Forensic Science International: Genetics 13 (2014) 134–142

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing

W. Parson [a,b,*], L. Gusmão [c,d], D.R. Hares [e], J.A. Irwin [e], W.R. Mayr [f], N. Morling [g], E. Pokorak [e], M. Prinz [h], A. Salas [i], P.M. Schneider [j], T.J. Parsons [k]

[a] Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[b] Penn State Eberly College of Science, University Park, PA, USA
[c] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil
[d] IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, Portugal
[e] FBI Laboratory, Quantico, VA, USA
[f] Division of Blood Group Serology, Medical University of Vienna, Austria
[g] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[h] Department of Sciences, John Jay College for Criminal Justice, New York, NY, USA
[i] Unidade de Xenética, Departamento de AnatomíaPatolóxica e CienciasForenses, and Instituto de CienciasForenses, Grupo de MedicinaXenómica (GMX), Facultade de Medicina, Universidade de Santiago de Compostela, 15872 Galicia, Spain
[j] Institute of Legal Medicine, Medical Faculty, University of Cologne, Cologne, Germany
[k] International Commission on Missing Persons, Alipasina 45a, 71000 Sarajevo, Bosnia and Herzegovina

ARTICLE INFO

Article history:
Received 15 July 2014
Accepted 19 July 2014

Keywords:
Phylogeny
Haplogroups
Quasi-median networks
Heteroplasmy
Alignment
Database searches

ABSTRACT

The DNA Commission of the International Society of Forensic Genetics (ISFG) regularly publishes guidelines and recommendations concerning the application of DNA polymorphisms to the question of human identification. Previous recommendations published in 2000 addressed the analysis and interpretation of mitochondrial DNA (mtDNA) in forensic casework. While the foundations set forth in the earlier recommendations still apply, new approaches to the quality control, alignment and nomenclature of mitochondrial sequences, as well as the establishment of mtDNA reference population databases, have been developed. Here, we describe these developments and discuss their application to both mtDNA casework and mtDNA reference population databasing applications. While the generation of mtDNA for forensic casework has always been guided by specific standards, it is now well-established that data of the same quality are required for the mtDNA reference population data used to assess the statistical weight of the evidence. As a result, we introduce guidelines regarding sequence generation, as well as quality control measures based on the known worldwide mtDNA phylogeny, that can be applied to ensure the highest quality population data possible. For both casework and reference population databasing applications, the alignment and nomenclature of haplotypes is revised here and the phylogenetic alignment proffered as acceptable standard. In addition, the interpretation of heteroplasmy in the forensic context is updated, and the utility of alignment-free database searches for unbiased probability estimates is highlighted. Finally, we discuss statistical issues and define minimal standards for mtDNA database searches.

1. Introduction

Mitochondrial DNA (mtDNA) is present in much higher copy number in the human cell than nuclear DNA. For this reason, mtDNA analysis can provide useful results in (forensic) samples that fail to yield successful nuclear DNA profiles. MtDNA haplotypes are uniparentally inherited and therefore reflect information from a non-recombining maternal lineage that may be shared by numerous people. As a result, mtDNA data cannot be applied to the identification of individuals in the same way that analyses from recombining nuclear DNA markers can. The mtDNA control region (CR), the largest non-coding portion of the molecule, nevertheless represents one of the most discriminatory single genetic markers known to forensics. Previous recommendations on the use of mtDNA sequence data in forensics addressed the need

* Corresponding author at: Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria. Tel.: +43 512 9003 70640; fax: +43 512 9003 73640.
E-mail address: walther.parson@i-med.ac.at (W. Parson).

http://dx.doi.org/10.1016/j.fsigen.2014.07.010
1872-4973/

for appropriate laboratory practice, the use of negative and positive controls, basic nomenclature guidelines for sequence differences and heteroplasmy, as well as guidance on interpretation, reporting and statistics [1–3]. In this article, we take a closer look at the challenges surrounding mtDNA casework and population databasing that have arisen since the establishment of these original guidelines, and we update the guidelines where appropriate.

Many of the challenges were encountered during the establishment of the EDNAP Mitochondrial DNA Population Database (EMPOP; www.empop.org; [4]) over the past 15 years. In its early stages, EMPOP was envisioned and designed to serve as a reference population database for use in the evaluation of mtDNA evidence worldwide, with its primary goal of providing the highest quality mtDNA data. The architecture of the EMPOP search engine and the various analysis tools provided *via* the website have evolved over the years. However, the emphasis of the EMPOP database continues to be on the importance of mtDNA data quality. As a result, EMPOP not only serves as a reference population database, but also as quality-control tool for scientists in forensic genetics and other disciplines. Though a number of high-quality reference population databases exist for forensic comparisons, EMPOP is the most comprehensive resource from the standpoint of populations represented. We therefore recommend EMPOP for its data quality control tools and its resource of global population data.

While the topics addressed in the following sections are generally applicable to both population genetic studies and forensic evidentiary sample handling, particular matters and recommendations are more relevant to one application than the other. In these cases, we treat forensic casework and population databasing separately, and describe the differences.

2. Generation of mtDNA data, good laboratory practice

2.1. General recommendations

MtDNA testing is extremely sensitive and thus contamination is of greater concern than it is with other forensic PCR-based methods. Pre-amplification measures that should be used to mitigate contamination were put forth in earlier guidelines [2,3,5] and still hold true. The following measures apply to both casework and databasing scenarios unless otherwise noted:

- Appropriate laboratory conditions that include dedicated spaces, instruments, chemistry and lab wear should be established for mtDNA testing.
- The use of controls (negative amplification, reagent blank and positive controls) should be carried all through the laboratory process.
- Low levels of contamination may be tolerated, since reliable results can be obtained in the presence of contamination.
- Redundant sequence information should be obtained for each reported mtDNA position to include both forward and reverse sequencing primers when possible.
- Sample consensus sequences should be determined using dedicated software for raw data alignment to the revised Cambridge Reference Sequence (rCRS; [6]).
- Consensus sequences should be confirmed by a second independent analysis of the raw data.
- Regular participation in proficiency testing programs for forensic mtDNA casework laboratories, such as GEDNAP [7], GHEP-ISFG [8–11], or the U.S. Quality Assurance Standards [12] is highly recommended.

In those cases for which forward and reverse coverage are not possible, replicate coverage from the same strand, preferably obtained from different primers, is acceptable. In the end, the quality of the raw electropherograms should dictate the forward and reverse coverage requirements. EMPOP data review of population datasets has shown that redundant coverage is sometimes missing or has only been performed in a subset of samples [13]. Lack of redundancy can lead to so-called phantom mutations (sequencing artifacts reported as actual mutations; [14]) that compromise data quality in forensic database searches. This has been shown in a systematic analysis of numerous mtDNA research studies [15], which demonstrated that erroneous calls tend to occur at particular positions. In addition, errors were shown to be more frequently observed downstream of homopolymeric sequence regions, such as the C-tracts in the hypervariable segments I and II (HVS-I/II; [13]). In both cases, these errors tend to reflect sequencing artifacts in a single strand of data that generally could have been resolved with data from either the complementary strand or a different primer for the same strand.

Recommendation #1
Good laboratory practice and specific protocols for work with mtDNA must be followed in accordance with previous guidelines.

Recommendation #2
Negative and positive controls as well as extraction reagent blanks must be carried through the entire laboratory process.

Recommendation #3
Reported consensus sequences must be based on redundant sequence information, using forward and reverse sequencing reactions whenever practical.

Recommendation #4
Manual transcription of data should be avoided and independent confirmation of consensus haplotypes by two scientists must be performed.

Recommendation #5
Laboratories using mtDNA typing in forensic casework shall participate regularly in suitable proficiency testing programs.

2.2. Targeted region, amplification and sequencing ranges

The CR harbors the vast majority of quickly evolving sites in the mitochondrial genome (mtGenome) that are relevant for the discrimination of haplotypes in the forensic context. Traditional protocols have targeted the hypervariable segments of the CR (usually delimited as follows: HVS-I, 16024–16365; HVS-II, 73–340 and HVS-III, 340–576) using independent amplification and different combinations of primers. This restricted approach has a number of implications. In population studies, the independent amplification of the hypervariable regions, combined with the manual processing of multiple samples at one time, has been shown to lead to an increased risk of chimeric haplotypes or so-called "artificial recombinants" that are caused by the inadvertent mix-up of mtDNA segments from different individuals [16,17]. Laboratory protocols have been established and improved over the past years that support the amplification of the entire CR in a single amplicon (16024-576, [17–20]) for high-quality samples routinely used for databasing purposes. This strategy eliminates the risk of chimeric sequences in databasing applications, and given the ease with which entire CR data can be developed from high-quality samples, the generation of only HVS-I/II data is no longer accepted for reference population databasing purposes.

In forensic casework, artificial recombination is mitigated by strict sample handling measures that include many or all of the

following measures: the processing of only one evidentiary specimen at a time, replicate testing, barcoding of samples and tubes, standardized placement and sample tracking on automated platforms, and other procedures that are deliberately put in place to eliminate such issues. Nevertheless, the recovery of only HVS-I/II data in forensic casework still has its limitations, with the greatest one being that profiles reflecting only portions of the CR restrict the discriminatory power of mtDNA testing. Because sequence information from the entire CR generally results in greater resolution, the recovery of more data is always desirable. In addition, the CR harbors positions outside the two hypervariable segments that display important phylogenetic information, such as position 489, which is a reliable marker to distinguish macro-haplogroups M and N(xJ). As a result, laboratory protocols based on the amplification of shorter overlapping fragments, often in a multiplex configuration, have been developed to facilitate the acquisition of full mtDNA CR sequences from even highly degraded DNA [21–23].

Because of the clear value of targeting the entire mtDNA CR for both casework and population databasing applications, we recommend the generation of entire CR data in both situations when possible, acknowledging that it takes time to validate new methods for casework, and that entire CR recovery may not be feasible in many cases due to sample quality. Moreover, as the field may progress toward the use of "next generation" or "massively parallel" sequencing approaches, it is recommended that protocols are developed to target at least the entire CR.

Recommendation #6
In population genetic studies for forensic databasing purposes, the entire mitochondrial DNA control region should be sequenced.

3. Data analysis, alignment and interpretation

3.1. Reference sequence

For convenient reporting of mtDNA haplotypes in both casework and population databasing applications, a format encoding only the differences relative to the revised version (rCRS (NC001807), [6]) of the first human mtDNA sequence (CRS, [24]) has become routine practice in forensics and other fields of mtDNA research. The rCRS was established from a woman of European descent and thus represents a modern lineage in the human mitochondrial DNA phylogeny. As an alternative to the rCRS, a hypothetical "Reconstructed Sapiens Reference Sequence" (RSRS, [25]) was proffered that was positioned near the root of the human phylogeny (between haplogroups L0 and L1′2′3′4′5′6) with the aim of facilitating haplogroup assignment and the identification of derived mtDNA states. While this approach has appealing features, it would have substantial detrimental consequences for the forensic practitioner, not the least of which is the requirement to translate between old and new nomenclature versions for mostlegacy data – a process that bears a high risk of introducing error [26,27]. Further, the RSRS may be subject to change when additional sequence information becomes available (e.g. [28]). As a result, it currently lacks the required "clerical" stability that would make it durable over time. Continued use of the rCRS has therefore been recommended for the time being [26].

When reporting mtDNA haplotypes relative to the rCRS, the sequence range (excluding primer sequence information) must be reported so that the reader can understand and reconstruct the full sequence. This is of course also critical for database searches, as the specified range restricts the query to identical or overlapping sequence ranges.

Recommendation #7
MtDNA sequences should be aligned and reported relative to the revised Cambridge Reference Sequence (rCRS, NC001807), and should include the interpretation range (excluding primer sequence information).

3.2. Alignment and notation

The previously suggested naming convention of mtDNA substitutions and insertions is still valid with the exception that deletions should be indicated by "DEL", "del", or "−" rather than "D" or "d", since the latter is already used as the International Union of Pure and Applied Chemistry (IUPAC) code for a mixture of G, A and T. In addition, "N" designations should only be used following the IUPAC convention in those cases where all four nucleotides are present at a single position. Other (heteroplasmic) mixtures should be reported using the nomenclature scheme provided by IUPAC. For example, a mixture of A and G should be represented by an R, while a mixture of C and T should be represented by a Y.

The IUPAC code uses capital letters, which allows for the necessary extension of the existing nomenclature to small letters for describing (heteroplasmic) mixtures of deleted/undeleted and inserted/non-inserted bases, respectively [29]. As such 152c would indicate the (heteroplasmic) mixture of a transition and a deletion at nucleotide position T152, while 315.1c describes a mixture of two sequences one of which harbors an insertion at 315.1C while the other does not. An example of a mixture of a deletion and the undeleted variant at position C299 is shown in Fig. 1. Note, that an observed difference to the rCRS is presented as suffix to the position (e.g. 152C) whereas the rCRS variant is notated as prefix (e.g. T152).

One consequence of reporting mtDNA sequences relative to a reference is that multiple alignments of the same string of nucleotides are feasible. For example, the insertion of a C in a homopolymeric tract of seven Cs (e.g. between positions 302 and 310) could be reported as 302.1C or 309.1C (or at any position in between). Because there is no simple molecular approach to determine the "correct" alignment, conventions are required to prevent identical sequences from being reported in different ways. The forensic community traditionally aligns insertions and deletions (indels) 3′ relative to the affected position in homopolymeric sequence tracts and other repeat regions (e.g. the AC repeat between 514 and 524; [2]). Although this convention is in contrast to medical genetics where indels are recorded 5′, the 3′ notation should be maintained to provide consistency with earlier data.

Previously, in an effort to standardize nomenclature of mtDNA haplotypes, a set of formal rules based on (unweighted) maximum parsimony relative to the rCRS was proposed [30] and later modified [31]. Over time, however, it was found that this approach would occasionally produce a greater number of differences between two similar sequences than would exist if the sequences were aligned directly to each other. For example, the HVS-II (73–340) haplotype of sample AKA058 (EMPOP, R9) A73G **G247A A249DEL** A263G A290DEL A291DEL −315.1C differs only by a single transition at position 247 from the standard haplogroup C1 motif A73G **A249DEL** A263G A290DEL A291DEL −315.1C that is e.g. present in ABS184 [32]. Under maximum parsimony AKA058 would need to be aligned as A73G **G247DEL** A263G A290DEL A291DEL −315.1C, which results in two differences (G247DEL A249DEL) between those neighboring sequences and thus creates a so-called jumping alignment (Table 1, see [33] for details and other examples).

It is also the case that the most parsimonious representation of a sequence is often difficult to determine by hand. Consider

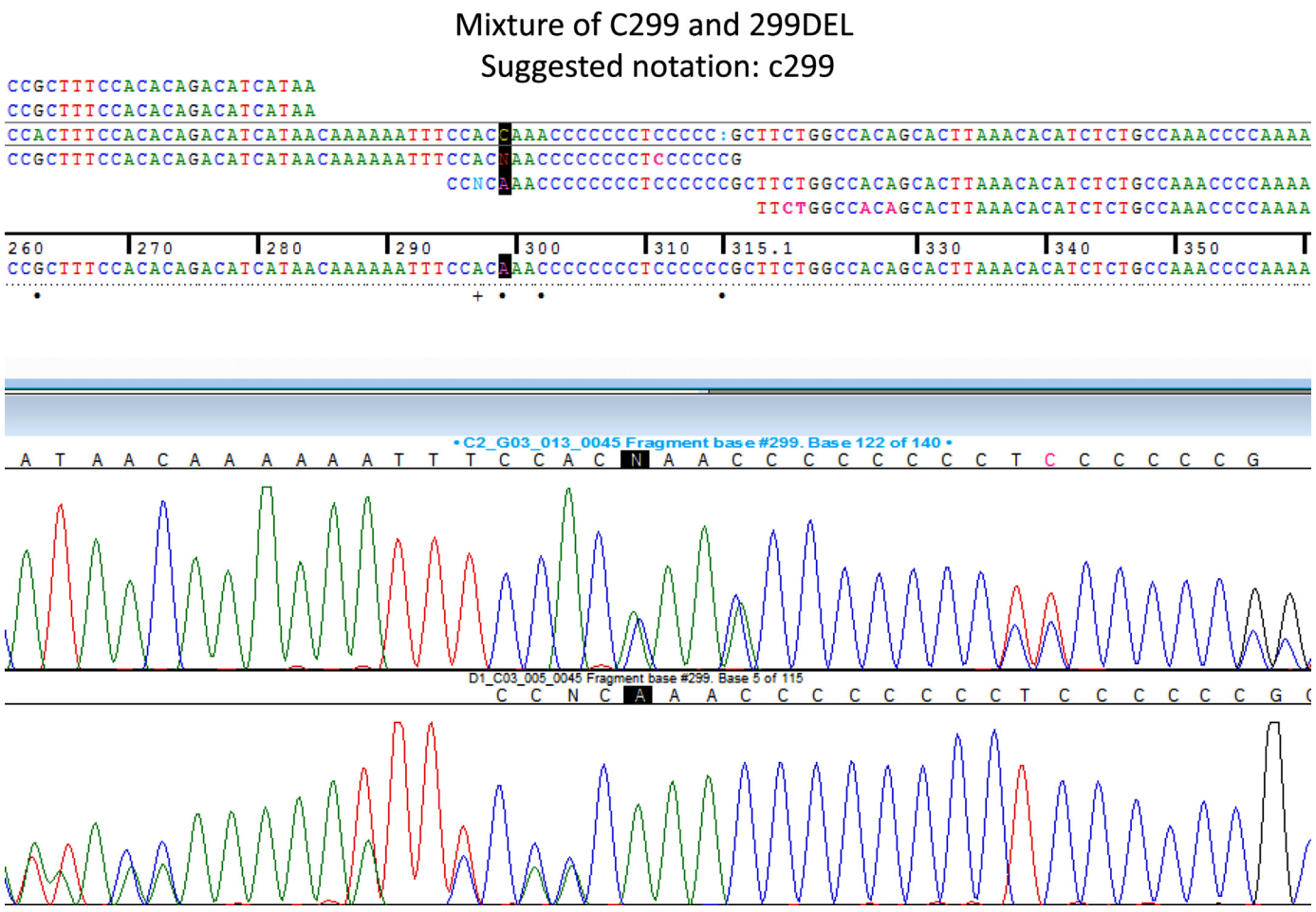

Fig. 1. Example showing a mixture of a deletion at 299 (HVS-II) and the undeleted variant C (=rCRS). The IUPAC code does not provide acronyms for mixtures of deleted/undeleted (inserted/non-inserted) bases. We suggest extending the IUPAC code with lower case letters to describe such mixtures. In this example the assignment would be c299. The upper epg shows the forward sequencing result where the heteroplasmic C/– position is indicated by a C/A base call, following by subsequent base overlaps. The lower epg shows the reverse sequencing result where base overlaps occur upstream from c299.

e.g. the HVS-I motif of haplogroup B4c2 C16147T **A16183C C16184A**T16189C T16217C A16235G C16294T (seven differences to rCRS, Phylotree [34], build 16; e.g. MAS019, EMPOP R9) that would be annotated as C16147T 16182.1C T16189DEL T16217C A16235G C16294T under maximum parsimony (six differences to the rCRS; ESM 1). The latter haplotype however, is difficult to find when performing manual raw data review.

To avoid jumping alignments and facilitate the determination of haplotype representation, phylogenetic nomenclature guidelines have been proposed that base the alignment of difference-coded haplotypes on the established mutation patterns of the mitochondrial phylogeny. Phylogenetic alignment is based on the following rules (from [33]).

1. *Phylogenetic rule*: Sequences should be aligned with regard to the current knowledge of the phylogeny. The most comprehensive repository is Phylotree (www.phylotree.org; [34]). The rCRS-oriented version of Phylotree should be used. In the case of multiple equally plausible solutions, one should strive for maximum (weighted) parsimony. Variants flanking long C tracts, however, are subject to sequence-specific conventions.
2. *C tract conventions*: The long C tracts of HVS-I and HVS-II should always be scored with 16189C and 310C, respectively. Length variation of the short A tract preceding 16184 should be notated preferring transversions unless the phylogeny suggests otherwise.
3. *Indel scoring*: Indels should be placed 3′ with respect to the light strand unless the phylogeny suggests otherwise.

Although no nomenclature scheme can easily address the full complexity of (particularly yet unknown) mtDNA diversity, the phylogenetic approach provides a biological basis for the

Table 1
Example of jumping alignment under maximum parsimony. The haplotype of sample AKA058 shows one difference (247) to the haplotype of sample ABS184 when the two are aligned directly, or aligned to the rCRS following phylogenetic rules. If most parsimonious alignment is applied AKA058 differs at two positions (247, 249) to ABS184 resulting in jumping alignment.

Sample	Haplogroup	Alignment	73	247	249	263	290	291	315.1	Diff. to rCRS	Diff. to AKA058 phylogenetic
AKA058	C1	Phylogenetic	G	A	DEL	G	DEL	DEL	C	7	–
ABS184	C1	Phylogenetic	G	–	DEL	G	DEL	DEL	C	6	1
AKA058	C1	Parsimonious	G	DEL	–	G	DEL	DEL	C	6	2

representation that is desirable for a number of reasons. The phylogenetic alignment conventions are concordant to the practices in other genetic fields with which the forensic community is sharing data on a regular basis. In addition, the phylogenetic notation of haplotypes attempts to describe the actual biological mutations, which will in turn permit the estimation of positional mutation rates and forms the basis for finer-scale, scientifically and biologically driven interpretation of forensic evidence. Tools to assist with the notation of mtDNA sequences are available at http://empop.org/. These practices are used for storing haplotypes in the EMPOP database and have also been adopted by the Scientific Working Group on DNA Methods (SWGDAM) in the United States [5].

While nomenclature issues have negligible practical impact in forensic casework scenarios involving one-to-one matching between known and questioned items analyzed in the same laboratory, standardized reporting of mtDNA variation is fundamental in forensic science as it enables the comparison of mtDNA data from different studies or forensic cases. However, it is important to note that for any alignment convention, including the phylogenetic approach, individual exceptions can be found where multiple alignments are still feasible after application of the rules. As a result, it is desirable to perform alignment-free database searches for estimating the rarity of a given mtDNA haplotype. While practitioners are generally cognizant of motifs that yield multiple potential alignments, and searches of standard rCRS differences-based databases that consider these nuances can be performed, string based searches avoid this issue altogether and enhance the quality of the process. As a result, and in order to ensure that matching haplotypes are not missed in a search due to different nomenclature between query and database sequences, the EMPOP database [35] has implemented this search feature (see also Section 5).

Recommendation #8
IUPAC conventions using capital letters shall be used to describe differences to the rCRS and (point heteroplasmic) mixtures. Lower case letters should be used to indicate mixtures between deleted and non-deleted (inserted and non-inserted) bases. N-designations should only be used when all four bases are observed at a single position (or if no base call can be made at a given position). For the representation of deletions, "DEL", "del" or "−" shall be used.

Recommendation #9
The alignment and notation of mtDNA sequences should be performed in agreement with the mitochondrial phylogeny (established patterns of mutations). Tools to assist with the notation of mtDNA sequences are available at http://empop.org/.

3.3. Heteroplasmy

The coexistence of multiple mtDNA molecules observed in a single source sample is known as heteroplasmy. Systematic studies have revealed that heteroplasmy occurs more often in tissues with high metabolic activity [36], and when mtDNA molecules pass through narrow bottlenecks during development (e.g. in hairs, [37–39]). Tissue-specific heteroplasmy has been further characterized by occasional findings (e.g. T72Y in liver [40] or G71DEL in brain [41]). Here, we distinguish between point (=sequence) heteroplasmy (PHP) and length heteroplasmy (LHP). Because the two differ in frequency and cause, they require individual consideration.

3.3.1. Point heteroplasmy

Point heteroplasmy is detected in Sanger-type sequencing data usually as mixture of two bases at a single position. While single instances of point heteroplasmy across the mtDNA CR are most commonly encountered, observations of two and even three heteroplasmic sites in a single sample have been reported [42,43]. The dye terminator sequencing chemistry that is most frequently used in forensic genetic laboratories is known to display PHP at varying ratios depending on the affected position and primer used, which is why in some cases peak height ratios do not correspond with the relative quantity of the contributing sequence variants. In accordance with the earlier guidelines, it is therefore not feasible to define fixed numerical threshold values for the detection of PHP. Instead, PHP is called as a result of the overall sequence quality and when the majority of sequence strands, ideally in both forward and reverse directions, confirm the presence of the heteroplasmic mixture (ESM 2). In contrast to the earlier guidelines, PHP should be designated with the appropriate IUPAC designation, rather than just "N", to provide a more precise description of the mixed bases.

The occurrence of PHP was described in systematic studies in buccal and blood samples [42] and in hair [43] at varying ratios of 6% and 11.4%, respectively. The affected positions coincided with evolutionary hot spot mutations. The only exceptions to this pattern were the transitional mixtures at positions 214 and 215 which are known to display a higher inter-generational mutation rate [44] but do not become fixed at the population level. The list of heteroplasmic occurrences in Irwin et al. [42] is a useful resource to compare observations in newly generated data. Recent studies employing more sensitive sequencing technologies have suggested that the incidence of heteroplasmy may be higher than previously reported [45]. However, evaluation and validation of these observations for forensic purposes is subject to future research.

Heteroplasmic positions can be searched in EMPOP to evaluate their frequency and phylogenetic background. If this is done, the search needs to be performed in *literal* search format (e.g. 16093Y matches only 16093Y in the sequence range 16093–16093). We note that standard EMPOP database searches involving point heteroplasmic sequences should be performed in a *pattern* mode so that the heteroplasmic mixture at a position does not result in exclusion to either of the contributing bases (e.g. 152Y matches Y, C and T at that position).

The current recommendations address the handling of heteroplasmic samples from the standpoints of raw data analysis and haplotype reporting. Laboratories should evaluate heteroplasmy based on the limitation of the Sanger sequencing and electrophoresis technology, the quality of the sequence data and the experience of the laboratory. It is important to note that from the standpoint of interpretation, differences in both PHP and LHP do not constitute evidence for excluding two otherwise identical haplotypes as deriving from the same source or same maternal lineage. In fact, the same PHP observed in identical haplotypes may increase the strength of the evidence [46].

3.3.2. Length heteroplasmy

Length heteroplasmy (LHP) is typically observed when homopolymeric tracts exceed eight identical nucleotides [47]. This is the case in HVS-I when a transition at position T16189C leads to an uninterrupted stretch of 10 Cs, resulting in extensive LHP with fragments consisting of up to 14 Cs as well as overlaid and generally uninterpretable sequences downstream of this position (ESM 3). In West Eurasian populations this phenomenon is observed in about 14% of samples tested (based on a review of 852 high quality EMPOP sequences) and can reach much higher values in phylogenetic backgrounds where T16189C constitutes a haplogroup-specific signature mutation (e.g. hg B, East Asia, 36%; $N = 637$ EMPOP sequences).

LHP is even more frequently observed in the HVS-II C tract between positions 302 and 310 where the stretch of seven Cs can be prolonged by a single or up to four C insertions (74% in 1887

specifically investigated EMPOP sequences showed insertions in this region). The transition at position T310 (0.5% EMPOP, R9) also leads to extensive LHP as the two adjacent C tracts of 7 and 5C (rCRS-variant) are combined into an uninterrupted stretch of 13 Cs. However, 13C residues have rarely been observed in that region and a prolongation of the C tract has not been reported for such instances. On the contrary, this C tract usually tends to be contracted down to nine Cs in total (ESM 3).

In HVS-III, length heteroplasmy is typically induced by C insertions between positions 568–573 (ESM 3) and, unlike 16193 and 309 which occur throughout the phylogeny, often has a phylogenetic signature, e.g. in haplogroup I. As a result, its occurrence also depends on the phylogenetic background of the population at hand (0.3% in EMPOP, R9).

Although LHP in the above-mentioned regions is quite common, there is no routine practice for its interpretation and reporting in forensic casework. Some laboratories disregard the C tract in HVS-I by skipping 10 bases from position 16184 and reporting differences to the rCRS only further downstream. Other laboratories report all observed LHP variants. In either case, it is important to understand that the detection of LHP strongly depends on the sequencing technology and chemistry used, and that the data may not necessarily represent all the LHP variants and their proportions in the tissue. Even more importantly, the distribution of LHP variants is known to vary within and between tissues of an individual [37]. As a result of this variability, insertions and deletions in these polycytosine regions are ignored in direct forensic comparisons and database searches; and it is therefore up to the individual laboratories to establish internal guidelines for the representation and reporting of LHP in forensic casework.

For population genetic databasing in the forensic context, on the other hand, and in the interest of standardization and consistency in the representation of database profiles, we recommend the reporting of the dominant LHP variant (major molecule) in the data [48]. This can be achieved by selecting the highest representation of a single, non-repetitive nucleotide downstream from the heteroplasmic region that shows the detected distribution of length variants and therefore serves as a signpost for the dominant variant. For example, G16196 may be used as a signpost for HVS-I, T310 and G316 for HVS-II, G527 for the dimeric AC repeat in HVS-III, and G577 for the HVS-III C tract [49]. The reverse sequencing reaction displays such indicators at positions G16156, G275, G513 and G564, respectively. Alternative signposts may need to be found in another sequence context. The dominant LHP type can be identified in the majority of cases. However, in those rare instances where two or more LHP variants are equally dominant, it has become practice to determine the shortest dominant LHP variant (e.g. EMPOP).

Recommendation #10
In forensic casework, laboratories must establish their own interpretation and reporting guidelines for observed length and point heteroplasmy. The evaluation of heteroplasmy depends on the limitations of the technology and the quality of the sequencing reactions as well as the experience of the laboratory. Differences in both PHP and LHP do not constitute evidence for excluding two otherwise identical haplotypes as deriving from the same source or same maternal lineage.

Recommendation #11
For population database samples, length heteroplasmy in homopolymeric sequence stretches should be interpreted by calling the dominant variant, which can be determined by identifying the position with the highest representation of a non-repetitive peak downstream of the affected stretch.

4. Quality control of population data

Over the past decade, many types of errors have been detected in a number of mtDNA population genetic studies [50]. These errors have been traced back to sample handling (e.g. the artificial recombination; referenced in Section 2.2) as well as clerical and transcription mistakes. Most commonly, however, the errors have been attributed to sequence interpretation mistakes deriving from poor raw data quality [8,51]. Steps to minimize these types of errors are standard practice in forensic casework, and many of the same steps should routinely be employed in the generation of reference population data, e.g. redundant sequence information, independent double evaluation of the data, electronic data handling and transfer [52]. Notwithstanding such steps, given the numerous potential sources of error in mtDNA sequencing, additional quality control measures should be applied to mtDNA population data.

4.1. Quasi-median network analysis

Because established mutation patterns (phylogenetic motifs) based on currently known mtDNA variation serve as powerful means of uncovering data idiosyncrasies [29], phylogenetic analysis offers an *a posteriori* tool for detecting errors. When haplotypes are analyzed and reviewed this way, unusual or previously unobserved variants stand out because they cannot be found within the spectrum of known mtDNA variation. The phylogenetic analysis, then, serves to highlight those positions/variants that should be confirmed by double checking the raw sequence data. In recent years, a phylogenetic tool developed and designed by EMPOP specifically for these purposes and based on quasi-median network analysis (QMN) has been successfully employed as a quality control measure for mtDNA reference population datasets [4,13,53]. Because of the effectiveness of phylogenetic tests for data quality control, the editors of the forensic journals *Forensic Science International: Genetics* and *International Journal of Legal Medicine* now require authors to submit their mtDNA population data to EMPOP for quality control prior to manuscript submission [54,55]. The QMN application is freely available on the EMPOP website, and detailed information on QMN analysis and its QC applications can be found there.

4.2. Haplogrouping of mtDNA sequences

An additional tool for the evaluation of mtDNA datasets is haplogroup assignment. Properly ascribing an mtDNA sequence to its evolutionary haplogroup puts the pattern of observed differences to the rCRS into a phylogenetic context that is useful for both QC purposes and phylogeographic considerations [56]. The haplogroup affiliation of a sample together with the haplogroup distribution of the dataset provide useful information about the composition of the population at hand, and provides a basis for comparison to other datasets. In addition, the haplogroup status of an mtDNA sequence, when regarded in the light of the worldwide phylogeny, provides information about the distribution of the haplotype on a global scale that may impact the selection of database subsets for forensic searches (see Section 5). Haplogrouping requires experience and cognizance of the phylogeny. It has been greatly simplified with the provision of Phylotree ([34]; www.phylotree.org; current Build 16), a curated phylogenetic tree of established haplogroups based on full mtGenomes. The haplogroup-defining polymorphisms listed in Phylotree facilitate manual haplogroup assignment, but they also serve as basis for software applications to automate this task. Recently, a new software tool for haplogroup assignment, EMMA, has been implemented in EMPOP Ver. 3 [57]. EMMA is the only automated

tool that uses both the virtual Phylotree haplotypes as well as a curated selection of 14,990 real haplotypes (full mtGenomes) to assign haplogroups. The inclusion of a large number of naturally occurring haplotypes of known haplogroup affiliation in the analytical algorithm provides a more accurate haplogroup estimate. Details of the software can be found in Röck et al. [57] and on the EMPOP website. We also note that automated haplogroup assignment always requires review of the proposed results as in some cases additional information (geography) may be required to resolve ambiguous results (see ESM 8 in [57]).

Recommendation #12
MtDNA population data should be subjected to analytical software tools that facilitate phylogenetic checks for data quality control. A comprehensive suite of QC tools is provided by EMPOP.

5. Databases and database searches

5.1. Source population for forensic database searches

In order to assess the significance of an mtDNA match, it is common and accepted practice to determine the frequency of the mtDNA type in question in a relevant population database. The choice of the population database is an extremely important factor [58], with large effect on magnitude of the reported significance (ESM 4). Ideally, the database would accurately represent the pool of potential contributors of the mtDNA type without reference to the geographic origin of the suspect in question. This is straightforward in principle with homogenous populations, where a single local or regional database can be expected to represent the potential contributors. In cases where different geographic databases can be demonstrated to have insignificant population differentiation at the mtDNA level, databases may be pooled to increase the discrimination provided by rare haplotypes (ESM 4). However, many events or circumstances in which mtDNA testing is performed have potential contributors that come from populations that differ in mtDNA composition, for example in urban areas where the general population is an admixture of people of different geographic origins. In such cases, choice and availability of appropriate databases is more complicated. In such instances it can be appropriate to report the haplotype frequencies in various databases. Local, regional and continental databases may all be searched and reported to guide evaluation of the evidence. Additionally, frequencies from geographic databases of relevant differentiated populations may be reported (as in the United States where reports often list separately search results from major population groups). Mass fatality victim identification events will often involve multiple population origins, and are an example where the context of the case must be carefully considered in choice and evaluation of databases. Overall, the choice of database(s)used for reporting should be carefully considered in regard to the context of the case, and care taken to ensure that the significance of a match is presented conservatively with regard to circumstances of the case and the uncertainties that may exist. It should be born in mind that frequencies of mtDNA types can vary significantly at the very local level (e.g. [59,60]).

5.2. Statistical evaluation of database search result

With regard to reporting frequencies or probabilities of haplotypes based on database searches, there are a number of accepted practices reflecting different conceptual approaches, which can give different results, but with little effect on overall magnitude (Table 2). An extremely conservative approach with regard to avoiding underestimates of population frequency is to report the upper bound of a 95% confidence limit calculated by [61]. Alternatively, point estimates for probabilistic approaches are often given by $(X + 1)/(N + 1)$ or $(X + 2)/(N + 2)$ [62,63], where X is the number of observations in the database and N is the database size; the latter equation takes the conceptual approach of adding the case profiles to the database, under the hypothesis that the profiles observed in the case may have come from two different individuals, i.e. the suspect and the unknown offender (as is common in the denominator of Likelihood Ratio calculations). Likelihood ratios for mtDNA then are normally calculated by 1/(match probability).

Egeland et al. [64] further elaborate on the estimation of rare or unseen mtDNA haplotype frequencies and compare the classical approach with new proposals based on Principal Component Analysis; specific software is also provided. These authors also discuss on the importance of sample size and coverage for the estimation of haplotype frequencies. The population distribution of mtDNA haplotypes is typified by a large number of very rare or unique haplotypes, also termed 'singletons', and probabilistic approaches that empirically take into account the haplotype distribution of the database in question (e.g. the Kappa model; [65]) may best represent the strength of the mtDNA evidence.

5.3. Combining nuclear and mitochondrial DNA testing results

In some cases, usually involving either partial autosomal STR profiles and/or kinship analysis with distant relatives, it may be desirable to combine the results of mtDNA testing with those of autosomal STR profiles. If calculated as Likelihood Ratios, the LR's of the different systems may be combined by multiplication, if the databases for each can be justified to adequately represent the same source pool, and if the hypotheses involved in the mtDNA and autosomal LRs are the same, and take into consideration that maternal relatives cannot be distinguished by mtDNA.

5.4. Criteria for mtDNA database searches

5.4.1. Data quality and search method

The database should harbor high-quality mtDNA data in order to avoid biased estimates due to erroneous sequence entries. The

Table 2
Comparison of methods for assessing the significance of mtDNA matching with different accepted approaches. Match probability or frequency estimates and corresponding LRs are listed side by side. The number of singletons was determined by SAM [35].

Matches	Singletons	p_{uncorr}	p_{N+1}	LR p_{N+1}	p_{N+2}	LR p_{N+2}	$f_{CL\ Zero}$	LR $f_{CL\ zero}$	$LR_{(Kappa)}$
0/1000	900	–	0.999E−03	1001	1.996E−03	501	2.991E−03	334	10,000
0/1000	700	–	0.999E−03	1001	1.996E−03	501	2.991E−03	334	3333
0/1000	100	–	0.999E−03	1001	1.996E−03	501	2.991E−03	334	1111
0/50,000	45,000	–	1.999E−05	50,001	3.999E−05	25,001	5.991E−05	16,690	500,000
0/50,000	35,000	–	1.999E−05	50,001	3.999E−05	25,001	5.991E−05	16,690	166,667
0/50,000	5000	–	1.999E−05	50,001	3.999E−05	25,001	5.991E−05	16,690	55,556

p_{uncorr} – uncorrected probability; p_{N+1} – probability with haplotype being added to database once; p_{N+2} – probability with haplotype being added to database twice; $f_{CL\ Zero}$ – frequency of confidence limit from zero proportion; $LR_{(Kappa)}$ – LR according to the Kappa model [65] taking into account the number of singletons in a database.

entire database should be targeted in a search to avoid biased results. The choice of a subset of the database query is strongly depending on the context of the case (see also Section 5.1).

Database queries require the definition of the sequencing (interpretation) range(s) and the observed differences relative to the rCRS. Because rCRS-coded haplotypes may result in more than one possible alignment, it is desirable to perform alignment-free database searches for estimating the rarity of a given mtDNA haplotype. In order to ensure that matching haplotypes are not missed in a search due to different nomenclature between query and database sequences, the EMPOP database has implemented this search feature.

5.4.2. Highly variable positions and heteroplasmy for mtDNA searches

Polymorphisms that are known to be highly variable within tissues of an individual should be ignored from a database search (default settings in EMPOP); otherwise the search result may be non-conservative. This involves the homopolymeric C-tracts around positions 16189, 310 and between positions 568 and 574 as well as T insertions around position 452.

Heteroplasmic mixtures should be called using IUPAC conventions and searched in pattern format, e.g. T152Y matches T152, T152C and T152Y. We note that the literal search mode (i.e. T152Y matches only T152Y) is a useful option to examine the database for the presence of sequences that show heteroplasmy at position 152.

Recommendation #13
The entire database of available sequences should be searched with respect to the sequencing (interpretation) range to avoid biased query results.

Recommendation #14
Laboratories must be able to justify the choice of database(s) and statistical approach used in reporting.

Recommendation #15
Laboratories must establish statistical guidelines for use in reporting an mtDNA match between two samples.

Recommendation #16
Highly variable positions such as length variants in homopolymeric stretches should be disregarded from searches for determining frequency estimates. Heteroplasmic calls should be queried in a manner that does not exclude any of the heteroplasmic variants.

Acknowledgements

The research leading to this publication was funded in part by the Austrian Science Fund (FWF) [P22880-B12] and TR L397, as well as by the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 285487 (EUROFORGEN-NoE). It was also supported by Award No. 2011-MU-MU-K402 to Jodi A. Irwin, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The funding was administered by the American Registry of Pathology. Neither the U.S. Department of Justice nor the American Registry of Pathology had any role in study design; collection, analysis or interpretation of data; in the writing of this report; or in the decision to submit this paper for publication. A.S. received funding from the Ministerio de Ciencia e Innovación (SAF2011-26983) and from the Xunta de Galicia (EM 2012/045; Modalidad REDES: 2012-PG226). The authors would like to thank the staff of the Institute of Legal Medicine Innsbruck (GMI) for outstanding technical assistance throughout the past 15 years of EMPOP IT development, data generation and sequence review. We would like to thank Bettina Zimmermann and Gabriela Huber for collecting raw data for figural examples in this publication and Rebecca S. Just for useful discussions. Arne Dür (Institute of Mathematics, University of Innsbruck) and Alexander Röck are greatly acknowledged for the development of specific EMPOP software tools including SAM and EMMA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.07.010.

References

[1] M.M. Holland, T. Parsons, Mitochondrial DNA sequence analysis – validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 21–50.
[2] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M.M. Holland, G. Tully, M. Wilson, DNA commission of the International Society of Forensic Genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.
[3] SWGDAM, Guidelines for mitochondrial DNA (mtDNA) Nucleotide sequence information, Forensic Sci. Commun. 5 (2003).
[4] W. Parson, A. Dür, EMPOP – a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88–92.
[5] Scientific Working Group on DNA Analysis Methods (SWGDAM), Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories, 2013, pp. 1–23 http://swgdam.org/SWGDAM%20mtDNA_Interpretation_Guidelines_APPROVED_073013.pdf.
[6] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 14.
[7] C. Hohoff, K. Schnöink, B. Brinkmann, P.M. Schneider, The GEDNAP (German DNA Profiling Group) Proficiency Testing System, 2013 http://www.gednap.org.
[8] A. Salas, L. Prieto, M. Montesino, C. Albarran, E. Arroyo, M.R. Paredes-Herrera, A.M. Di Lonardo, C. Doutremepuich, I. Fernandez-Fernandez, A.G. de la Vega, C. Alves, C.M. Lopez, M. Lopez-Soto, J.A. Lorente, A. Picornell, R.M. Espinheira, A. Hernandez, A.M. Palacio, M. Espinoza, J.J. Yunis, A. Perez-Lezaun, J.J. Pestano, J.C. Carril, D. Corach, M.C. Vide, V. Alvarez-Iglesias, M.F. Pinheiro, M.R. Whittle, A. Brehm, J. Gomez, D.N.A. Mitochondrial, error prophylaxis: assessing the causes of errors in the GEP'02-03 proficiency testing trial, Forensic Sci. Int. 148 (2005) 191–198.
[9] M. Montesino, A. Salas, M. Crespillo, C. Albarrán, A. Alonso, V. Alvarez-Iglesias, J.A. Cano, M. Carvalho, D. Corach, C. Cruz, A. Di Lonardo, R. Espinheira, M.J. Farfán, S. Filippini, J. García-Hirschfeld, A. Hernández, G. Lima, C.M. López-Cubría, M. López-Soto, S. Pagano, M. Paredes, M.F. Pinheiro, A.M. Rodríguez-Monge, A. Sala, S. Sóñora, D.R. Sumita, M.C. Vide, M.R. Whittle, A. Zurita, L. Prieto, Analysis of body fluid mixtures by mtDNA sequencing: an inter-laboratory study of the GEP-ISFG working group, Forensic Sci. Int. 168 (2007) 42–56.
[10] L. Prieto, B. Zimmermann, A. Goios, A. Rodriguez-Monge, G.G. Paneto, C. Alves, A. Alonso, C. Fridman, S. Cardoso, G. Lima, M.J. Anjos, M.R. Whittle, M. Montesino, R.M. Cicarelli, A.M. Rocha, C. Albarrán, M.M. de Pancorbo, M.F. Pinheiro, M. Carvalho, D.R. Sumita, W. Parson, The GHEP-EMPOP collaboration on mtDNA population data – a new resource for forensic casework, Forensic Sci. Int. Genet. 5 (2011) 146–151.
[11] L. Prieto, C. Alves, B. Zimmermann, A. Tagliabracci, V. Prieto, M. Montesino, M.R. Whittle, M.J. Anjos, S. Cardoso, B. Heinrichs, A. Hernandez, A.M. Lopez-Parra, A. Sala, V.G. Saragoni, G. Burgos, M. Marino, M. Paredes, C.A. Mora-Torres, R. Angulo, G. Chemale, C. Vullo, M. Sanchez-Simon, D. Comas, J. Puente, C.M. Lopez-Cubria, N. Modesti, M. Aler, S. Merigioli, E. Betancor, S. Pedrosa, G. Plaza, M.V. Masciovecchio, P.M. Schneider, W. Parson, GHEP-ISFG proficiency test 2011: paper challenge on evaluation of mitochondrial DNA results, Forensic Sci. Int. Genet. 7 (2013) 10–15.
[12] Federal Bureau of Investigation Quality Assurance Standards for Forensic DNA Testing Laboratories, 2011.
[13] B. Zimmermann, A.W. Röck, A. Dür, W. Parson, Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software, Croat. Med. J. 55 (2014) 115–120.
[14] H.-J. Bandelt, L. Quintana-Murci, A. Salas, V. Macaulay, The fingerprint of phantom mutations in mitochondrial DNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.
[15] A. Brandstätter, T. Sänger, S. Lutz-Bonengel, W. Parson, E. Beraud-Colomb, B. Wen, Q.P. Qong, C.M. Bravi, H.-J. Bandelt, Phantom mutation hotspots in human mitochondrial DNA, Electrophoresis 26 (2005) 3414–3429.
[16] H.-J. Bandelt, A. Salas, S. Lutz-Bonengel, Artificial recombination in forensic mtDNA population databases, Int. J. Legal Med. 118 (2004) 267–273.
[17] W. Parson, H.-J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (2007) 13–19.
[18] A. Brandstätter, C.T. Peterson, J.A. Irwin, S. Mpoke, D.K. Koech, W. Parson, T.J. Parsons, Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database, Int. J. Legal Med. 118 (2004) 294–306.
[19] A. Brandstätter, R. Klein, N. Duftner, P. Wiegand, W. Parson, Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm), Int. J. Legal Med. 120 (2006) 310–314.

[20] J.A. Irwin, J.L. Saunier, K.M. Strouss, K.A. Sturk, T.M. Diegoli, R.S. Just, M.D. Coble, W. Parson, T.J. Parsons, Development and expansion of high-quality control region databases to improve forensic mtDNA evidence interpretation, Forensic Sci. Int. Genet. 1 (2007) 154–157.
[21] C. Eichmann, W. Parson, 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples, Int. J. Legal Med. 122 (2008) 385–388.
[22] C. Berger, W. Parson, Mini-midi-mito: adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples, Forensic Sci. Int. Genet. 3 (2009) 149–153.
[23] M.N. Gabriel, E.F. Huffine, J.H. Ryan, M.M. Holland, T.J. Parsons, Improved MtDNA sequence analysis of forensic remains using a mini-primer set amplification strategy, J. Forensic Sci. 46 (2001) 247–253.
[24] S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, Nature 290 (1981) 457–465.
[25] D.M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.L. Loogvali, N.M. Silva, T. Kivisild, A. Torroni, R. Villems, A "Copernican" reassessment of the human mitochondrial DNA tree from its root, Am. J. Hum. Genet. 90 (2012) 675–684.
[26] A. Salas, M. Coble, S. Desmyter, T. Grzybowski, L. Gusmão, C. Hohoff, M.M. Holland, J.A. Irwin, T. Kupiec, H.Y. Lee, B. Ludes, S. Lutz-Bonengel, T. Melton, T.J. Parsons, H. Pfeiffer, L. Prieto, A. Tagliabracci, W. Parson, A cautionary note on switching mitochondrial DNA reference sequences in forensic genetics, Forensic Sci. Int. Genet. 6 (2012) e182–e184.
[27] H.-J. Bandelt, A. Kloss-Brandstätter, M.B. Richards, Y.G. Yao, I. Logan, The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies, J. Hum. Genet. 59 (2014) 66–77.
[28] B.A. Malyarchuk, Improving the reconstructed sapiens reference sequence of mitochondrial DNA, Forensic Sci. Int. Genet. 7 (2013) e74–e75.
[29] H.-J. Bandelt, A. Dür, Translating DNA data tables into quasi-median networks for parsimony analysis and error detection, Mol. Phylogenet. Evol. 42 (2007) 256–271.
[30] M.R. Wilson, M.W. Allard, K.L. Monson, K.W.P. Miller, B. Budowle, Recommendations for consistent treatment of length variants in the human mtDNA control region, Forensic Sci. Int. 129 (2002) 35–42.
[31] B. Budowle, D. Polanskey, C.L. Fisher, B.K. Den Hartog, R.B. Kepler, J.W. Elling, Automated alignment and nomenclature for consistent treatment of polymorphisms in the human mitochondrial DNA control region, J. Forensic Sci. 55 (2010) 1190–1195.
[32] M.C. Bobillo, B. Zimmermann, A. Sala, G. Huber, A.W. Röck, H.J. Bandelt, D. Corach, W. Parson, Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database, Int. J. Legal Med. 124 (2010) 263–268.
[33] H.-J. Bandelt, W. Parson, Consistent treatment of length variants in the human mtDNA control region: a reappraisal, Int. J. Legal Med. 122 (2008) 11–21.
[34] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2009) E386–E394.
[35] A.W. Röck, J. Irwin, A. Dür, T. Parsons, W. Parson, SAM: string-based sequence search algorithm for mitochondrial DNA database queries, Forensic Sci. Int. Genet. 5 (2011) 126–132.
[36] C.D. Calloway, R.L. Reynolds, G.L. Herrin Jr., W.W. Anderson, The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age, Am. J. Hum. Genet. 66 (2000) 1384–1397.
[37] A. Salas, M.V. Lareu, A. Carracedo, Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report, Int. J. Legal Med. 114 (2001) 186–190.
[38] T. Melton, Mitochondrial DNA Heteroplasmy, Forensic Sci. Rev. 16 (2004) 2–19.
[39] G. Tully, S.M. Barritt, K. Bender, E. Brignon, C. Capelli, N. Dimo-Simonin, C. Eichmann, C.M. Ernst, C. Lambert, M.V. Lareu, B. Ludes, B. Mevag, W. Parson, H. Pfeiffer, A. Salas, P.M. Schneider, E. Staalstrom, Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts, Forensic Sci. Int. 140 (2004) 1–11.
[40] D.C. Samuels, C. Li, B. B. Li, Z. Song, E. Torstenson, H. Boyd Clay, A. Rokas, T.A. Thornton-Wells, J.H. Moore, T.M. Hughes, R.D. Hoffman, J.L. Haines, D.G. Murdock, D.P. Mortlock, S.M. Williams, Recurrent tissue-specific mtDNA mutations are common in humans, PLoS Genet. 9 (2013) e1003929.
[41] E.E. Jazin, L. Cavelier, I. Eriksson, L. Oreland, U. Gyllensten, Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 12382–12387.
[42] J. Irwin, J. Saunier, H. Niederstätter, K. Strouss, K. Sturk, T. Diegoli, A. Brandstätter, W. Parson, T. Parsons, Investigation of point heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from over 5000 global population samples, J. Mol. Evol. 68 (2009) 516–527.
[43] T. Melton, G. Dimick, B. Higgins, L. Lindstrom, K. Nelson, Forensic mitochondrial DNA analysis of 691 casework hairs, J. Forensic Sci. 50 (2005) 73–80.
[44] T.J. Parsons, D.S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner, M.R. Wilson, D.L. Berry, K.A. Holland, V.W. Weedn, P. Gill, M.M. Holland, A high observed substitution rate in the human mitochondrial DNA control region, Nat. Genet. 15 (1997) 363–368.
[45] M.M. Holland, M.R. McQuillan, K.A. O'Hanlon, Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy, Croat. Med. J. 52 (2011) 299–313.
[46] P.L. Ivanov, M.J. Wadhams, R.K. Roby, M.M. Holland, V.W. Weedn, T.J. T.J. Parsons, Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II, Nat. Genet. 12 (1996) 417–420.
[47] W. Parson, T.J. Parsons, R. Scheithauer, M.M. Holland, Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case, Int. J. Legal Med. 111 (1998) 124–132.
[48] C. Berger, P. Hatzer-Grubwieser, C. Hohoff, W. Parson, Evaluating sequence-derived mtDNA length heteroplasmy by amplicon size analysis, Forensic Sci. Int. Genet. 5 (2011) 142–145.
[49] L. Forster, P. Forster, S.M. Gurney, M. Spencer, C. Huang, A. Rohl, B. Brinkmann, Evaluating length heteroplasmy in the human mitochondrial DNA control region, Int. J. Legal Med. 124 (2009) 133–142.
[50] A. Salas, A. Carracedo, V. Macaulay, M. Richards, H.J. Bandelt, A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics, Biochem. Biophys. Res. Commun. 335 (2005) 891–899.
[51] W. Parson, A. Brandstätter, A. Alonso, N. Brandt, B. Brinkmann, A. Carracedo, D. Corach, O. Froment, I. Furac, T. Grzybowski, K. Hedberg, C. Keyser-Tracqui, T. Kupiec, S. Lutz-Bonengel, B. Mevag, R. Ploski, H. Schmitter, P. Schneider, D. Syndercombe-Court, E. Sorensen, H. Thew, G. Tully, R. Scheithauer, The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives, Forensic Sci. Int. 139 (2004) 215–226.
[52] A. Brandstätter, H. Niederstätter, M. Pavlic, P. Grubwieser, W. Parson, Generating population data for the EMPOP database – an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example, Forensic Sci. Int. 166 (2007) 164–175.
[53] B. Zimmermann, A. Röck, G. Huber, T. Krämer, P.M. Schneider, W. Parson, Application of a west Eurasian-specific filter for quasi-median network analysis: sharpening the blade for mtDNA error detection, Forensic Sci. Int. Genet. 5 (2011) 133–137.
[54] W. Parson, L. Roewer, Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine, Int. J. Legal Med. 124 (2010) 505–509.
[55] A. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, Update of the guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 10 (2014) A1–A2.
[56] H.-J. Bandelt, M. Oven, A. Salas, Haplogrouping mitochondrial DNA sequences in Legal Medicine/Forensic Genetics, Int. J. Legal Med. 126 (2012) 901–916.
[57] A.W. Röck, A. Dür, M. van Oven, W. Parson, Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA), Forensic Sci. Int. Genet. 7 (2013) 601–609.
[58] A. Salas, H.-J. Bandelt, V. Macaulay, M.B. Richards, Phylogeographic investigations: the role of trees in forensic genetics, Forensic Sci. Int. 168 (2007) 1–13.
[59] A. Gómez-Carballa, A. Ignacio-Veiga, V. Alvarez-Iglesias, A. Pastoriza-Mourelle, Y. Ruíz, L. Pineda, A. Carracedo, A. Salas, A melting pot of multicontinental mtDNA lineages in admixed Venezuelans, Am. J. Phys. Anthropol. 147 (2012) 78–87.
[60] A. Gómez-Carballa, J. Pardo-Seco, L. Fachal, A. Vega, M. Cebey, N. Martinón-Torres, F. Martinón-Torres, A. A. Salas, Indian signatures in the westernmost edge of the European Romani diaspora: new insight from mitogenomes, PLOS ONE 8 (2013) e75397.
[61] C. Clopper, E.S. Pearson, The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika 26 (1934) 404–413.
[62] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci. Int. 64 (1994) 125–140.
[63] T. Egeland, A. Salas, Estimating haplotype frequency and coverage of databases, PLoS ONE 3 (2008) e3988.
[64] T. Egeland, H.M. Bøvelsta, G.O. Storvik, A. Salas, Inferring the most likely geographical origin of mtDNA sequence profiles, Ann. Hum. Genet. 68 (2004) 461–471.
[65] C.H. Brenner, Fundamental problem of forensic mathematics – the evidential value of a rare haplotype, Forensic Sci. Int. Genet. 4 (2010) 281–291.