Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 2 (2008) 1–8
FSI GENETICS
www.elsevier.com/locate/fsig

Evaluation of mitochondrial DNA coding region assays for increased discrimination in forensic analysis

Martina Nilsson, Hanna Andréasson-Jansson, Max Ingman, Marie Allen *

Uppsala University, Department of Genetics and Pathology, Rudbeck Laboratory, 751 85 Uppsala, Sweden

Received 15 May 2007; received in revised form 9 July 2007; accepted 18 July 2007

Abstract

There is an increasing trend to use mitochondrial DNA (mtDNA) analysis in criminal investigations where only limited amounts of DNA are available. However, analysis of the mtDNA control region has the drawback of low discrimination power, due to the lack of recombination that results from uniparental (maternal) inheritance. As a strategy to increase discrimination, a number of typing assays detecting variation in the mitochondrial coding region have been developed. In this study, several of these assays are evaluated for their discriminatory capacity using data obtained from 495 complete Caucasian mtDNA sequences. In order to add a local geographic perspective to this evaluation, we have also sequenced and analysed the entire mtDNA from 20 individuals of Swedish origin. We find that the coding region assays are very useful for resolving sequences with identical HVI/HVII regions. The best-performing coding region assay was able to discriminate 46% of the resolvable sequences, compared to 20–30% for the other coding region assays we evaluated.

Keywords: mtDNA; Coding region; Haplogroup; Discrimination; Forensic

1. Introduction

Forensic evidence materials are often subjected to harsh environments that degrade the DNA. Moreover, some samples (e.g. shed hairs, teeth or bone) often contain low amounts of DNA. When nuclear DNA cannot be obtained, sequencing of mitochondrial DNA (mtDNA) is a commonly used alternative for these samples, due to its high DNA copy number per cell [1]. In addition, the mitochondrial genome has a higher nucleotide substitution rate than the nuclear genome and the two hypervariable regions HVI and HVII, located in the non-coding control region, are dense in polymorphic positions [2,3]. The extensive sequence variability found within the control region makes this region suitable for routine analysis in human identification. However, due to maternal inheritance and the resulting lack of recombination of this genome, there is a relatively high risk of mtDNA sequence identity by chance. Although a majority (84%) of the HVI/HVII mtDNA sequences are found to be unique in a large sample set, several individuals share identical sequences [4]. Mitochondrial DNA is characterised by clusters of closely related lineages, or haplogroups, that are distinguished by certain diagnostic polymorphisms. The diagnostic SNPs that define these haplogroups are commonly located in the coding region, and some are more common in certain populations. Among Caucasian populations, haplogroup H is the most common and is present at frequencies between 40 and 50% [5,6]. The most common HVI/HVII type (H1, an H sub-haplogroup) is present in approximately 7% of individuals of Caucasian origin [7]. Other HVI/HVII types also occur at high frequencies, rendering some individuals indistinguishable by HVI/HVII analysis.

Besides the polymorphism-dense non-coding region, a substantial number of polymorphic positions are found in the 15,447 bp coding region, but with a lower average variability [3]. Therefore, in addition to the HVI/HVII regions, analysis of positions in the coding region has been suggested as a valuable tool to increase the discrimination power in forensic mtDNA analysis [1,4]. In recent years, several mtDNA typing assays, based on analysis of informative SNPs or selected fragments of DNA in the coding region, have been developed. The first coding region assay described is based on pyrosequencing technology and allows the analysis of all the

* Corresponding author. Fax: +46 18 471 4808.
E-mail address: marie.allen@genpat.uu.se (M. Allen).

1872-4973/$ – see front matter
doi:10.1016/j.fsigen.2007.07.004

variation present in 15 short fragments [8,9]. One of the first assays utilising coding region SNPs, rather than sequencing of short regions, was developed to identify and resolve nine Caucasian haplogroups [5]. This assay is based on single nucleotide extension in a SNaPshot analysis and covers 16 SNPs in the coding region. The SNaPshot system has the advantage of having a relatively high multiplexing capability. In a second SNaPshot based assay, 11 SNPs is used to resolve the most common Caucasian HVI/HVII mtDNA type (H1) [10]. For this assay, informative SNPs for H1 resolution were chosen from 59 coding region positions revealed by sequence analysis of 241 European Caucasian individuals [7,10]. Another SNaPshot assay for increased mtDNA discrimination and assigning sequences into haplogroups is based on analysis of 17 coding region SNPs [6]. More recently, three additional SNaPshot assays for sub-typing haplogroup H sequences have been described (in total 45, 16 and 9 coding region positions, respectively) [11–13].

A growing number of completely sequenced mtDNA genomes are emerging, adding valuable population data and information for future forensic applications. The human mitochondrial genome database (mtDB) consists of 1865 complete human mtDNA sequences and 839 coding region sequences [14]. In the mtDB dataset with sequences from global populations, 505 of the total 2704 coding region sequences belong to haplogroup H (19%), 253 to U (9%), 120 to T (4%) and 130 to haplogroup K (5%). Samples belonging to these haplogroups often show low control region variability and were therefore investigated in further detail in this study.

Here, an extensive investigation of the variation in the mitochondrial coding region among some haplogroups commonly seen among Caucasians (H, U, T and K) has been performed. Furthermore, the discrimination capacities of several mtDNA coding region typing assays were evaluated by pairwise comparisons of 495 sequences of Caucasian origin. To further investigate the discrimination on regional data, we also evaluated these assays in a dataset of 20 complete mitochondrial genomes from individuals of Swedish origin.

2. Materials and methods

2.1. Database sequences and pairwise comparisons

When this study was initiated (February 2006), 1624 complete human mtDNA sequences and 837 coding region sequences were available in the mtDB dataset [14]. Of these, a subset consisting of sequences of Caucasian origin (495 samples) was compiled, comprising 241 European Americans [7], 192 Finnish [15] and 62 Italian individuals [16]. The European American sequences as well as most of the Italian sequences had been selected because they are difficult to discriminate from each other, whereas the Finnish dataset represents the normal mtDNA type distribution in Finland [7,15–17]. All sequences were imported into the Sequencher™ 4.1.2 software (Gene Codes Corporation, Ann Arbor, MI, USA). Due to commonly observed length heteroplasmy in the homopolymeric C-stretch in the HVII region (nucleotide 303–315), this region was deleted [18]. Furthermore, positions in any sequence with interpretation difficulties were labelled 'N' and regarded in further analyses as possibly being any of the four nucleotides.

The sequences were divided into three separate datasets; the HVI/HVII region (HVII, nucleotide position (np) 73–340 and HVI, 16024–16365), non-hypervariable region (NonHV, np 1–72, 341–16023 and 16366–16569), and complete genome (np 1–16569). Furthermore, the sequences were sub-divided into four common haplogroups. In total, 189 of the sequences belong to haplogroup H, 50 to haplogroup T, 38 to haplogroup K, 31 to haplogroup U while 187 of the sequences did not match any of these four haplogroups (Table 1). Pairwise comparisons of the sequences in each of the three datasets (HVI/HVII, NonHV and complete genome) were performed to evaluate the extent of genetic diversity in the different regions of mtDNA and among sequences of different haplogroups.

To investigate the discriminatory capacity among eight different mtDNA coding region typing assays, we compiled further partial genome datasets from the 495 sequences. The sequences were trimmed to create datasets containing the HVI/HVII region (16024–16365, 73–340) and also the specific positions covered by each of the eight different mtDNA coding region typing assays, ranging from 9 to 59 SNPs or 15 short fragments (Table 2). Of these assays, seven are based on analysis of SNPs and have been described in detail by Pereira et al. [13], Vallone et al. [10], Brandstätter et al. [5], Grignani et al. [12], Quintans et al. [6], Brandstätter et al. [11], and Coble et al. [7]. An additional assay described by Andreasson et al. [9] for sequence analysis of short stretches of DNA rather than identification of individual SNPs, was also included. In Table 2, only the positions covered by the seven SNP assays (in total 113 positions) are shown. The discriminatory capacities of the different typing systems were evaluated by pairwise analysis of the trimmed sequences. As a consequence of the interpretational guidelines for forensic HVI/HVII analysis, at least two pairwise differences (substitutions or indels) between sequences are required to classify them as unique (resolved). These guidelines have been developed to handle situations when the evidence and reference material differ by a single nucleotide in the control region. Due to heteroplasmy, a

Table 1
Haplogroup distribution of the 495 sequences

Reference	Total number of individuals	H[a]	T[b]	K[c]	U[d]	Other
Coble et al. [7]	241	104	39	28	0	70
Moilanen et al. [15]	192	31	11	10	31	109
Achilli et al. [16]	62	54	0	0	0	8
	495	189	50	38	31	187

[a] Haplogroup H defined by nucleotide 7028C [5].
[b] Haplogroup T defined by nucleotide 7028T, 14766T, 1719G, 12372G, 13708G and 8697A [5].
[c] Haplogroup K defined by nucleotide 7028T, 14766T, 1719G, 12372A, and 14798C [5].
[d] Haplogroup U defined by nucleotide 7028T, 14766T, 1719G, 12372A and 14798T [5].