Ann. Hum. Genet. (2002), **66**, 49–60 © *University College London*
DOI: 10.1017/S0003480001001002 *Printed in the United Kingdom*

Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade

A. BREHM[1], L. PEREIRA[2,3], H.-J. BANDELT[4], M. J. PRATA[2,3] AND A. AMORIM[2,3]
[1] *Centre of Biological and Geological Sciences, University of Madeira, Campus of Penteada, 9000 Funchal, Portugal*
[2] *IPATIMUP (Instituto de Patologia e Imunologia Molecular da Universidade do Porto) R. Dr. Roberto Frias s/n 4200 Porto, Portugal*
[3] *Faculdade de Ciências da Universidade do Porto. Praça Gomes Teixeira. 4099-002 Porto, Portugal*
[4] *Fachbereich Mathematik, Universität Hamburg, Hamburg, Germany*

(Received 11.7.01. Accepted 28.11.01)

SUMMARY

In order to study the matrilineal genetic composition in Cabo Verde (Republic of Cape Verde), an archipelago that used to serve as a Portuguese entrepôt of the Atlantic slave trade, we have analysed a total of 292 mtDNAs sampled from the seven inhabitated islands for the hypervariable segment I (HVS-I) and some characteristic RFLPs of the coding regions. The different settlement history of the northwestern group of the islands is well reflected in the mtDNA pool. The total Cabo Verde sample clearly displays the characteristic mitochondrial features of the Atlantic fringe of western Africa and testifies to almost no mitochondrial input from the Portuguese colonizers.

INTRODUCTION

The Portuguese discovered the Cabo Verde islands in 1460–1462, although there have been sporadic reports that at least the island of Sal (referred to as Aulil or Ulil) was already known to the Moors, and perhaps also to Wolof, Serer, and Lebu fishermen who took salt from there (Carreira, 1983). In any case, the islands were uninhabited and the first settlement occurred in Santiago in 1461–1462 and then in the nearby island of Fogo. The first settlers were an assortment of Portuguese nobles, Genovese adventurers, exiles and convicts, as well as Jews. No women were among the settlers, who went to Cabo Verde without their families and formed liaisons with slave women, thus creating a new class of individuals, the 'mulattos' or 'crioulos', who would become the majority of the population (Godinho, 1965). Some of the first slaves brought to the islands were Africans acquired along the Senegal River by the Moors, sold to the Wolof and then resold to the Portuguese (Barry, 1998). There are records that among these first slaves there were also Guanches from the Canary Islands. Together with some banished Portuguese, the mulattos played a pivotal role as intermediate slave traders on the coast of Guinea known as Senegambia. In 1466 King Afonso V granted the settlers of Santiago with a special permit to trade Africans from the Guinean coast, excluding Arguin (on today's Mauritanian coast), and sell them wherever they wanted. The population grew quickly, mainly due to the slaves captured or traded on the coast ('resgatados') (Carreira, 1983). By 1470 the Portuguese had already explored the Gulf of Guinea and São Tomé and Príncipe islands (Russell-Wood, 1998; Barry, 1998). Soon Santiago became an obligatory stop for ships going to the coast of Angola, São Tomé Island, and later Brazil and the Antilles. With

Correspondence: Prof. Dr. Hans-Jürgen Bandelt, Fachbereich Mathematik der Universität, Bundesstr. 55, 20146 Hamburg, Germany. Fax: +49-40-42838-5190 and +49-441-883424. E-mail: bandelt@math.uni-hamburg.de and bandelt@yahoo.com

Table 1. *Population data of the Cabo Verde Islands*

Code	Island	Relative location	First settlement	Population census 1999
BRA	Brava	Windward, SE	Late 17th century	6975
FOG	Fogo	Windward, SE	1470–1480	33902
SAN	Santiago	Windward, SE	1462–1466	175691
MAI	Maio	Windward, SE	Early 17th century	4969
BOA	Boavista	Leeward, SE	Early 17th century	3452
SAL	Sal	Leeward, SE	Middle 20th century	7715
NIC	São Nicolau	Leeward, NW	Middle 17th century	13665
VIC	São Vicente	Leeward, NW	Middle 19th century	51277
ANT	Santo Antão	Leeward, NW	Late 16th century	43845

Source: Carreira (1983), National Institute of Statistics, Praia, Cabo Verde.

the construction of São Jorge da Mina Fortress in the Gulf of Guinea (1485), Cabo Verde served as a central pivot for a complex transatlantic commercial slave network (Russell-Wood, 1998).

The European population in Cabo Verde has never been numerous; Santiago and Brava in 1582 had little more than 100 European men, while there were more than 13,000 African slaves. During most of the 16th century and the first half of the 17th century, most of the slaves brought to Santiago were en route to Brazil and the West Indies (Antilles and Central America) (Godinho, 1965; Russell-Wood, 1998). Many of the slaves were shipped to Lisbon, and fewer to the Canary Islands and Sevilha, the remaining ones being used to populate the southeastern islands. The origin of slaves going to the West Indies from the Guinean coast through Cabo Verde is well documented: 1503, first arrival of Guineans to Santo Domingo; 1521, Cuba; 1620, (US state of) Virginia; and 1650, French Antilles (Russell-Wood, 1998). It was only just prior to or during the 17th century that the northwestern islands of Santo Antão and São Nicolau received the first settlers, composed of fugitive slaves from the southeastern islands. The European settlers were very few, and this group of islands never harboured a large number of people until the 18th century (see Table 1). During the past five centuries the islands suffered from several periods of drought and famine, which substantially decreased the number of inhabitants and promoted some inter-island migrations (1810–1814). About the middle of the 19th century, Cabo Verdeans from Fogo and Brava started to immigrate en masse to New England, recruited by the whaling ships from Massachusetts and Rhode Island in the USA.

The main aim of the present work is to analyse the mtDNA pool of the present-day Cabo Verde population, which would be expected to mirror the mtDNA pool of West Africans from Senegambia (south of the Senegal river as far as Serra Leone). Some data have been gathered concerning mtDNA haplotypes of Senegalese (Graven *et al.* 1995; Rando *et al.* 1998), but these do not cover, by far, the ethnic groups that populated Cabo Verde. It will be interesting to see whether there exist mitochondrial differences between the two main groups of islands, populated at different periods of time, given that only the slaves of the southeastern group were in transit to other destinations. mtDNA analysis will also give an idea of the extent of the maternal European input brought to the archipelago, today regarded as a typically creole population. Finally, the mtDNA haplotypes found in Cabo Verde will help to trace the matrilineal origins of Afro-American populations.

Material and methods

Sampling

A total of 292 blood samples were collected from unrelated Cabo Verdeans whose maternal ancestors were known to be originally from one of the nine islands of the archipelago for at least three generations. The samples were collected in military camps with the permission of the Cabo

Table 2. *HVS-I Haplotypes and their Regional Distribution in the Cabo Verde Islands*

	Island distribution										Characteristic RFLP sites												
	SE						NW																
Haplotype	BRA	FOG	SAN	MAI	BOA	SAL	NIC	VIC	ANT	HVS-I Motif (minus 16000)	3 2 2 e	6 7 9 c	2 3 4 9 9 j	4 1 5 7 a	4 5 7 7 q	5 2 6 0 b	5 5 8 4 a	10 8 4 l	10 9 4 c	10 3 7 a	10 3 1 z	14 8 7 j	16 3 9 b	Haplogroup
CV1			2							129 148 168 172 187 188G 189 223 230 311 320														L1a
CV2			2							093 126 187 189 223 256 264 270 278 293 311														L1b
CV3			2							187 189 223 264 270 278 293T 311														L1b
CV4			2							126 187 189 223 256 264 278 293 311														L1b
CV5			1							126 187 189 223 270 278 311														L1b
CV6			1	1						104 187 189 223 270 278 289 293 311														L1b
CV7			1							126 187 189 223 264 270 278 293 311														L1b
CV8								3	4	126 187 189 223 264 270 278 293 311 362														L1b
CV9									1	126 145 187 189 223 264 270 278 293 311														L1b
CV10									1	126 187 189 223 239 254 264 270 278 293 311														L1b
CV11		2								093 126 187 189 223 239 264 270 278 293 311														L1b
CV12		1								114G 126 187 189 223 264 270 274 278 293 311														L1b
CV13		1								126 187 189 223 264 270 274 278 293 311														L1b
CV14							1			086 129 187 189 219 223 261 278 293 311 360 362														L1c
CV15			1							093 129 169 172 187 189 223 239 261 278 293 311 360														L1c
CV16	1		1	1				5	7	(017) 129 163 187 189 223 278 293 294 304 311 360														L1c
CV17		1								086 129 169 172 183C 189 192 223 261 278 293 311 360														L1c
CV18			1							086 129 169 172 183C 189 223 261 278 293 311 360														L1c
CV19			1							129 183C 189 223 235 274 278 293 294 311 360														L1c
CV20			1							180del 223 256 278 294 309 368 (390)														L2a
CV21		1								189 192 223 278 291 294 311														L2a
CV22			1							182C 183C 189 223 278 294 362 (390)														L2a
CV23			1							111G 223 278 294 (390)														L2a
CV24		1	1							093 189 223 264 278 294 309 (390)														L2a
CV25			2							189 192 223 278 294 309 (390)														L2a
CV26			1							183C 223 278 294 362 (390)														L2a
CV27				1						093 223 274 278 294 309														L2a
CV28		1								093 223 278 290 294 309 362 (390)														L2a
CV29									1	111 278 294 309 (390)														L2a
CV30			2							189 223 278 294 309 (390)														L2a
CV31		1								111 223 278 294														L2a
CV32							1			223 278 294 309				−										L2a
CV33			1							179 223 278 290 294 309														L2a
CV34			1							093 223 278 290 294 309														L2a
CV35			1							223 278 290 294 309														L2a
CV36			1							223 278 294 309 368														L2a
CV37		3								086 223 278 294 309 (390)														L2a
CV38	1									140 223 278 290G 294 355														L2a
CV39	3	2	4			1	3	4	14	111 223 278 294 309 (390)														L2a
CV40			5							093 148 223 278 294 355 (390)														L2a
CV41			1							114A 129 213 223 278 355 362														L2b
CV42			1							114A 188 189 223 278 304 311														L2b
CV43			2							114A 129 169 213 223 278 362														L2b
CV44	1									114A 213 223 278 362														L2b
CV45		1	6							114A 129 213 223 278 (390)				+										L2b
CV46									1	093 223 278 (390)	−	−												L2c
CV47								2	10	177 223 278 311	−	−											+	L2c
CV48			1							086 223 264 278 (390)	−	−												L2c
CV49			3							126 189 223 269 278 293 (390)	−	−												L2c
CV50			1							173 217 223 254 264 278	−	−											+	L2c
CV51			1							093 223 264 278	−	−											+	L2c

Table 2. (*Cont.*)

	Island distribution										Characteristic RFLP sites													
	SE						NW																	
Haplotype	BRA	FOG	SAN	MAI	BOA	SAL	NIC	VIC	ANT	HVS-I Motif (minus 16000)	3 2 2 e	6 7 9 c	2 4 9 j	4 5 7 a	4 1 7 q	5 2 6 b	5 5 8 a	10 8 4 l	10 3 4 c	10 3 7 a	10 8 1 z	14 8 9 j	16 3 9 b	Haplogroup
CV52									1	093 223 278 318 (390)	–	–												L2c
CV53							1		1	223 264 278 (390)	–	–												L2c
CV54		1								051 217 223 264 278	–	–											+	L2c
CV55		1	1							192 223 278 (390)	–	–												L2c
CV56	1		6			1		1	1	223 278 (390)	–	–												L2c
CV57			2							084 093 220 223 264 278 311	–	–											+	L2c
CV58			2							172 223 241 278	–	–											+	L2c
CV59		1	1						1	223 278 311 (390)	–	–												L2c
CV60			1							223 261 278	–	–											+	L2c
CV61			2							051 223 278 (390)	–	–												L2c
CV62			2							223 278 320	–	–	–									+	+	L2c
CV63			1							111A 145 184 223 239 278 292 311				–										L2*
CV64			1							189 223 259 274 278	+	+											+	L2*
CV65		1	5	3				1		124 223 278 362														L3b
CV66			2							124 183C 189 214 223 278 362														L3b
CV67			2			1			1	086 124 223 278 362														L3b
CV68			1							124 183C 189 209 223 278 362														L3b
CV69			1							124 183C 189 214 223 278 325 362														L3b
CV70			1							048 124 223 278 362														L3b
CV71		1	3							124 183C 189 223 278 362 (390)														L3b
CV72			1							124 223 278 355 362														L3b
CV73			1							124 278 362														L3b
CV74		1	1							124 223 234 278 362														L3b
CV75			1							086 124 223 278 311 362								+						L3b
CV76		1								124 212 223 278 311 362								+						L3b
CV77		1								189 223 278 362								+			–			L3b
CV78			1							051 223 278 318 362	+	+						+			–		–	L3b
CV79		1								124 223 256														L3d
CV80			1							124 223 245 319														L3d
CV81		1								124 223 288														L3d
CV82			2							124 223 257														L3d
CV83			1	1					1	093 124 223														L3d
CV84			1							111 124 223 291														L3d
CV85			1						1	124 223 311														L3d
CV86								1		111 124 189 223														L3d
CV87	1									086 124 223														L3d
CV88			3							124 223								–						L3d
CV89			1						2	111 124 223														L3d
CV90	1							1		124 148 223 293 311 362								–						L3d
CV91			2							172 183C 189 223 320														L3e2
CV92			1							086 223 320	+	+	+									–		L3e2
CV93		1	3				1		1	223 320	+	+	+			–	+					–		L3e2
CV94			1							093 172 189 223 320												+		L3e2
CV95			1							126 172 183C 189 223 320												+		L3e2
CV96					1					051 223 264 299 311	+	+	+			+	–							L3e4
CV97			1					2	9	051 223 264 299						+	–							L3e4
CV98			1							051 093 223 247 264 311	+	+	+			+	–							L3e4
CV99			1							051 093 223 264 311						+	–							L3e4
CV100			2					4	11	051 223 264						+	–							L3e4

Table 2. (*Cont.*)

	Island distribution										Characteristic RFLP sites													
	SE						NW																	
Haplotype	BRA	FOG	SAN	MAI	BOA	SAL	NIC	VIC	ANT	HVS-I Motif (minus 16000)	3222 e	679 c	2349 j	4157 a	4577 q	5260 b	5584 a	10084 l	10394 c	10397 a	10871 z	14869 j	16390 b	Haplogroup
CV101		1								117 129 223 205A 311 372									+	–	–			L3*
CV102		1								129 154 223 256A 311 362	–	–	–						+	–	–			L3*
CV103			1							129 223 256A 311 362	–	–	–						+	–	–			L3*
CV104									2	129 223 295	–	–	–						+	–	–			L3*
CV105			1							209 223 292 311									+	–	–			L3*
CV106			1							223 355	–	–	–						+	–	–			L3*
CV107			1						2	129 223	–	–	–						+	–	–			L3*
CV108			1							223 290 355									+	–	–			L3*
CV109			1							093 209 223 292 311									+	–	–			L3*
CV110			1						1	172 219 278														U6a
CV111			1							172 183C 189 219 274 278														U6a
CV112			1							172 182C 183C 189 219 278														U6a
CV113			1							111 172 183C 189 219 278														U6a
CV114		1	3							172 183C 189 219 278														U6a
CV115			1							189 192 234 270 320														U5
CV116			1							145 222					–									H
CV117									1	111 223									–	–	+			N*
CV118			2							183C 189 223 278 311											+		–	X
CV119			1							126 209 294 296 304														T
CV120		1								126 362														pre-HV
CV121									2	298					–									pre*V
CV122					1					182C 183C 189 217 247 261 (390)														B
	9	30	133	7	2	3	7	24	77															

Note: Sequence positions are numbered according to the reference sequence CRS (Anderson et al. 1981).

Numbers in the motifs refer to transitions unless a suffix indicates a deletion (del) or a transversion to that nucleotide. Numbers in brackets are outside the common reading frame. The following single-letter codes for restriction enzymes are used: a = *Alu*I; b = *Ava*II; c = *Dde*I; e = *Hae*III; j = *Mbo*I; l = *Taq*I; q = *Nla*III; z = *Mnl*I.

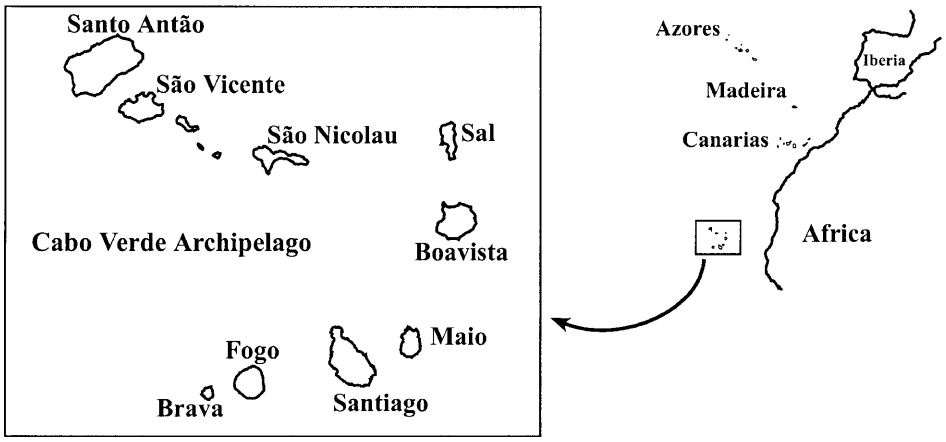

Fig. 1. Geographic map of the Cabo Verde archipelago.

Verde Joint Chiefs of Staff. Every participant gave his consent in an individual interview after a detailed explanation of the project. Sample sizes and origins (along with additional information) are specified in Tables 1 and 2. On the basis of geography (see Fig. 1), the archipelago is divided into two groups: Sotavento, the leeward group, comprises Brava, Fogo, Santiago, and Maio, whereas Barlavento, the windward group, consists of Boavista, Sal, São Nicolau, São Vicente, and Santo Antão. On historical grounds, however, the following division is preferable: the southeastern group (SE), which includes the leeward group as well as the (late setttled) islands Boa Vista and Sal, is contrasted to the northwestern group (NW), constituted by the remaining three windward islands.

HVS-I sequencing

The leukocyte fraction of whole blood was used to extract DNA by the Chelex method (Lareu *et al.* 1994). The mtDNA hypervariable segment I (HVS-I) of the control region was amplified using primers L15996 and H16401 of Vigilant *et al.* (1989) with the conditions specified therein. PCR products were sequenced with the same primers used for amplification in an ABI-DNA Automated Sequencer (AB Applied Biosystems) according to the manufacturer instructions. The sequences obtained were aligned with the Cambridge reference sequence (CRS, Anderson *et al.* 1981). All sequences could be read unambiguously between 16025 and 16389.

RFLP testing

All digestions were carried out according to the manufacturer's instructions (Promega). Digested fragments were resolved in gels with specific acrylamide percentages depending on the fragment sizes, and visualized after silver staining. The following polymorphic restriction sites were screened: 322 *Hae*III, 679 *Dde*I, 2349 *Mbo*I, 4157 *Alu*I, 4577 *Nla*III, 5260 *Ava*II, 5584 *Alu*I, 10084 *Taq*I, 10394 *Dde*I, 10397 *Alu*I, 10871 *Mnl*I, 14869 *Mbo*I, and 16390 *Ava*II. The polymorphic position 10873 was determined by digesting a fragment amplified with primers H10672–H10683 and L10959–L10978 (Toomas Kivisild, pers. comm.). Positions 5260 and 5584 were checked by amplifying a fragment with a newly designed primer, H4790–H4810, and the known L5898–L5917 (Torroni *et al.* 1996), thus diminishing fragment length and avoiding multiple fragments.

Phylogeographic analysis

For the most part, we used the same classification scheme for African and European mtDNAs as Quintana-Murci *et al.* (1999), Alves-Silva *et al.* (2000), and Richards *et al.* (2000). In particular, haplogroups L1a, L1b, L1c, L2, L3b, L3d, L3e, and the paragroup L3* (which constitutes the paraphyletic cluster harbouring all other L3 mtDNAs not grouped into the Eurasian haplogroups M or N) are specific to sub-Saharan

Africa (albeit with minor diffusion into geographically proximate areas). Haplogroup U6 (Rando *et al.* 1998; Macaulay *et al.* 1999) is autochthonous to North Africa (with partial diffusion into the Sahel zone). From Chen *et al.* (2000) we adopted a subdivision of haplogroup L2 into L2a (characterized by 16294), L2b (characterized by 16114A and 16213), and L2c (characterized by 325 in HVS-II), and the remainder L2*. Here, and in what follows, characteristic mutations are indicated by the site relative to CRS in the case of transitions; otherwise, a suffix specifies a transversion. Haplogroup L3e is partitioned (with the characteristic sites in brackets) into L3e1 (16327), L3e2 (16320), L3e3 (16265T), and L3e4 (16264), following Bandelt *et al.* (2001). mtDNA haplogroup status was primarily assessed via HVS-I motifs and subsequently, in potentially ambiguous cases, confirmed by partial RFLP analysis.

For each population, the (relative) frequency vectors ('haplogroup profiles') of the haplogroups under consideration were then recorded and subjected to principal component (PC) analysis, by using the program popstr (ftp://mombasa.anthro.utah.edu/pub/).

RESULTS

Haplogroup profiles

The 122 different haplotypes of the Cabo Verde sample represent all major haplogroups observed so far in western Africa (Table 2). More than 93% of the sampled mtDNAs belong to sub-Saharan African haplogroups, which are well characterised (L1a, L1b, L1c, L2, L3b, L3d, and L3e), except for a few minor ones lumped together in the paragroup L3*. Only 3% are members of the North African haplogroup U6, so that the total African fraction of the mtDNA pool amounts to nearly 97%. Two mtDNAs (1%) are from European haplogroups (H and U5) but belong to specific clades distributed at very low frequencies across western Africa (Rando *et al.* 1998, 1999). Seven mtDNAs (2%) are classified as N*, X, T, pre-HV, and pre*V, and could have arrived from northwestern Africa or the Iberian Peninsula. A single mtDNA, an erratic in the Cabo Verde mtDNA pool, belongs to the Asian haplogroup B and, quite surprisingly, has the so-called Polynesian motif. According to the family's oral tradition of this subject, the great-grandmother (in the maternal line) came from North America to Boavista and was assumed to be a Navajo. In view of the mtDNA a Hawaiian origin, for example, would seem more plausible.

It is instructive to compare the haplogroup profiles of Cabo Verde with those of other (sufficiently large) samples from Africa and America. The available data sets on Senegal (Mandenka: Graven *et al.* 1995; mixed, mainly Wolof and Serer: Rando *et al.* 1998), Niger/Nigeria (Songhai, Tuareg, Yoruba, Hausa, Fulbe, and Kanuri: Watson *et al.* 1997), Dominican Republic (A. Torroni, unpubl. data), Brazil (Alves-Silva *et al.* 2000), Mozambique (Pereira *et al.* 2001), and !Kung/Khwe (Chen *et al.* 2000) are taken for comparison. Note that the distinction between L2c and L2* in the samples from Senegal (mixed) and Niger/Nigeria is problematic because of the lack of sufficient RFLP/HVS-II information. Table 3 lists the resulting haplogroup profiles. Haplogroups L1a, L3e1, and L3e3 are either absent or have very low frequency in Cabo Verde and Senegal, but are quite frequent (29% together) in both Mozambique and the African fraction of the Brazilian pool. On the other hand, L1b, L2c, and L3e4 are nearly absent in the latter two mtDNA pools and among !Kung/Khwe, but are frequent to dominant in the pools of Cabo Verde (35%) and Senegal (Mandenka: 61%). Specifically, highest percentages are reached for haplogroup L3e4 in NW Cabo Verde , for L3b in SE Cabo Verde and Niger/Nigeria, L1b and L2c in Senegal (Mandenka), L2b in Senegal (mixed), L3d and U6b in the Dominican Republic, L3* in Niger/Nigeria, L1c, L2*, L3e2, L3e3, and U6a in the African fraction of the Brazilian pool, L1a, L2a, and L3e1 in Mozambique, and for L1* (embracing the Khoisan-specific haplogroups) in the !Kung/Khwe.

The PC plot (not shown) for the raw haplogroup profiles from Table 3 is dominated by the

Table 3. *Haplogroup profiles in Cabo Verde* (*SE*, *NW*), *Senegal* (*Mandenka*, *mixed*), *Niger/Nigeria*, *Dominican Republic* (*Caribbean*), *Brazil*, *Mozambique*, *and !Kung/Khwe*

Sample		Haplogroup (%)																		
Origin	Size	L1a	L1b	L1c	L1*	L2a	L2b	L2c	L2*	L3b	L3d	L3e1	L3e2	L3e3	L3e4	L3*	U6a	U6b	Eurasian	American
Cabo Verde SE	184	1	8	4	0	20	7	15	1	16	8	0	5	0	3	4	4	0	4	0
Cabo Verde NW	108	0	8	12	0	21	0	18	0	2	6	0	2	0	24	4	1	0	3	0
Senegal (Mand.)[1]	119	2	21	4	0	12	3	36	1	5	11	0	1	0	4	0	1	0	0	0
Senegal (mixed)[2]	121	0	17	3	0	21	13	7	2	12	6	1	2	1	2	8	1	1	3	0
Niger/Nigeria[3]	160	2	14	3	0	23	1	4	1	16	7	1	9	3	0	9	3	1	6	0
Dominican Rep.[4]	127	3	3	7	0	15	7	13	2	9	12	2	10	1	1	2	0	2	2	9
Brazil[5]	247	3	1	5	0	3	1	1	1	0	2	4	3	1	0	1	2	0	39	33
Mozambique[6]	109	15	1	5	9	43	2	1	0	4	2	12	3	2	1	2	0	0	0	0
!Kung/Khwe[7]	99	5	1	0	62	4	5	0	0	0	8	0	7	0	0	8	0	0	0	0

[1]Graven *et al.* (1995); [2]Rando *et al.* (1998); [3]Watson *et al.* (1997); [4]Torroni, unpubl. data; [5]Alves-Silva *et al.* (2000); [6]Pereira *et al.* (2001); [7]Chen et al. (2000).

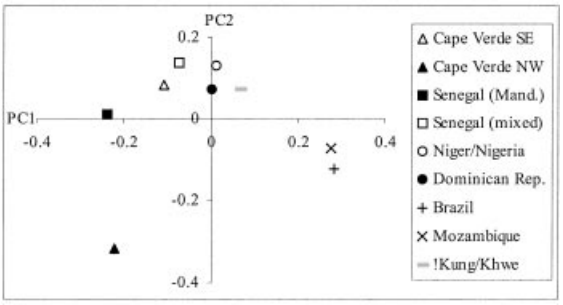

Fig. 2. PC map for the sub-Saharan African fractions of several African and American mtDNA pools (input vectors truncated to the coordinates representing monophyletic groups).

Table 4. *Diversity in the Cabo Verde Islands*

Island(s)	$1-\Sigma_i x_i^2$
Cabo Verde SE	0.983
Santiago	0.983
Fogo	0.951
Cabo Verde NW	0.900
Santo Antão	0.900
São Vicente	0.865

trivial signal (mainly incurred by L1*) that distinguishes the !Kung/Khwe from the other populations. Moreover, the Eurasian/American mtDNA contribution to the Brazilian and Dominican populations would give a minor signal that is not relevant for their affinities to the populations of the African continent. We have therefore restricted our main analysis to the sub-Saharan fractions (viz. columns L1a to L3* of Table 3) and the corresponding relative haplogroup frequencies. For better focus, we have then dropped the coordinates for the paragroups L1*, L2*, and L3* (which are potentially mixed bags of different clades) from the input vectors for the PC analysis. Figure 2 displays the resulting two-dimensional PC plot. The first component reflects a clear NW-SE cline, where NW Cabo Verde is paired with Senegal (Mandenka), SE Cabo Verde with Senegal (mixed), the Dominican Republic with Niger/Nigeria, and Brazil with Mozambique. This gives some hints at the ancestry of the populations that were ultimately created by slave trade. Not surprisingly, the haplogroups that contribute most to the first PC are L1b, L2c, and L3e4 on the negative side and L1a and L3e1 on the positive side. For the second PC, L3e4 and L1c have the strongest negative impact, which explains the anomalous placement of NW Cabo Verde together with Brazil and Mozambique on the same halfplane of the PC map. The third PC (not shown) highlights the distinction between Brazil and Mozambique (mainly incurred by L2a and L1c).

Major haplotypes

We regard a haplotype as major when its sample frequency exceeds 3%. Seven haplotypes, covering together 37% of the mtDNA pool, then qualify as major: CV16 (L1c), CV39 (L2a), CV47 (L2c), CV56 (L2c), CV65 (L3b), CV97 (L3e4), and CV100 (L3e4). Except for CV47, which is restricted to São Vicente and Santo Antão, these haplotypes were found in at least three islands each. CV56 and CV65, being the probable ancestral types of their respective haplogroups, are widespread in Africa, whereas CV97 is typically found in Senegal. CV16 has hitherto been observed in a single Lebu individual from Senegal (Rando *et al.* 1998), and CV39, the most frequent haplotype (11%), in a single São Tomean (Mateu *et al.* 1997). The other two major haplotypes have not yet been sampled elsewhere.

Each of these seven haplotypes also occur at frequency $> 3\%$ within the two separate mtDNA pools, SE and NW. In addition, the SE and NW pools each harbour one further haplotype (each sampled 7 times) that occurs at $> 3\%$ in either pool: CV45 (L2b) in SE and CV8 (L1b) in NW. These nine major haplotypes make up 22% of the SE pool but 75% of the NW pool. This contrast is also well reflected in the diversity values (average 'heterozygosity'); see Table 4. For the SE pool and Santiago alone, the values are in the general range of mainland Senegalese populations (cf. Pereira *et al.* 2001), whereas the diversity of the NW pool, and the islands of São Vicente and Santo Antão separately, are comparable to the diversity observed in islands such as La Gomera and La Palma (Rando *et al.* 1999).

For comparison, we use the 3% criterion for major haplotypes also in other African mtDNA pools, where we would require sizes $\geqslant 67$ in order

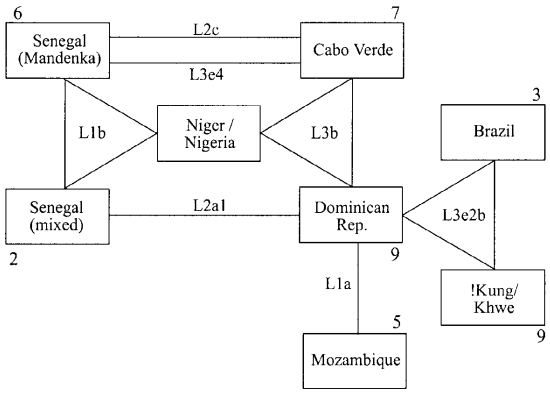

Fig. 3. Major haplotype sharing between mtDNA pools (African fractions): each line/triangle signifies one major haplotype in common (indicated by the respective haplogroup to which it is ancestral); numbers of major haplotypes in the pools are indicated.

to have a major haplotype sampled at least three times. We investigate the African fractions (to which the 3% threshold is applied) of the mtDNA pools listed in Table 3. Then, in all cases, the numbers of major haplotypes range between 2 and 9. When we focus on matched major haplotypes across populations, we find seven shared ones (Fig. 3), which are indeed widespread (though not ubiquitous) in Africa and constitute potential ancestral types of the haplogroups L1b (with 16293), L2c, L3b, and L3e4, or of the pronounced clades L2a1 (defined by 16309) and L3e2b (defined by 16172 and 16189; Bandelt *et al.* 2001). Note that clade L2a1 is, however, difficult to recognize because 16309 seems to be prone to back mutation. A further data set from Africa that could be taken for comparison here is the one from Sierra Leone, for which, however, only very limited information has been released (Budowle *et al.* 1999): two major haplotypes of frequency $>4\%$ were reported, one is the ancestral type of L2c shared with Cabo Verde and Senegal, and the other is a L3* haplotype. When we repeat the same analysis with the Cabo Verde sample split into the SE and NW pools, the picture does not change in that either local pool behaves exactly as the total pool in comparison with the other populations, but as many as six major haplotypes link the two SE and NW pools, thus clearly pointing to a common regional ancestry.

To see whether at least the nine major haplotypes in the SE and NW pools could have been carried to Cabo Verde from the African mainland, we tentatively calculated the estimated (average) age of the founder event. We focused on 1-step descendants exclusively because it would be very rare to observe more than one mutation in a line of descent within less than one millennium. Since CV56 and CV65 constitute the ancestral types of haplogroups L2c and L3b (highly frequent in Senegal), we could not predict which 1-step descendants of those two haplotypes originated in Cabo Verde. For the remaining seven major haplotypes, we found only two 1-step descendants (one in SE, one in NW): this yields an average transitional distance of 2/103 to the seven founders and their descendants. According to Saillard *et al.* (2000), this translates into an age of 390 ± 280 years – in perfect agreement with the historical records.

Discussion

In our mtDNA sample from the seven islands of Cabo Verde, ten out of 292 mtDNAs are of non-African origin: one Asian mtDNA (as explained above) and three mtDNAs that are of west Eurasian origin (in the Holocene) but most probably entered the sub-Saharan mtDNA pool (via North Africa) before the slave trade began. This leaves only six mtDNAs (2%) that could have been directly transmitted from European female immigrants, thus implying a minute female contribution from Europe to the peopling of the present-day population of Cabo Verde. This is in complete agreement with historical sources that strongly limit the participation of European women among the colonizers. Although there is a measurable mtDNA input from North Africa, as evidenced by U6a haplotypes, it is rather likely that the three U6a haplotypes not yet found in Senegal or Mauritania are also present in some Senegalese and Guinean populations. In fact, the first slaves taken by the Portuguese were from the Mauritanian coast ('white slave trade') but most of those were brought to the Iberian Peninsula and not to

Cabo Verde. A Guanche mtDNA input from the Canary Islands is not discernible in our sample from Cabo Verde in view of the lack of haplogroup U6b and any founder type reconstructed for the Guanche mtDNA pool (Rando *et al.* 1999).

The differences in haplotype frequencies (as well as some haplogroup frequencies) between SE Cabo Verde and NW Cabo Verde are striking, but parallel the significant serological differences found between Santiago and Santo Antão/São Vicente (Lessa & Ruffié, 1960). Nevertheless, the major haplotypes in both regions are nearly the same, and no haplogroups were found in the NW group that were not already sampled in the SE group of islands. On the other hand, the former islands seem to be missing three minor haplogroups of the latter islands. The evident reduction in NW diversity indicates rather strong founder effects. This agrees with the historical records, stating that fugitive slaves from the more populated islands (e.g. Santiago) escaped to the NW islands. These escapees may have formed whole clans (organized according to the place of origin of their ancestors), thus having enforced drift effects. It is tempting to infer from the haplogroup profiles that a considerable number of those escapees had their maternal ancestry in the Mandenka. It is plausible that all the NW major haplotypes were already involved in the pioneer settlement of Santo Antão. As to the mtDNAs observed, São Vicente is almost indistinguishable from the earlier settled Santo Antão. Bottlenecks in the course of the famines (in the 19th century and even later) may have had a stronger effect in the NW pool compared to the SE pool.

The haplotypes found in just one ethnic group (Mandenka) of Senegal do not show very similar frequencies to those found in Cabo Verde. Clearly a better and more thorough mtDNA survey of mainland West Africa is needed to understand fully the haplotype composition of Cabo Verde. Comparing the few available mtDNA samples of African populations, we find characteristic broad regional imprints in both the haplogroup profiles and the short-lists of major haplotypes. In particular, haplogroups L2c and L3e4 are conspicuous in the Atlantic fringe of western Africa, thus allowing the recognition of (matrilineal) Senegambian influence in the Americas or remote African areas. For instance, the L3e4 haplotype CV98 matches a haplotype from São Tomé (Mateu *et al.* 1997), and also the L2c haplotypes CV51 and CV54 were sampled in São Tomé. Moreover, the most frequent Cabo Verdean haplotype (CV39) has been found, so far, only there. This points to some direct link between Cabo Verde and São Tomé, which is supported by historical sources. For instance, it has been reported that between 1950 and 1970 more than 30,000 Cabo Verdeans emigrated to São Tomé (Carreira, 1983).

The authors are grateful for the precious help of the Cabo Verde Joint Chiefs of Staff, ICCTI (Lisbon, Portugal) and CITMA (Madeira, Portugal) for financial support to AB. LP was supported by a Ph.D. grant from Fundação para a Ciência e a Tecnologia (PRAXIS XXI/BD/13632/97). We thank Antonio Torroni for access to unpublished data.

REFERENCES

Alves-Silva, J., Santos, M., Guimarães, P., Ferreira, A. C., Bandelt, H.-J., Pena, S. D. & Prado, V. F. (2000). The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* **67**, 444–461.

Anderson, S., Bankier, A. T., Barrel, B. G., De Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Bandelt, H.-J., Alves-Silva, J., Guimarães, P., Santos, M., 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. J. (2001). Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann. Hum. Genet.* **65**, 549–563.

Barry, B. (1998). *Senegambia and the Atlantic slave trade*. CUP, Cambridge, England.

Budowle, B., Wilson, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., Holland, M. M. & Monson, K. L. (1999). Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci. Int.* **103**, 23–35.

Carreira, A. (1983). *Migrações nas Ilhas de Cabo Verde*. Instituto Caboverdeano do Livro, 2nd ed., Lisboa.

Chen, Y.-S., Olckers, A., Schurr, G., Kogelnik, A. M., Huoponen, K. & Wallace, D. (2000). mtDNA variation in the south African Kung and Khwe and their genetic relationships to other African populations. *Am. J. Hum. Genet.* **66**, 1362–1383.

Godinho, V. M. (1965). *Os descobrimentos e a economia mundial*. Arcádia ED., Vol. 1 and 2, Lisboa.

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.

Lareu, M. V., Phillips, C., Carracedo, A., Lincoln, P., Syndercombe, D. & Thompson, J. (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.

Lessa, A. & Ruffié, J. (1960). *Seroantropologia das Ilhas de Cabo Verde. Mesa redonda sobre o homem Cabo-Verdiano.* Junta de Investigações do Ultramar (ed.). Estudos, Ensaios e Documentos, Vol. 32. Lisbon.

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**, 232–249.

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**, 507–518.

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**, 439–458.

Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999). Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nature Genetics* **23**, 437–441.

Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H.-J. (1998). Mitochondrial DNA analysis of North-west African populations reveals genetic exchanges with European, Near-Eastern, and Sub-Saharan populations. *Ann. Hum. Genet.* **62**, 531–550.

Rando, J. C., Cabrera, V. M., Larruga, J. M., Hernández, M., González, A. M. Pinto, F. & Bandelt, H.-J. (1999). Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann. Hum. Genet.* **63**, 413–428.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Gölge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Calì, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Nørby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Torroni, A. & Bandelt, H.-J. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276.

Russell-Wood, A. J. (1998). *The Portuguese empire, 1415-1808. A World on the Move.* Johns Hopkins Univ. Press, Baltimore.

Saillard, J., Forster, P. Lynnerup, N., Bandelt, H.-J. & Nørby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* **67**, 718–726.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. & Wallace, D. C. (1996). Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.

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.

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