International Congress Series 1261 (2004) 416–418

www.ics-elsevier.com

Subdividing mtDNA haplogroup H based on coding-region polymorphisms—a study in Iberia

L. Pereira[a,*], M. Richards[b], A. Alonso[c], C. Albarrán[c], O. Garcia[d], V. Macaulay[e], A. Amorim[a,f]

[a] *IPATIMUP (Instituto de Patologia e Imunologia Molecular da Universidade do Porto), Porto, Portugal*
[b] *Department of Chemical and Biological Sciences, University of Huddersfield, Huddersfield, UK*
[c] *Instituto de Toxicología, Sección de Biologia, Madrid, Spain*
[d] *Area de Laboratorio Ertzaintza, Gobierno Vasco, Bilbao, Spain*
[e] *Department of Statistics, University of Oxford, UK*
[f] *Faculdade de Ciências da Universidade do Porto, Portugal*

Abstract. Information on complete human mitochondrial DNA has revealed polymorphisms located in the coding region, which are useful for a more secure mtDNA haplogroup subdivision. We sequenced 1580 bp in the coding region informative for subclassification of haplogroup H, the most poorly resolved European haplogroup when studying only control-region variation. This revealed nine coding polymorphisms (at positions 3010, 3915, 3992, 4024, 4336, 4745, 4769, 4793 and 6776, besides the diagnostic position 7028) that allowed subclassification of around 80% of the 306 Iberian H samples in eight subhaplogroups. Frequencies for these haplogroups vary widely, with half of the samples sharing the substitution at position 3010 (but some structure is already apparent within this subhaplogroup). The HVRI diversity suggested the following age estimates for some subhaplogroups: 3010: 13,400 ± 3000 years; 6776: 8600 ± 2800 years; 4336: 4700 ± 2700 years; 3915: 17,300 ± 9100 years. Those ages indicate that these H subhaplogroups may be informative for different time scales of European demographic history.

Keywords: MtDNA; Coding region; Haplogroup H; Iberia

1. Introduction

The most common strategy used in the past to study the variation in population samples of mtDNA, the sequencing of one or both of the hypervariable regions located in the control region, has been insufficiently informative to reveal clear phylogeographic patterns in about a half of the European mtDNA pool. This is because the phylogeny of the most common European haplogroup, H, is poorly resolved. Recent data from mtDNA complete sequencing [1,2] have revealed informative polymorphisms at some coding-region positions, namely in the ND1, ND2 and CO1 genes and some tRNAs. We sequenced

* Corresponding author.
E-mail address: lpereira@ipatimup.pt (L. Pereira).

0531-5131/
doi:10.1016/S0531-5131(03)01651-0

four segments encompassing these diagnostic positions in three samples from the Iberian Peninsula, in the subset already classified as haplogroup H.

2. Material and methods

Iberian samples (194 from Portugal, 37 from Madrid in Central Spain and 75 from the Basque Country in Northeastern Spain) previously classified (by HVRI screening and checking of the diagnostic position 7028) as belonging to haplogroup H were sequenced for the following coding regions: nps 3001 – 3360, 3661 – 4050, 4281 – 4820 and 6761 – 7050. Primers used were, respectively: L2978 5′-GTC CAT ATC AAC AAT AGG GT-3′ and H3361 5′-CGT TCG GTA AGC ATT AGG AA-3′; L3640 5′-TCT AGC CAC CTC TAG CCT AG-3′ and H4051 5′-TAG AGT TCA GGG GAG AGT GC-3′; L4264 5′-CAT TCC CCC TCA AAC CTA AG-3′ and H4821 5′-AGA GGG GTG CCT TGG GTA AC-3′; L6740 5′-TGG TCT GAG CTA TGA TAT CA-3′ and H7051 5′-GAT GGC AAA TAC AGC TCC TA-3′. The temperature profiles for the PCR were: 95 °C for 10 s, 64 °C for 30 s and 72 °C for 30 s for 35 cycles for the third pair of primers, and the same except 58 °C as annealing temperature for the others. Automated sequencing was carried out in an ABI 3100 using the Kit Big-Dye Terminator Cycle Sequencing Ready Reaction (AB Applied Biosystems).

The HVRI diversity observed between 16,090–16,365 bp was used for age estimation, by assuming a mutation rate of one transition per 20,180 years and applying the rho statistics as described in [3]; calculations were carried out using the Network version 3.0 software.

3. Results

The haplogroup H frequencies in the Iberian samples analysed here are: 41% in Portugal; 49% in Madrid; and 50% in the Basque Country.

The sequencing of the 1580 bp in the coding region in the 306 H Iberian samples showed that nine polymorphisms were present in at least two of those populations, allowing the discrimination of eight H sub-haplogroups, the frequencies of which are displayed in Table 1.

There is still a wide range of subhaplogroup frequencies, with half of the individuals sharing the substitution at np 3010. But this subhaplogroup shows some diversity in the coding regions studied here, and potential further subdivisions within it can be defined. In Iberia, there is a lineage 3010 4733, the frequency of which was 0.015 in Portugal, 0.135 in Madrid and 0.107 in the Basque Country; and a subset of this with an additional substitution, 3010 4733 3849, with relative frequencies of 0.005, 0.027 and 0.067, respectively.

The second most frequent subhaplogroup in Portugal and the Basque country, defined by the substitution at position 6776, shows, curiously, a very low frequency in Madrid.

Table 1
Iberian frequencies of the H sub-haplogroups based on coding region polymorphisms (all transitions)

	Characteristic variants								
	3010	4769	6776	3992 4024	4336	3915	4745	4793	other/no
Portugal	0.459	0.041	0.191	0.021	0.031	0.031	0.010	0.026	0.190
Madrid	0.541	0.054	0.027	0.027	0.081	0.027	–	0.054	0.189
Basque Country	0.507	0.120	0.160	0.027	0.053	–	0.013	–	0.080

The Basque country presents a high frequency for the subhaplogroup defined by the 4769 polymorphism, and half of the samples presented an additional coding polymorphism at position 4592, that was not observed in the other two populations.

A proportion of 10 – 20% of the H samples were not further classifiable on the basis of variation in these coding-region segments.

This study showed that HVRI sites that have been previously used to define subclades (e.g., position 16,304) seem to have mutated more than once in haplogroup H. Nevertheless, the HVRI position 16,304 was present in all but one of the 13 individuals that presented the coding substitution 4336, while 16362 was present in all of the seven individuals bearing the 3915 substitution.

Using the HVRI diversity within the most frequent of these subhaplogroups, we estimated the following times for their most recent common ancestors: 3010: 13,400 ± 3,000 years; 6,776: 8600 ± 2800 years; 4336: 4700 ± 2700 years; 3915: 17,300 ± 9100 years.

4. Conclusions

This initial screening, by sequencing a considerable region (1580 bp long) around the coding positions potentially informative for haplogroup H subclassification, indicated a total of nine coding positions that enable the classification of 80% of the Iberian H samples in eight subhaplogroups. A more practical screening strategy (RFLP or minisequencing) can be designed for population genetic/forensic surveys of these polymorphisms, whenever HVRI diversity is poorly informative. In fact, the HVRI CRS haplotype appeared in different coding-defined haplotypes, probably due to the high recurrence in the control region. It is a priority to learn more about them in a wider geographical area, since the different ages seem to indicate that these H subhaplogroups may be informative for different time scales of European prehistory.

Acknowledgements

This work was partially supported by a research grant to LP (SFRH/BPD/7121/2001) from Fundação para a Ciência e a Tecnologia and IPATIMUP by Programa Operacional Ciência, Tecnologia e Inovação (POCTI), Quadro Comunitário de Apoio III. This study is included in the Project POCTI/ANT/45139/2002 financed by Fundação para a Ciência e a Tecnologia (Eixo 2, Medida 2.3 do POCTI, QCA III). VM has a Wellcome Trust Research Career Development Fellowship.

References

[1] M. Ingman, H. Kaessmann, S. Paabo, U. Gyllensten, Mitochondrial genome variation and the origin of modern humans, Nature 408 (2000) 708–713.

[2] C. Herrnstadt, J.L. Elson, E. Fahy, G. Preston, D.M. Turnbull, C. Anderson, S.S. Ghosh, J.M. Olefsky, M.F. Beal, R.E. Davis, N. Howell, Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups, Am. J. Hum. Genet. 70 (2002) 1152–1171.

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