Available online at www.sciencedirect.com
SCIENCE DIRECT
Forensic Science International 159 (2006) 43–50
Forensic Science International
www.elsevier.com/locate/forsciint

Evaluating the forensic informativeness of mtDNA haplogroup H sub-typing on a Eurasian scale

Luísa Pereira [a,*], Martin Richards [b], Ana Goios [a,c], Antonio Alonso [d], Cristina Albarrán [d], Oscar Garcia [e], Doron M. Behar [f], Mukaddes Gölge [g], Jiři Hatina [h], Lihadh Al-Gazali [i], Daniel G. Bradley [j], Vincent Macaulay [k], António Amorim [a,c]

[a] *Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal*
[b] *Schools of Biology and Computing, University of Leeds, Leeds, UK*
[c] *Faculdade de Ciências da Universidade do Porto, Porto, Portugal*
[d] *Instituto de Toxicología, Sección de Biologia, Madrid, Spain*
[e] *Area de Laboratorio Ertzaintza, Gobierno Vasco, Bilbao, Spain*
[f] *Bruce Rappaport Faculty of Medicine and Research Institute, Technion and Rambam Medical Center, Haifa, Israel*
[g] *Department of Physiology, University of Kiel, Kiel, Germany*
[h] *Charles University, Medical Faculty in Pilsen, Institute of Biology, Pilsen, Czech Republic*
[i] *Department of Paediatrics, Faculty of Medicine and Health Sciences, UAE University, Al Ain, United Arab Emirates*
[j] *Smurfit Institute of Genetics, Trinity College, Dublin, Ireland*
[k] *Department of Statistics, University of Glasgow, Glasgow, UK*

Received 30 March 2005; received in revised form 17 June 2005; accepted 18 June 2005
Available online 1 August 2005

Abstract

The impact of phylogeographic information on mtDNA forensics has been limited to the quality control of published sequences and databases. In this work we use the information already available on Eurasian mtDNA phylogeography to guide the choice of coding-region SNPs for haplogroup H. This sub-typing is particularly important in forensics since, even when sequencing both HVRI and HVRII, the discriminating power is low in some Eurasian populations. We show that a small set (eight) of coding-region SNPs resolves a substantial proportion of the identical haplotypes, as defined by control-region variation alone. Moreover, this SNP set, while substantially increasing the discriminating efficiency in most Eurasian populations by roughly equal amounts, discloses population-specific profiles.

Keywords: Human mtDNA; Haplogroup H; Haplotype diversity; HVRI and HVRII; Coding-region SNPs; Eurasia

1. Introduction

The advent of new methodological strategies for SNP typing, allied to the publication of complete mtDNA sequence population data, has given birth to a new phase in the application of mtDNA to forensic casework, characterized by the incorporation of coding-region information.

* Corresponding author. Tel.: +351 22 5570700; fax: +351 22 5570799.
E-mail address: lpereira@ipatimup.pt (L. Pereira).

0379-0738/$ – see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2005.06.008

Searches for the most informative coding-region polymorphisms are currently in progress, either by randomly screening some mtDNA genes in specific population samples [1], or by performing complete mtDNA sequencing on some of the more common HVRI/HVRII haplotypes [2,3]. This latter strategy has led to the identification of several coding-region polymorphisms, requiring a set of eight multiplex panels. Up to now, only the first one of these panels has been adapted to the SNaPshot methodology [4]. Other multiplex assays for haplogroup classification have been developed (also based on SNaPshot), suitable not only for forensic but also for anthropological purposes: Brandstatter et al. [5] designed an assay of 16 SNPs for the characterization of western Eurasian haplogroups; Quintáns et al. [6] combined 17 SNPs in two multiplexes, the first including SNPs that define common European haplogroups, and the second for polymorphisms defining sub-haplogroups within haplogroup H.

It is widely agreed that this strategy, at least in forensics, will not replace HVRI/HVRII sequencing. It can be used, nonetheless, to increase the discrimination power in cases of a HVRI/HVRII haplotype match.

It remains to be determined to what extent this gain of information due to coding-region typing holds throughout different west Eurasian populations. The best candidate for this evaluation is haplogroup H due to (1) its poor phylogenetic resolution in HVRI, and (2) its being by far the most common west Eurasian haplogroup, accounting for 40–50% of the mtDNA pool in most of Europe, and ~20–30% in the Near East and the Caucasus region [7–9].

Recently, using polymorphism information derived from the growing complete mtDNA sequence database, we sequenced 1580-bp of targeted coding-region segments of the mtDNA genome in 649 individuals harbouring mtDNA haplogroup H from populations throughout Europe, the Caucasus and the Near East [10]. This screening included, besides the haplogroup H diagnostic polymorphism 7028C, all of the seven SNPs (3010A, 4769A, 6776C, 3992T, 4336C, 3915A and 4793G, defining, respectively, sub-haplogroups H1–H7) combined in multiplex 2 by Quintáns et al. [6], and the polymorphism 4745G that defines the additional sub-haplogroup H13 [8]. This enhanced genealogical resolution clearly showed that sub-clades of haplogroup H have highly distinctive geographical distributions throughout Eurasia.

This improvement in phylogeographic resolution could provide information relevant to forensics, namely to guide the choice of SNPs to be typed. To test this hypothesis, we have evaluated the potential of sub-typing these eight coding-region polymorphisms to increase the haplotype discrimination power inside haplogroup H in populations throughout west Eurasia. Framed in a more practical forensic context, we have also analysed, for the same purpose, a subset of these population samples for which combined HVRI/HVRII information, along with the above-mentioned SNPs, was available. This is an attempt to extend the use of phylogeographic approaches to mtDNA forensics, which have been largely limited up to now to the quality control of published sequences and databases [11].

2. Materials and methods

2.1. Samples and sequencing

We dissected haplogroup H variation in 649 samples from 20 populations from Europe, the Caucasus and the Near East (see Supplementary material), previously analysed only for HVRI sequence variation and some haplogroup-diagnostic RFLPs. The populations were grouped as: SW (Portugal [12], Spain-Madrid [10] and Basque Country [10]), NW (France [7], Ireland [7] and Norway [7]), Med (Italy [7], Sardinia [7] and Crete [7]), NE (Finland [13], Russia [7] and Chuvash [7]), CSE (Poland [7], Czech Republic [7], Romania [7] and Bulgaria [7]), AJ (Ashkenazi Jews [14]), CA (Caucasus [7]) and NRE (Palestine [7], Kurd [7] and Gulf states [7]). We sequenced four mtDNA coding-region segments encompassing the principal diagnostic positions in haplogroup H samples: 3001–3360, 3661–4050, 4281–4820, and 6761–7050 (a total of 1580 bp) according to the numbering of the reference sequence (CRS) [15]. The new sequences generated for this work have been deposited in GenBank, accession nos. AY776364–AY778959. We also included 31 complete sequences from Finland [13], available in GenBank, accession nos. AY339402–AY339432. In Supplementary material, we give tables of the diversity in these coding regions, together with diversity in HVRI and, in some cases, also in HVRII.

For amplifying and sequencing the four coding-region segments, the following four sets of primers were used: 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 with 58 °C as annealing temperature, for the others. We carried out automated sequencing in an ABI 3100, using the Big-Dye Terminator Cycle Sequencing Ready Reaction kit (AB Applied Biosystems).

2.2. Genetic analysis

Control-region positions considered for analysis were between nucleotide positions 16024–16365 for HVRI, and 73–340 for HVRII. The hypervariable positions 16182 and 16183 in HVRI and the indels at positions 309 and 315 in HVRII were not considered in the analysis [16]. Assignment