Ann. Hum. Genet. (2000), **64**, 491–506
Printed in Great Britain

Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation

L. PEREIRA",#, M. J. PRATA",# *and* A. AMORIM",#
"*Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias s}no, 4200 Porto, PORTUGAL*
#*Faculdade de Ciências da Universidade do Porto, Pr. Gomes Teixeira, 4050 Porto, PORTUGAL*

(Received 11.1.00. Accepted 26.7.00)

summary

The analysis of the hypervariable regions I and II of mitochondrial DNA in Portugal showed that this Iberian population presents a higher level of diversity than some neighbouring populations. The classification of the di erent sequences into haplogroups revealed the presence of all the most important European haplogroups, including those that expanded through Europe in the Palaeolithic, and those whose expansion has occurred during the Neolithic. Additionally a rather distinct African influence was detected in this Portuguese survey, as signalled by the distributions of haplogroups U6 and L, present at higher frequencies than those usually reported in Iberian populations. The geographical distributions of both haplogroups were quite di erent, with U6 being restricted to North Portugal whereas L was widespread all over the country. This seems to point to di erent population movements as the main contributors for the two haplogroup introductions. We hypothesise that the recent Black African slave trade could have been the mediator of most of the L sequence inputs, while the population movement associated with the Muslim rule of Iberia has predominantly introduced U6 lineages.

introduction

Since the description of the mitochondrial DNA sequence by Anderson *et al.* (1981), this peculiar genome, which is maternally inherited, non-recombining and fast-evolving, has been intensively investigated and applied to population studies. The initial screening based on restriction fragment length polymorphisms spread all over the molecule, was soon enlarged by direct sequencing of two hypervariable regions located in the control region (D-Loop): HVRI and HVRII.

Correspondence: Luísa Pereira, Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias s}n, 4200 Porto PORTUGAL. Tel: › 351 22 5570700; Fax: › 351 22 5570799.
E-mail: lpereira@ipatimup.pt

Present day sequence variation of mtDNA is a valuable tool for making what Avise *et al.* (1987) referred to as phylogeographic inferences. MtDNA sequences can be used to construct networks or used in other methodological approaches which a ord information about prehistoric population size and patterns of gene flow. The evolutionary history of haplogroups, their inferred origin and expansion through the world, provide the basis for reconstructing and dating major prehistoric and historic population movements.

Many published studies based on HVRI sequence diversity focus on the history of European populations (Richards *et al.* 1996; Côrte-Real *et al.* 1996; Richards *et al.* 1998). Richards *et al.* (1998), applying a phylogeographic approach to western Europe mtDNA diversity, concluded that the majority (85%) of

European sequences must have originated during the Upper Palaeolithic and su ered a considerable post-glacial expansion; about 15% of the sequences reflect a restricted Neolithic input, from the Near East toward the West of Europe, and only 1% of the sequences represent more recent influences of Asian and African mtDNA pools.

Focussing on the westernmost edge of Europe, the Iberian Peninsula, some studies (Côrte-Real *et al*. 1996; Salas *et al*. 1998) have pointed to a common origin of all Iberian populations in the Upper Palaeolithic. For most populations, diversity levels were found to be lower than the values reported for central European countries, a feature that was thought to support the expansion model of modern humans from the Middle East in the direction of Western Europe. The lowest Iberian diversity value was observed in Basques, reflecting the uniqueness of this Iberian population (Côrte-Real *et al*. 1996).

Another peculiarity of the Iberian mitochondrial pool is the presence of sequences belonging to the U6 group (Richards *et al*. 1998), signalling a North African influence that has not been detected elsewhere in other European populations.

In this work we have analysed HVRI and HVRII diversity in Portugal, the westernmost country of the Iberian Peninsula, with the aim of obtaining a better characterisation of European mtDNA variability. We have considered three main regions in Portugal: North, Central and South. This was done in parallel with a study of Y chromosome biallelic markers that has revealed statistical di erences between the south compared to the north and central regions (Pereira *et al*. 2000).

Our main approach regarding the analysis of mtDNA diversity was the evaluation of patterns of mismatch distribution within the major haplogroups found in Portugal. We intended to assess whether inferences regarding the history of

nos. AF277997–AF278237 and AF278238–AF278478).

Molecular diversity indexes and mismatch distributions were executed using the software ARLEQUIN 1.1 (Schneider *et al*. 1997).

results and discussion

HVRI and HVRII diversity in Portugal

Some diversity parameters obtained in the three Portuguese regions studied for HVRI and}or II are presented in Table 1. HVRI presented a higher mean number of nucleotide di erences than region II. However, when corrected for fragment sizes both regions showed a similar mean number of nucleotide pairwise di erences: HVRII}HVRI mean ratios were 0.87, 0.90 and 1.05 in North, Central and South Portugal, respectively.

Table 1. *MtDNA diversity parameters in North* (*NP*), *Central* (*CP*) *and South* (*SP*) *Portugal*,

Fig. 1. Observed (black) and expected (grey) mismatch distributions for HVRI, HVRII and HVRI › HVRII in Portugal.

et al. 1996), maintaining the level of diversity while population size was reduced in other Iberian regions, is a possible explanation, but in contradiction with the known palaeoclimatic

Table 3. *Sequence diversity observed in HVRI in several populations. N, sample size; K, number of di erent sequences found; A, number of variable nucleotides positions; B, mean nucleotide pairwise di erences; C, percentage average pairwise di erence per nucleotide;* p, *nucleotide diversity*.

	N	K	A	B	C	p	References
Basque	106	52	52	2-95	0-82	0-008	1, 2
Galician	92	53	56	3-13	0-87	0-009	3
Portuguese	54	38	46	3-60	1-00	0-010	2
Catalonian	15	11	16	3-73	1-04	0-010	2
British	100	71	67	4-45	1-24	0-012	4
S. Portuguese	59	41	54	4-51	1-25	0-013	This study
N. Portuguese	100	67	71	4-78	1-33	0-013	This study
C. Portuguese	82	62	66	4-87	1-35	0-014	This study
Spanish	89	70	69	5-02	1-39	0-014	2, 5
Tuscan	49	40	55	5-03	1-40	0-014	6
Turkish	96	79	82	5-45	1-51	0-015	7, 8
Middle-Eastern	42	38	59	7-08	1-97	0-020	9
S. Tomean	50	32	53	7-56	2-10	0-021	10

" Bertranpetit *et al.* (1995); # Côrte-Real *et al.* (1996); $ Salas *et al.* (1998); % Piercy *et al.* (1996); &

Fig. 2. Mismatch distributions for the haplogroups observed in Portugal (North, Central and South) considering only HVRI diversity. A- haplogroup U5 with and without U5a1a sequences. B- haplogroups H, V and K. C- haplogroup J. D- haplogroups T* and T1. E- haplogroup U6. F- haplogroups L1b › L2, L3* and all simultaneously.

Haplogroup U5 had the highest mode for the number of pairwise di erences distribution, and showed a regular unimodal pattern (Figure 2A). Both features are in accordance with its being the oldest haplogroup in Europe which has registered a regional development. As expected, and depicted in Figure 2A, a slight bimodality appeared when the more recent sub-haplogroup U5a1a was considered in the analysis.

The three haplogroups H, V and K, all considered to be post-glacially expanded European haplogroups (Figure 2B), showed clear unimodal distributions but with low means, reflecting the shorter period since the accumulation of variation began.

Haplogroups J and T, said to have had a common origin in the Near East, presented very distinct mismatch distribution patterns. Haplogroup J, which seems to have been recently introduced into Europe during the Neolithic,

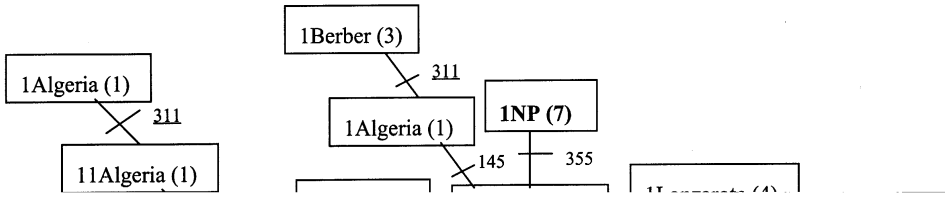

Fig. 3. A most parsimonious tree of sequences belonging to cluster U6. Root motif 172-219 indicated with an asterisk. Branches are labelled by the nucleotide positions in HRVI (minus 16000) to designate transitions; transversions are further specified and positions underlined represent parallel mutations. Numbers in brackets represent bibliographic reference (1) Côrte-Real *et al.* (1996), (2) Mateu *et al.* (1996), (3) Rando *et al.* (1998), (4) Rando *et al.* (1999), (5) Salas *et al.* (1998), Watson *et al.* (1996), (7) this work.

showed a very irregular curve with a ragged shape (Figure 2C). This pattern additionally suggests a high level of diversification of the founder sequences and also expresses the heterogeneity within the haplogroup: some well-defined J sub-haplogroups are distinctively spread through Europe, paralleling the clines observed for Y chromosomal, and some autosomal, markers.

In contrast, haplogroup T presented a unimodal and bell-shaped curve (Figure 2D), which can be explained by a more ancient introduction into Europe with subsequent accumulation, *in loco*, of homogenising mutations, starting from a

groups, U6 and L, that have been reported as occurring sporadically in other European populations, were detected with comparatively high frequency. Both haplogroups were characterised by high levels of diversity and displayed very irregular mismatch distributions (Figure 2 E and F). Moreover, haplogroup U6 was found to be restricted to the North region of the country, whereas the L sequences were spread all over the country.

These haplogroups have been reported to be characteristic of African populations, where their frequency is inversely correlated with the North-South axis: the frequency of U6 is high in North Africa and decreases in a southerly direction, being almost absent south of the equator; the L cluster has an opposite distribution (Rando *et al.* 1998, 1999; Watson *et al.* 1996; Mateu *et al.* 1996).

In Portugal, as well as generally in Iberia, many migration waves from both North and sub-Saharan African populations are well documented. The geographical proximity of North Africa and the Iberian Peninsula certainly a orded many opportunities for mutual population contacts. Among them, we stress the movement of Berbers and Arabs that took place during the very recent Muslim rule of Iberia (from the 8th century to the end of the 15th, in some regions). In addition, many sub-Saharan individuals entered the region during the slave trade period, from its very beginning (middle 15th century) until its total ban in the late 19th century.

As it would be interesting to find out the origin of the L and U6 sequences detected in Portugal, we have tried to compare the motifs of the sequences observed in Portugal with those described in the literature for several populations (Figures 3 and 4). However most of the matches found for the Portuguese sequences were with sequences widely distributed in Africa, and no clear pattern of geographic clustering was detected.

A striking aspect observed for the U6 haplogroup was that 5 out of 7 of the Portuguese

Fig. 4. A phylogeny of Portugese sequences belonging to African clades L1b and L2 and to the default cluster L3* (some members of which may have a non-African origin). The sequence with a transition from the CRS at np 16223 is indicated with an asterisk. Sequence matches in other populations are shown. Numbers in brackets represent bibliographic reference (1) Côrte-Real *et al.* (1996), (2) Graven *et al.* (1995), (3) Mateu *et al.* (1996), (4) Pinto *et al.* (1996), (5) Rando *et al.* (1998), (6) Rando *et al.* (1999), (7) Vigilant *et al.* (1991), (8) Watson *et al.* (1996), (9) this work. Solid circles represent sequences observed in the mtDNA database or branching nodes in the mtDNA phylogeny which aid in the identification of parallel mutations, which are shown with the position underlined.

sequences the absence of matches can be due to the present bias in the description of sub-Saharan mtDNA variability. Broad areas corresponding to Ivory Coast, Angola and Mozambique, which represented very important sources of African slaves, remain uncharacterised.

There were more African slaves in Portugal than in any other European country: in 1550, Lisbon boasted 10000 resident slaves in a population of 100000, and Portugal as a whole probably had over 40000 (Thomas, 1998). In the mid-sixteenth century the birth of slaves' children was stimulated in Portugal for internal tra c purposes. Inter-breeding between autochthonous individuals and African slaves certainly occurred and the predominant mating must have been between slave African females and autochthonous males, due to social pressures and also for legal reasons: o spring of slave females would be slaves, whereas o spring of slave males would not. Therefore, breeding between slave African males and white females, besides being socially repressed, would not bring any economic profit. If the pattern of genetic admixture was markedly sex influenced, the signature of this recent African influence would be expected to be very

di erent in the maternally inherited gene pool and in the paternally inherited one. In a recent study based on Y chromosome biallelic markers (Pereira *et al*. 2000) we have reported the absence of typical sub-Saharan haplogroups in the Y chromosome Portuguese pool. This finding, and the detection of L sequences at 7.1% in the mitochondrial pool, both seem to support the above-mentioned pattern of admixture with African slaves.

conclusions

Studies of large population samples, designed to characterise the molecular diversity in restricted geographical contexts, can produce valuable insights concerning specific demographic features that would remain undetectable in broader scale surveys. In this work we have studied mtDNA variability in Portugal, considering North, Central and South regions as micro-screening sample units.

The level of mtDNA diversity found, although characteristic of European populations, is high when the westernmost location of the country in Europe, and the reported European tendency for reduction of diversity toward Western Iberia (Corte-Real *et al*. 1996; Salas *et al*. 1998), are considered. The observed HVRI and HVRI › HVRII mismatch distributions were unimodal but smoother than others previously found in neighbouring populations.

This finding, as well as the high level of haplogroup diversity, suggests the influence of specific demographic factors acting in the Portuguese population, and led us to hypothesise that an important modulator of the present Portuguese mtDNA variability could have been the influx of distinct mtDNA lineages at historically quite di erent times.

Sharing the features of mtDNA diversity generally registered in Europeans (all European

reference sequence for human mitochondrial DNA. *Nature Genet.* **23**, 147.

Arnaiz-Villena, A., Martínez-Laso, J., Gómez-Casado, E., Díaz-Campos, N., Santos, P., Martinho, A., Breda-Coimbra, H. (1997). Relatedness among Basques, Portuguese, Spaniards and Algerians studied by HLA allelic frequencies and haplotypes. *Immunogenetics* **47**, 37–43.

Avise, J., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E. *et al.* (1987). Intraspecific phylogeography: the molecular bridge between population genetics and systematics. *Ann. Ver. Ecol. Syst.* **18**, 489–522.

Bertranpetit, J., Sala, I., 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.

Calafell, F., Underhill, P., Tolun, A., Aangelicheva, D. & Kaladjieva, L. (1996). From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann. Hum. Genet.* **60**, 35–49.

Comas, D., Calafell, F., Mateu, E., Pérez-Lezaun, A., Bosch, E., Martínez-Arias, R. *et al.* (1998). Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am. J. Hum. Genet.* **63**, 1824–1838.

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 population. *Mol. Biol. Evol.* **13**, 1067–1077.

Côrte-Real, H., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A. *et al*

Thomas, H. (1998). *The slave trade – the history of the Atlantic slave trade 1440–1870*. London: Macmillan Publishers Ltd.

Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., *et al*. (1998). MtDNA analysis reveals a major late Palaeolithic population expansion from southwestern to northeastern European populations. *Genetics* **62**, 1137–1152.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R. *et al*. (1996). Classification of

appendix

Haplotypes and their geographical distribution in North (*N*), *Central* (*C*) *and South* (*S*) *Portugal*.

HVRI	HVRII	N	C	S	Hap
051 162	73 263 311.1	—	1	—	H
051 257	152 263 303.1 311.1	1	—	—	H
075 183$^{A/C}$ 189 249 356	263 303.3 311.1	—	—	1	H
092 129 239	73 263 311.1	—	—	1	H?
093 126	199 263 303.1 311.1	1	1	—	H?
093 213 215 263$^{T/A}$	263 303.1 311.1	1	—	—	H
093 263	263 311.1	—	1	—	H
124	146 263 303.1 311.1	—	1	—	H
124 256	146 263 303.1 311.1	—	1	—	H
129	146 263 311.1	1	—	—	H
129	152 263 303.1 311.1	—	1	—	H
162 209	73 263 311.1	—	—	1	H
162 209 293	73 263 303.1 311.1	—	—	1	H
163	263 311.1	—	—	1	H
172	263 311.1	1	—	—	H
176	195 204 263 311.1				

HVRI	HVRII	N	C	S	Hap
CRS	152 263 303.1 311.1	1	1	1	H
CRS	151 152 263 311.1	1	—	—	H
CRS	152 263 311.1	1	2	—	H
CRS	263 269C/A 303.2 311.1	1	—	—	H
CRS	185 263 303.1 311.1	1	—	—	H
CRS	195 257 263 303.2 311.1	1	—	—	H
CRS	263 303.2 311.1	—	1	2	H
CRS	146 263 303.1 311.1	—	1	—	H
CRS	150 263 311.1	—	1	1	H
CRS	151 262 263 303.2 311.1	—	—	1	H
CRS	151 263 303.2 311.1	—	—	1	H
CRS	150 263 303.1 311.1	—	—	1	H
CRS	195 257 263 303.1 311.1	—	—	1	H
CRS	263 303.2 311.1 338	—	—	1	H
129 223 278 311 (391)	73 199 204 250 263 303.2 311.1	1	—	—	I
129 172 223 311	73 199 203 204 250 263 311.1	—	—	1	I
063 069 126	73 228 263 295 311.1	—	1	—	J*
069 126 172	73 228 263 295 311.1	1	—	—	J*

HVRI	HVRII	N	C	S	Hap
223	73 150 195 263 303.1 311.1	—	—	1	L3*
129 183A/C 189 223 249 311	73 195 263 303.1 311.1	—	1	—	M1
037 126 186 189 222	73 152 263 303.1 311.1	1	—	—	T*?

HVRI	HVRII	N	C	S	Hap
189 298	(72) 263 311.1	—	—	1	V
242	(72) 152 263 303.1 311.1	—	1	—	V?