
- FELSENSTEIN, J. (1989). PHYLIP – Phylogeny Inference Package. *Cladistics* **5**, 164–166.
- FRANCALACCI, P., BERTRANPETIT, J., CALAFELL, F. & UNDERHILL, P. (1996). Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am. J. Phys. Anthropol.* **100**, 443–460.
- GRAVEN, L., PASSARINO, G., SEMINO, O., BOURSOT, P., SANTACHIARA-BENERECETTI, S., LANGANEY, A. & EXCOFFIER, L. (1995). Evolutionary correlation between control region sequence and restriction polymorphism in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol. Biol. Evol.* **12**, 334–345.
- HARPENDING, H. C., SHERRY, S. T., ROGERS, A. R. & STONEKING, M. (1993). The genetic structure of ancient human populations. *Curr. Anthropol.* **34**, 483–496.
- VON HAESLER, A., SAJANTILA, A. & PÄÄBO, S. (1996). The genetical archaeology of the human genome. *Nature Genetics* **14**, 135–140.
- HASEGAWA, M., DI RIENZO, A., KOCHER, T. D. & WILSON, A. C. (1993). Toward a more accurate time scale for the human mitochondrial DNA tree. *J. Mol. Evol.* **37**, 347–354.
- HOWELL, N., KUBACKA, I. & MACKEY, D. A. (1996). How rapidly does the human mitochondrial genome evolve? *Am. J. Hum. Genet.* **59**, 501–509.
- KOLMAN, C. J., SAMBUUGHIN, N. & BERMINGHAM, E. (1996). Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142**, 1321–1334.
- KRINGS, M., STONE, A., SCHMITZ, R. W., KRAINITZKI, H., STONEKING, M. & PÄÄBO, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 1–20.
- MARTÍN DEL MOLINO, A. (1993). *Los Bubis. Ritos y creencias*. Madrid, Labrys 54.
- MOUNTAIN, J. L., HEBERT, J. M., BHATTACHARYYA, S. S., UNDERHILL, P., OTTOLENGHI, C., GADGIL, M. & CAVALLI-SFORZA, L. L. (1995). Demographic history of India and mitochondrial DNA sequence diversity. *Am. J. Hum. Genet.* **56**, 979–992.
- NEI, M. & TAJIMA, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145–163.
- NEI, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.

- PÄÄBO, S. (1996). Mutational hot spots in the mitochondrial microcosm. *Am. J. Hum. Genet.* **59**, 493–496.
- PARSONS, T. J., MUNIEC, D. S., SULLIVAN, K., WOODDYATT, N., ALLISTON-GREINER, R., WILSON, M. R., BERRY, D. L., HOLLAND, K. A., WEEDN, V. W., GILL, P. & HOLLAND, M. M. (1997). A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* **15**, 363–368.
- PIERCY, R., SULLIVAN, K. M., BENSON, N. & GILL, P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Leg. Med.* **106**, 85–90.
- PINTO, F., GONZÁLEZ, A. M., HERNÁNDEZ, M., LARRUGA, J. M. & CABRERA, V. (1996). Genetic relationship between the Canary islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann. Hum. Genet.* **60**, 321–330.
- RICHARDS, M., CÔRTE-REAL, H., FORSTER, P., MACAULAY, V., WILKINSON-HERBOTS, H., DEMAINE, A., PAPIHA, S., HEDGES, R., BANDELT, H.-J. & SYKES, B. (1996). Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**, 185–203.
- ROGERS, A. R. & HARPENDING, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569.
- SAITOU, N. & NEI, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- SAJANTILA, A., LAHERMO, P., ANTINEN, T., LUKKA, M., SISTONEN, P., SAVONTAUS, M.-L., AULA, P., BECKMAN, L., TRANEBJAERG, L., GEDDE-DAHL, T., ISSSEL-TARVER, L., DI RIENZO, A. & PÄÄBO, S. (1995). Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Research* **5**, 42–52.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular cloning. A laboratory manual*. Second Edition. New York: Cold Spring Harbor Laboratory Press.
- SANTOS, M., WARD, R. H. & BARRANTES, R. (1994). mtDNA variation in the Chibcha Amerindian Hueter from Costa Rica. *Hum. Biol.* **66**, 963–977.
- SLATKIN, M. & HUDSON, D. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562.
- SOODYALL, H., VIGILANT, L., HILL, A. V., STONEKING, M. & JENKINS, T. (1996). mtDNA Control-Region sequence variation suggests multiple independent origins of an 'Asian-Specific' 9-bp deletion in sub-Saharan Africans. *Am. J. Hum. Genet.* **58**, 595–608.
- STONEKING, M. (1993). DNA and recent human evolution. *Evol. Anthropol.* **2**, 60–73.
- TAJIMA, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- TISHKOFF, S. A., DIETZSCH, E., SPEED, W., PAKSTIS, A. J., KIDD, J. R., CHEUNG, K., BONNÉ-TAMIR, B., SANTACHIARA-BENERECETTI, A. S., MORAL, P., KRINGS, M., PÄÄBO, S., WATSON, E., RISCH, N., JENKINS, T. & KIDD, K. K. (1996). Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380–1387.
- VARA, S. & BOLEKIA, J. (1993). Bioko, tierra de los Bubis. *Africa 2000* **20**, 16–21.
- VIGILANT, L., PENNINGTON, R., HARPENDING, H., KOCHER, T. D. & WILSON, A. C. (1989). Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci. USA* **86**, 9350–9354.
- VIGILANT, L., STONEKING, M., HARPENDING, H., HAWKES, K. & WILSON, A. C. (1991). African populations and the evolution of mitochondrial DNA. *Science* **253**, 1503–1507.
- WARD, R. H., FRAZIER, B. L., DEW-JAGER, K. & PÄÄBO, S. (1991). Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* **88**, 8720–8724.
- WAKELEY, J. (1993). Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. *J. Mol. Evol.* **37**, 613–623.
- WATTERSON, G. A. (1975). On the number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* **7**, 256–276.
- WATSON, E., BAUER, K., AMAN, R., WEISS, G., VON HAESSELER, A. & PÄÄBO, S. (1996). mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.* **59**, 437–444.