Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 1 (2007) 13–19
FSI GENETICS
www.elsevier.com/locate/fsig

Extended guidelines for mtDNA typing of population data in forensic science

Walther Parson [a,*], Hans-Jürgen Bandelt [b]

[a] Institute of Legal Medicine, Innsbruck Medical University, Muellerstrasse 44, 6020 Innsbruck, Austria
[b] Department of Mathematics, University of Hamburg, Bundesstr. 55, 20146 Hamburg, Germany

Received 16 May 2006; received in revised form 16 November 2006; accepted 19 November 2006

Abstract

Mitochondrial DNA analysis has become a vital niche in forensic science as it constitutes a powerful technique for low quality and low quantity DNA samples. For the forensic field it is important to employ standardized procedures based on scientific grounds, in order to have mtDNA evidence be accepted in court. Here, we modify and extend recommendations that were spelled out previously in the absence of solid knowledge about the worldwide phylogeny. Refinement of those earlier guidelines became necessary in regard to sample selection, amplification and sequencing strategies, as well as a posteriori quality control of mtDNA profiles. The notation of sequence data should thus reflect this growing knowledge.

Keywords: Mitochondrial DNA; Haplogroup; Sequencing; Statistical analysis; Quality control; Network; EMPOP

1. Introduction

MtDNA analysis has found widespread application in forensic medicine in the past 15 years, as it often constitutes the last chance for successful DNA typing when only very limited or severely degraded DNA is present in a sample. In other cases, mtDNA screening is applied as its mode of inheritance allows testing for a putative exclusion scenario in human identification. While recently developed mtDNA screening methods provide a reliable and cost-effective supplementary technique for routine applications, the golden standard still constitutes direct sequencing of PCR products or cloned amplicons. Laboratory guidelines and general considerations were established in order to standardize the mtDNA typing process [1–3]. The increasing body of population data that became available in the past few years improved our understanding of the worldwide mitochondrial phylogeny—but the wealth of data also revealed that some laboratory practice still met difficulties in dealing with mtDNA sequences, both at the sequencing and documentation stages [4–10]. The earlier forensic recommendations did not detail amplification and sequencing schemes, which are, however, crucial for the generation of high-quality sequencing data and a meaningful interpretation of the results. Moreover, a reappraisal of the interpretation rules for length variants in polycytosine tracts, which were previously defined by Refs. [11,12] in a first attempt, became necessary as well.

We briefly discuss the essential steps to take from sampling to sequencing and finally documenting the results. We specifically address laboratory procedures that are crucial for the generation of high quality data. A number of errors can be avoided by applying appropriate quality control procedures a posteriori to the sequencing process. Good laboratory practice demonstrates that the additional analysis of coding-region sites helps to confirm control-region haplotypes and increases our understanding of haplogroup-specific variation [13]. Finally, we draw attention to a modified interpretation of length variants that will be featured in a separate paper [14].

2. Sampling scheme and sample information

Sampling for population databases in the forensic field does not seem to follow any standardized scheme. Most samples would constitute convenience samples of ‘‘laboratory staff members, or from blood donor samples with the cooperation of a local blood bank, or from samples from victims and suspects

* Corresponding author. Tel.: +43 512 9003 70640; fax: +43 512 9003 73640.
E-mail address: walther.parson@i-med.ac.at (W. Parson).

1872-4973/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2006.11.003

examined in the course of casework'' [15, p. 44]. Convenience samples are not representative for the targeted national populations or ethnic groups and cannot be regarded as ''random'', inasmuch as they are typically pre-structured in various ways that may or may not influence the mtDNA profiles. Instead, one should strive for avoiding the adverse influence of hidden stratification (social stratum, ethnic origin, etc.), although this seems hard to achieve in practice for mixed (e.g. urban) populations.

The sample information should be as detailed as possible concerning matrilineal ancestry in order to achieve maximum flexibility for future choices of reference databases tailored to a specific forensic case. The broad regional geographic information about the birth place of the maternal grandmother should be provided whenever retrievable. Regional databases should then not include mtDNAs of recent immigrants without identifying the matrilineal origin (within the last two generations), since immigrant groups will be insufficiently represented even in large samples. Reference to specific ethnic or social groupings may be relevant insofar as such strata may partially reflect immigration events of the distant past. Reference to hair or skin colour (or whatever concept of so-called ''race'') of the mtDNA donor, however, is not really the information one should aim at: ''Geographical origin (ancestry) appears to be more relevant than a person's self-identified race'' [16]. In particular, the vague classifier ''Caucasian'' or ''Caucasoid'' should no longer be used to label samples, say, from specific European or recently admixed populations, since such outdated concepts are derived from ill-based preconceptions about human genetic variation; see also Section 6 (Names and Labels) in the American Heritage® Book of English Usage (www.bartleby.com/64/6.html).

In EMPOP, the EDNAP MtDNA Population Database (www.empop.org), metapopulations are distinguished at the most basal level of classification, viz. the sub-Saharan African, West Eurasian (including North African), South Asian, East Asian, Southeast Asian, Oceanian, Native American metapopulations, and the Central Asian population (which originated from mainly West and East Asian groups) as well as other populations (e.g. of the Americas) that were the result of colonization, slave trade, and more recent large-scale immigration. The next level is the national/ethnic group. Future analyses of population stratification will then permit fine-grained sampling schemes that can specifically focus on geographic and social strata within a national state or ethnic group.

3. Sequencing of separate segments in the control region

The mtDNA control region spans over 1122 base pairs from nucleotide position 16,024 (the origin of heavy-strand replication) to position 576. Traditionally, only the first two hypervariable segments (HVS-I and HVS-II, alias HV1 and HV2) of the control region have been targeted for forensic purposes, because a good portion of the differences between two mtDNAs can be found in these two short segments, with maximum ranges of roughly 16024–16,400 and 40–400. In some cases, only HVS-I typing (which is still popular in population genetics) has been exercised, which, however, can no longer be considered to be state of the art. The segment ranges that are sequenced may vary from study to study (depending on the employed primers), but too narrow ranges may miss a number of frequent nucleotide variants that are characteristic for several haplogroups, such as transitional changes at positions 16,051 (e.g. in haplogroup U2), 16,390 (in haplogroups L2, E, N1b1, U3a, etc.), 16,391 (in haplogroup I), 16,399 (in haplogroups L2d, F4a, and U5a1), 64 (in haplogroups A2, L0a1, etc.), 72 (in haplogroup HV0 = pre-V), and 357 (in haplogroup L1b).

In the late 1990s it became evident that there is yet another short stretch in the control region that also harbours notable variation, which has then been called the third hypervariable segment (HVS-III), with approximate range from about 430 until 576, the 3′-end of the control region [17–19]. There were only very few studies that targeted HVS-III besides the other two hypervariable segments [20–22]. HVS-III includes a few sites that make part of the motifs of some Eurasian haplogroups [23–26] and thus increase the discriminative power of forensic analyses, such as the 447G transversion in haplogroup M2, and the transitions at positions 431 (D4b1b), 456 (H5), 461 (M6), 462 (J1), 482 (M3, J1c1), 489 (M, J), 497 (K1a), 499 (B4b, U4′9), and 508 (U2e).

The separate analysis of the two or three hypervariable segments of an mtDNA sample inherently bears the risk of mix-up of these regions between different mtDNAs, leading to artificial recombinants [4,8,27,28]. Typically, the laboratory process of mtDNA analysis involves DNA extraction, amplification, purification of the amplification products, double-strand sequencing, purification of the sequencing products, and preparation of the samples for electrophoresis (e.g. [6,29]). Mix-up can be generated at the bench when handling the tubes or in the course of generating sample sheets for the individual laboratory steps. Each of these errors would result in recombinant mitochondrial haplotypes, consisting of segments that may belong to different haplogroups. Such errors cannot be unmasked by examining the raw data but only through a posteriori investigation using phylogenetic analysis [4,27–29]. A repeated full analysis of each sample (such as recommended for casework analysis [1,30]) would, of course, help minimizing this kind of error, but for the high through-put analysis of population samples this strategy might seem too laborious and too expensive.

4. Sequencing of the entire control region

Instead of performing multiple separate amplifications, it has proven most efficient to amplify the entire control region in a single PCR assay and to sequence with internal primers, because this scheme minimizes the risk of introducing sample mix-up at the stage of the PCR products. Due to the high nucleotide variability within the control region, internal primer annealing sites are prone to mutations which would lower or even inhibit PCR yield. Further, the polycytosine tracts incurred