The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives

Walther Parsona,*, Anita Brandstättera, Antonio Alonob, Nathalie Brandtc, Bernd Brinkmannd, Angel Carracedoe, Daniel Corachf, Olivier Fromentg, Ivana Furach, Tomasz Grzybowskii, Karin Hedbergj, Christine Keyser-Tracquik, Tomasz Kupiec, Sabine Lutz-Bonengelm, Bente Mevag, Rafal Ploskio, Hermann Schmitterp, Peter Schneiderq, Denise Syndercombe-Courtr, Eric Sørensen, Heather Thew, Gillian Tullyu, Richard Scheithauerab

a Institute of Legal Medicine, Medical University of Innsbruck, Muellerstrasse 44, Innsbruck 6020, Austria
b Instituto Nacional de Toxicologia, Ministry of Justice, Madrid, Spain
c National Institute of Criminalistics and Criminology, Bruxelles, Belgium
d Institute of Legal Medicine, University of Munster, Munster, Germany
e Institute of Legal Medicine, University of Santiago de Compostela, Santiago, Spain
f Facultad de Farmacia y Bioquimica, Buenos Aires, Argentina
g Institute of Pathology and Genetic, Gerpinnes, Belgium
h Department of Forensic Medicine, University of Zagreb, Zagreb, Croatia
i Institute of Molecular and Forensic Genetics, Medical University of Bydgoszcz, Bydgoszcz, Poland
j National Laboratory of Forensic Science, SKL, Linkoeping, Sweden
k Laboratoire CODGENE, Institute of Legal Medicine, Strasbourg, France
l Institute of Forensic Research, Krakow, Poland
m Institute of Legal Medicine, University of Freiburg, Freiburg, Germany
n Institute of Legal Medicine, University of Oslo, Oslo, Norway
o Department of Forensic Medicine, University of Warsaw, Warsaw, Poland
p Bundeskriminalamt Wiesbaden, Wiesbaden, Germany
q Institute of Legal Medicine, University of Mainz, Mainz, Germany
r Department of Haematology, Barts and The London, Queen Mary’s School of Medicine and Dentistry, London, UK
s Department of Forensic Genetics, Institute of Legal Medicine, University of Copenhagen, Copenhagen, Denmark
t Armed Forces DNA Identification Laboratory, Rockville, MD, USA
u Forensic Science Service, Trident Court, Birmingham, UK

Received 23 July 2003; received in revised form 27 October 2003; accepted 7 November 2003

Abstract

This paper presents an overview of the organisation and the results of the collaborative exercises (CE) of the European DNA Profiling (EDNAP) Group’s mitochondrial DNA population database project (EMPOP). The aim of the collaborative exercises was to determine whether uniformity of mtDNA sequencing results could be achieved among different laboratories. These were asked to sequence either the complete mtDNA control region or the two hypervariable regions HVI (16024–16365) and HVII (73–340) from DNA extracts, buccal swabs or bloodstains, proceeding in accordance with the protocol and strategies used in each individual laboratory. The results of the collaborative exercises were employed to identify possible sources of errors that could arise during the analysis and interpretation of mtDNA profiles. These findings were taken as a basis to tentatively make suitable arrangements for the construction of a high quality mtDNA database. One hundred fifty mtDNA profiles were submitted to the evaluating laboratory, and disaccording profiles were classified into four groups corresponding to the source of error:

* Corresponding author. Tel.: +43-512-507-3314; fax: +43-512-507-2764.
E-mail address: walther.parson@uibk.ac.at (W. Parson).

0379-0738/$ – see front matter © 2003 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2003.11.008
1. Introduction

Mitochondrial DNA (mtDNA) typing often is the last resort in forensic casework and identification, when nuclear markers fail to give conclusive results. The high copy number of mtDNA ensures a better PCR amplification success rate over nuclear loci. The mode of inheritance of mtDNA (maternal transmission and lack of recombination) has the consequence that mtDNA types are not unique but shared by maternally related individuals. Given an exact mtDNA sequence match between a crime stain and a suspect, it is of interest to determine the frequency of the sequence in the population in order to calculate the probability of a chance match. Generally, the expected frequency of an mtDNA haplotype is estimated by comparison to a collection of mtDNA haplotypes in a population database.

Recently, the quality of published mtDNA population data has been under severe criticism [1-5]. The authors demonstrated that published mtDNA sequence data are prone to contain errors, mainly due to misinterpretation of raw sequence data (phantom mutations) and/or the introduction of clerical errors in the process of data transcription. In the majority of cases the errors can be attributed to an isolated event within an mtDNA haplotype, which may lead to a biologically non-existing sequence. As a consequence an erroneous database would lower the bound for a frequency estimate of a particular mtDNA haplotype. Still, the data may provide valuable information for phylogeographic or population genetic studies. For forensic applications however, the employment of legally defensible mtDNA haplotypes is indicated, which requires high quality data involving competency testing as a basic need for monitoring the quality of the sequencing results.

At the ISFH conference in San Francisco (1999), the European DNA Profiling (EDNAP) Group expressed their commitment to support the development of a web-based online mtDNA database (EMPOP) holding high quality mtDNA population data from diverse populations. The data which are imported onto the EMPOP database are supplied by international forensic DNA laboratories, which have successfully participated in collaborative exercises (CE). The present work is a review of the EMPOP collaborative exercises on mtDNA typing undertaken by the European DNA Profiling Group, corresponding to the period 2000–2003. The exercises were developed in two main phases: In a pre-phase, ten DNA extracts from unrelated persons were sent to five laboratories in order to develop a strategy for the future testing activities. In the main phase of the exercises, three buccal swabs and two bloodstains were sent to each participating laboratory in order to perform the mtDNA sequencing analyses.

1.1. Basic principles of the EMPOP collaborative exercises (CE)

The basic principles of the EMPOP CE were modelled on existing systems of forensic quality control [6-8] and attempted to evaluate the following areas.

1. The generation of concordant mtDNA sequence results among different forensic DNA laboratories using their individual techniques and instrumentation.
2. The requirements to successfully establish a common ground of mtDNA sequence analysis for harmonization of data interpretation and nomenclature.
3. The introduction of an IT-based platform for logistical and organisational requirements for secure data transfer and storage.

On this account, the EMPOP CEs were designed as a tool for quality control, which a laboratory engaged in providing population data for EMPOP should complete successfully. The following issues were addressed.

1. Did the laboratory test the correct sample and were the safety precautions within the laboratory appropriate to avoid confusion or contamination of samples?
2. Were the results correct?
3. Did the laboratory interpret the results correctly?

2. Materials and methods

In a pre-phase of the EMPOP CE, buccal swabs obtained from ten persons were extracted at the Institute of Legal Medicine, University of Innsbruck, Austria, using Chelex
The DNA-extracts were sent to five EDNAP laboratories that participated in the pre-phase of the CE. The aim of the pre-phase was to address organisational questions concerning the set up of further exercises, such as the number and kind of samples submitted to further participating laboratories, mode of exchange of samples, data and results, and assessment of the quality of the results.

For the main phase of the EMPOP CE it was decided that each participating laboratory should receive a set of five samples of mixed biological sources (e.g. three bloodstains and two buccal swabs) from known and previously tested individuals. The samples were prepared according to the recommendations laid down by the ISFG to include a maximum of integrity of the samples and a minimum of contamination [10]. Three bloodstain samples and two buccal swab samples were obtained from volunteers. Blood was taken in sterile citrate containers and 200 μl of whole blood was dispensed onto absorbent paper (FTA; Fitzco Inc., Minnesota, USA) using a calibrated pipette. Bloodstains were prepared in such a way that there was sufficient blank paper for negative blank paper controls. Saliva was collected with sterile buccal swab devices (Omni Swab, Fitzco Inc.) and put in sterile 1.5 ml reaction tubes filled with absolute iso-propanol [11]. The laboratories were asked to type at least one of the hypervariable regions HV1/HV2/HV3 of the mtDNA control region in accordance with the protocol and strategies used in each individual laboratory.

Participating laboratories were assigned a unique identifier in chronological order of dispatch in order to avoid duplication of exercise samples for laboratories, which participated more than once. They were requested to provide detailed information on the processing of the samples by means of a questionnaire with all the technical details related to the analyses. It was defined that the international recommendations for mtDNA typing [12–15] were to be used to describe the results obtained. In addition to a result summary (tabular results), the participating laboratories were also asked to attach electropherograms in electronic form for comparison of sequence raw data submitted for the EMPOP CE, base calls of forward and reverse sequences were trimmed 5’ to nucleotide position (np) 16024, 3’ to np 16365 for HV1, 5’ to np 73 and 3’ to np 340 for HVII, respectively. Mean PHRED values were computed for the sequenced regions. In case of length heteroplasmy (polyC stretches), sequences were trimmed after the C-stretch, and the average PHRED value for this hypervariable fragment was computed as a mean over the PHRED values of the two sequence fragments covering the region of interest.

3. Results

The number of participants in the EMPOP collaborative exercises, which returned results by 1 May 2003 stands at 21 laboratories from 11 European countries, one north-American and one south-American country.

3.1. Primers and PCR strategy

Analyses of samples were done in accordance with the protocol and strategies used in each individual laboratory. Primers, polymerases and PCR strategies are summarised in Table 1. Most frequently used primer combinations for the amplification of HV1 were L15971/L15997 and H16401/H16410, and L15/L29 and H408/H484 for HV2, respectively. AmpliTaq Gold (AB, Applied Biosystems, Foster City, CA, USA) and Taq DNA Polymerase (Promega Corporation, Madison, WI, USA) were the most frequently used polymerases in the exercises. Concerning the PCR protocol, 30 up to 62 PCR cycles were applied, the majority of the laboratories performed direct PCR. For the purification of PCR-products before sequencing, most laboratories used Microcon 100 (Millipore, Billerica, MA, USA), QIAquick (Qiagen, Hilden, Germany), ExoSap-IT (USB Corporation, Cleveland, OH, USA), Microspin S300hr (Amersham, Buckinghamshire, UK) or Centricon 100 (Millipore).

Dye terminator cycle sequencing was applied by the majority of the participating laboratories. In general, each template was sequenced in both forward and reverse directions. Sequencers included gel electrophoresis (ABI377, 50% of the laboratories) and capillary electrophoresis (ABI310, ABI3100 and ABI3700). Most of the laboratories used the same primers as were used for the amplification, in some cases with universal tails. Approximately 70% of the laboratories used BigDye chemistry (AB) for cycle sequencing, the remaining labs applied dye primer cycle sequencing (AB), dichloro-rodhamines (AB), DYEnamicETTerminators (Amersham), or sequence (AB).
<table>
<thead>
<tr>
<th>Lab</th>
<th>HV1 primer</th>
<th>HV2 primer</th>
<th>Polymerase</th>
<th>Cycle</th>
<th>PCR strategy</th>
<th>Post-PCR purification</th>
<th>Sequencing chemistry</th>
<th>Electrophoresis/detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L-15971/H-16410</td>
<td>L-15/H-484</td>
<td>AmpliTAQ Gold (AB)</td>
<td>35</td>
<td>Direct PCR</td>
<td>Exo-SapIT (USB)</td>
<td>BigDye term (AB)</td>
<td>ABI3700</td>
</tr>
<tr>
<td>2</td>
<td>L-15997/H-16401</td>
<td>L-29/H-408</td>
<td>AmpliTAQ Gold (AB)</td>
<td>35 or 2*25</td>
<td>Direct and nested PCR</td>
<td>No comment</td>
<td>BigDye term, sequenase (AB)</td>
<td>ABI377, ABI3100</td>
</tr>
<tr>
<td>3</td>
<td>L-15997/H-16401</td>
<td>L-48/H-408</td>
<td>AmpliTAQ Gold (AB)</td>
<td>30 + 15</td>
<td>Nested PCR</td>
<td>Microspin S300hr (Amersham)</td>
<td>BigDye term (AB)</td>
<td>ABI310</td>
</tr>
<tr>
<td>4</td>
<td>L-15997/H-16401</td>
<td>L-48/H-408</td>
<td>Taq DNA Polymerase (Promega)</td>
<td>36</td>
<td>Direct PCR</td>
<td>Microspin S300hr (Amersham)</td>
<td>No comment</td>
<td>ABI3100</td>
</tr>
<tr>
<td>5</td>
<td>L-15997/H-16401</td>
<td>L-29/H-408</td>
<td>Taq Supreme (Helena Biosciences)</td>
<td>30</td>
<td>Direct PCR</td>
<td>Concert rapid PC (Invitrogen) Microspin 300 (Pharmacia)</td>
<td>BigDye term (AB)</td>
<td>ABI310</td>
</tr>
<tr>
<td>6</td>
<td>L-15997/H-16401</td>
<td>L-29/H-408</td>
<td>TAQ Polymerase (AB)</td>
<td>35</td>
<td>Direct PCR</td>
<td>QIAquick (Qiagen) Dye primer (AB)</td>
<td>ABI377</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>L-15971/H-16410</td>
<td>L-15/H-389</td>
<td>AmpliTAQ Gold (AB)</td>
<td>36</td>
<td>Direct PCR</td>
<td>Centricon 100 (Millipore) BigDye term (AB)</td>
<td>ABI377</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>L-15971/H-16410</td>
<td>L-15/H-381</td>
<td>TAQ Polymerase (AB)</td>
<td>32</td>
<td>Direct PCR</td>
<td>Concert (Gibco) BigDye term (AB)</td>
<td>ABI377</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>L-15971/H-408</td>
<td>L-15997/H-408</td>
<td>Pfu Polymerase (Promega)</td>
<td>32</td>
<td>Direct PCR</td>
<td>Microcon 100 (Millipore) BigDye term (AB)</td>
<td>ABI3100</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>L-15997/H-16239</td>
<td>L-29/H-255</td>
<td>AmpliTAQ Gold (AB)</td>
<td>30</td>
<td>Direct PCR</td>
<td>Overlapping fragments Microcon 100 (Millipore) BigDye term (AB)</td>
<td>ABI377</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>L-15995/H-16400</td>
<td>L-31/H-410</td>
<td>AmpliTAQ Gold (AB)</td>
<td>43</td>
<td>Direct PCR</td>
<td>Microcon 100 (Millipore) Dye primer (AB)</td>
<td>ABI377</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>L-15995*/H-16239**</td>
<td>L-182*/H-619**</td>
<td>TAQ Polymerase (AB)</td>
<td>30</td>
<td>Overlapping fragments QIAquick (Qiagen) BigDye term (AB)</td>
<td>ABI3100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>L-15971/H-16410</td>
<td>L-15/H-484</td>
<td>AmpliTAQ Gold (AB)</td>
<td>31</td>
<td>Direct PCR</td>
<td>QIAquick (Qiagen) BigDye term (AB)</td>
<td>ABI310</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>L-15997/H-16401</td>
<td>L-29/H-408</td>
<td>Taq DNA Polymerase (Promega)</td>
<td>32</td>
<td>Direct PCR</td>
<td>Microcon 100 (Millipore) BigDye term (AB)</td>
<td>ABI3100</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>L-15971/H-16410</td>
<td>L-15/H-484</td>
<td>AmpliTAQ Gold (AB)</td>
<td>30</td>
<td>Direct PCR</td>
<td>QIAquick (Qiagen) Dichloro rhodamines (AB)</td>
<td>ABI377</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Sample Code</td>
<td>Reagent/Enzyme</td>
<td>Cycles</td>
<td>Method</td>
<td>Column/Centricon</td>
<td>Additives/Primer</td>
<td>Machine</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>--------</td>
<td>-------------------------</td>
<td>------------------</td>
<td>-----------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>L-15997/H-16395</td>
<td>L-48/H-408</td>
<td>AmpliTAQ Gold (AB)</td>
<td>34</td>
<td>Direct PCR</td>
<td>Centricon 100 (Millipore)</td>
<td>Dichloro rhodamines (AB)</td>
<td>ABI377</td>
</tr>
<tr>
<td>19</td>
<td>L-15877/H-16432</td>
<td>L-27/H-469</td>
<td>Mastermix (Promega)</td>
<td>36</td>
<td>Direct PCR</td>
<td>Microcon 100 (Millipore)</td>
<td>BigDye term (AB)</td>
<td>ABI3100</td>
</tr>
<tr>
<td>20</td>
<td>L-15926/H-580</td>
<td>L-15926/H-580</td>
<td>Taq DNA Polymerase (Promega)</td>
<td>30 + 32</td>
<td>Nested PCR</td>
<td>Microcon 100 (Millipore)</td>
<td>BigDye primer (AB)</td>
<td>ABI377</td>
</tr>
<tr>
<td>21</td>
<td>L-15997/H-16401</td>
<td>L-47/H-408</td>
<td>AmpliTAQ Gold (AB)</td>
<td>36</td>
<td>Direct PCR</td>
<td>Exo-SapIT (USB)</td>
<td>DYEnamic ET (Amersham)</td>
<td>ABI3100</td>
</tr>
<tr>
<td>22</td>
<td>L-15997/H-16427</td>
<td>Not typed</td>
<td>TAQ Polymerase (Perkin-Elmer)</td>
<td>35</td>
<td>Direct PCR</td>
<td>Microcon 100 (Millipore)</td>
<td>DYEnamic ET (Amersham)</td>
<td>ABI377</td>
</tr>
<tr>
<td>25</td>
<td>L-15977/H-16410</td>
<td>L-15/H-484</td>
<td>Taq DNA Polymerase (Promega)</td>
<td>32</td>
<td>Direct PCR</td>
<td>QIAquick (Qiagen)</td>
<td>BigDye term, sequenase (AB)</td>
<td>ABI310</td>
</tr>
</tbody>
</table>

Asterisk marks primers with the universal forward tail (–21)M13; two asterisks mark primers with the universal reverse tail. For nested PCR, the numbers of cycles of both first and second PCR runs are given.
Analyses of the sequences were performed using the commercial software packages Sequence Navigator (AB, 60% of the laboratories), Sequencher (GeneCodes Corporation, 20%), Sequence Editor (AB, 20%) and SeqScape (AB, one lab).

3.2. Sequence quality score values

For 11 randomly selected laboratories quality values (QVs) were calculated for the raw sequencing data. We selected samples which were analysed by all of these laboratories (n = 5), and computed the mean quality values from DNA sequence stretches covering HV1 (16024–16365) and HV2 (73–340), both forward and reverse direction for each sample from each laboratory. Thus, one bar in Fig. 1 corresponds to a mean PHRED value of a sample sequenced in HV1 and HV2 by both forward and reverse reactions. Generally, based on our experience, base calls with quality values higher than 20 (corresponding to a maximum error probability of 1%) can be considered high sequence quality for the interpretation of the predominant sequence. For the data submitted by the 11 laboratories quality scores in the range of 40 were observed with only few exceptions (between 29 and 40). We observed an association between the heights of the quality score values and the sequencing platforms used. Interestingly, quality values were higher for sequences obtained by gel electrophoresis instruments than those obtained by capillary analysers, even if the sequence backgrounds of the gel electrophoresis sequences were obviously higher. Our interpretation of this phenomenon is that the base spacing of gel electrophoresis instruments generates broader peaks than the base spacing of capillary instruments, and the “better” base spacing masks the fact that the sequence itself shows a higher sequence background. Fig. 2 demonstrates how the quality score values differed with respect to the levels of background noise using identical primers, sequencing chemistry, and electrophoresis instrumentation but different samples (various levels of length heteroplasmy 5′ to the region shown).

3.3. Evaluation of the exercise results and classification of errors

In the course of evaluation of the exercise results, both the listed data denoted as different to the CRS sequence as well as the raw data of each participating laboratory were inspected. With respect to the results reported in the tables, 14 disaccording haplotypes (16 individual errors) were observed out of a total of 150 submitted samples/haplotypes (approximately 150,000 bp). Deviations from the expected mtDNA results were classified into categories depending on the sources of errors. Table 2 lists a summary of the results...
Fig. 2. Selected electropherograms (AC—repeat in HV3) with PHRED quality score values. PHRED quality score values are shown above each base call.
and the discrepancies observed. Classes were defined as follows:

- Mistakes in the course of transcription of the results, i.e. clerical errors (Type I).
- Sample mix-up, “artificial recombinants” (Type II).
- Contamination (Type III).
- mtDNA nomenclature (Type IV).

3.3.1. Clerical errors (Type I)

Overall, 10 discrepancies between the expected values and the individual reported data (tabular results) were observed. In seven instances, a sequence variant was not reported in the results table, although it was called correctly by the analysis software. Positions involved were 16093C, 16189C, 16356C, 16362C, 16399G and 263G. In one instance, the original CRS base was reported instead of the variant base (positions 16311). In one case an “artifac-

tual transversion” (16358A) was introduced. Another instance involved a single base shift in the results table, when 16310C was reported instead of 16311C.

3.3.2. Sample mix-up, “artificial recombinants” (Type II)

In two cases mix-up of results with respect to the haplotypes obtained for the two hypervariable regions was observed. One laboratory reported a correct HV1 haplotype,
whereas the HV2-haplotype was not in agreement with the exercise sample in question. This was also supported by the raw data for this individual sample. None of the other four samples included in the batch had a corresponding HV2-haplotype; therefore, this haplotype must have been introduced by an unknown source. Finally, a second independent analysis of the same sample led to the expected HV2-result. In another case, a laboratory reported a correct HV1–HV2-haplotype, but the HV3-haplotype did not agree with the exercise sample in question. In this case, the HV3 region corresponded to the HV3 region of another collaborative exercise sample. Again, re-analysis of a new sample gave the correct result.

3.4. Interpreting heteroplasmy

3.4.1. Length heteroplasmy

Samples 1 and 5 included length heteroplasmy in HV1 (around position 16189) while samples 9, 10 and 13 showed length heteroplasmy in HV2 (around position 310, Table 2). A total of 20 out of 21 laboratories returned the questionnaire. Thirteen laboratories stated that they detected length heteroplasmy in the CE samples, four laboratories did not consider length heteroplasmy, and three laboratories did not specify on this subject. One of the thirteen laboratories, which claimed to identify length heteroplasmy, reported a result which was not in accordance with the international nomenclature [15,18]. The other 12 laboratories reported consistent results concerning the polyC stretches.

3.4.2. Point heteroplasmy

The samples were sent in batches in a way that every laboratory received at least one sample displaying point heteroplasmy. Sample 1 showed a mixture at np 234 (A ~ G) and sample 2 showed a mixture at np 195 (C ~ T). The latter was detected by all five laboratories that received that sample. The A ~ G point heteroplasmy at np 234 in sample 1 was described by 11 out of 21 labs. Three laboratories did not specify in the questionnaire concerning the question whether they would consider point heteroplasmy. Ten laboratories indicated the detection of point heteroplasmy, two laboratories only typed HV1 (missing the heteroplasmic position at np 234), and the remaining five laboratories declared that they did not consider point heteroplasmy. Examples for sequence alignments showing position 234R are given in Fig. 3. All laboratories, which stated that they would consider point heteroplasmy were able to correctly identify it.

4. Discussion

Collaborative exercises are a valuable means for evaluating and monitoring the quality of laboratory work as well as the interpretation and the transfer of results. In the EMPOP collaborative exercises the mtDNA sequencing results were evaluated by inspecting both the raw data and the tabular results submitted by each participating laboratory. Consequently, the results of the CE needed to be considered individually, as for example the raw data of the analysis were correct while the interpretation or transcription of the result were erroneous. A “simple” clerical error in a table does therefore not necessarily invalidate the laboratories’ ability to generate correct mtDNA profiles. The raw data serve as a quality measure of the sequence data, reflecting the entire laboratory process of a particular sample. The reported results (summary, table) monitor the quality of data interpretation and transcription, and are crucial for reporting evidence.

Clerical errors have shown to be the main source of all discrepancies observed in the CEs (10 out of 16 total errors, 62.5%). As outlined above, the raw data of these samples were in perfect agreement with the expected results, and the errors were instantly corrected by the laboratories when they were contacted for revision of the sample in question. The observation of clerical errors is not restricted to the analysis.
of mtDNA polymorphisms alone, but can be observed in any situation where data are transcribed manually. Clerical errors are likely to occur more frequently when the necessary control mechanisms are not in place. Such control mechanisms involve double reading of results, completely repeated analysis of a sample if appropriate or independent analysis of the sample in another laboratory for confirmation.

In routine casework analysis clerical errors are not expected to appear to such an extent because reference samples of the victim and/or the suspect(s) are usually analysed as well. Therefore, the triangle stain/victim/suspect serves as final control instance for clerical errors to be detected and corrected. If at all, it is much more likely that (clerical) errors produce false exclusion scenarios, which would be in favour of the suspect. A false match due to an error is very unlikely.

Collaborative exercises, proficiency tests and high sample-number experiments such as population studies however, are more prone to clerical errors. First, samples are typed anonymously without any correlation or possibility for
double-check, which might help discover clerical errors. Second, as a consequence of the high amount of samples to be analysed in population studies, the chance of introducing clerical errors raises due to the increased complexity of the data set.

In two instances, mix-up of the sequenced regions of two samples was observed (artificial recombinants): this phenomenon, which has also been discussed elsewhere [1,3], caused 12.5% of all observed errors. While this kind of error can easily be identified in the CE, where the result of the samples is known, it is more difficult to reveal this kind of errors in a set of population data. It has been described that phylogenetic methods as prefigured by [1] help to identify the erroneous haplotypes, although the methods therein were more focussed on the identification of phantom mutations. Further studies are on their way, in order to evaluate the sensitivity of these methods.

4.1. Sequence quality

The quality score values of the sequence data of 11 randomly selected laboratories were evaluated by means of qualitative sequence analysis software (PHRED), which assigns a quality value ranging from 4 to 60 to every individual base call. These values are linked to error probabilities in a way that a PHRED quality score of 10 corresponds to an accuracy of the base call of 90%, and a PHRED quality score of 50 matches a correctness of the base call of 99.999%. Based on our experience, quality scores exceeding the value of 20 can generally be regarded as acceptable sequence quality in forensic context (99% accuracy of the base call), even more so when the sequence information of both strands is available and the base call is confirmed. In an earlier study, we demonstrated that the interpretation of low level heteroplasmy requires a lower sequence background, which is usually associated with PHRED values exceeding 45 [19].

The algorithms involved in this calculation using PHRED mainly take two parameters into consideration, the signal-to-noise ratio at a given peak position and the base spacing of the sequence electropherogram. Individual properties of the instrumentation (gel or capillary-based) or sequencing chemistry are not considered for the PHRED score value. As a consequence, sequence data derived from different electrophoresis platforms and/or sequencing chemistries can only be compared with caution. Excellent base spacing may mask moderate levels of background noise, the latter being a more important value for the interpretation of point heteroplasmy. Sequence analysis programs calculating quality score values were originally designed for rapid high-volume sequence analysis in the context of high through-put genome sequencing [20,21]. Nevertheless, its application to the analysis mtDNA sequences in the forensic field may provide useful information regarding the objectiveness and the reliability of the individual base calls, and can serve as additional support to the assessment of high quality data.

4.2. Conclusions and impact on forensic casework analysis

The results of the EMPOP collaborative exercises may bear consequences with respect to the analysis of mtDNA in the forensic context. Although the detected sources of errors may not be of such importance in the field of mtDNA casework analysis, they would certainly impact the establishment of forensic mtDNA population databases for frequency estimations.

It is important to state that the error rate that has been delineated from the collaborative exercises cannot be extrapolated to a universally valid error rate for forensic mtDNA casework analysis. In a routine mtDNA casework example, the forensic laboratory usually types the mtDNA profiles of the stain material, the suspect and the victim. The interpretation in the report requires a direct comparison of the obtained sequences, which facilitates the identification of potential errors. However, this safety net does not relieve the laboratory from its responsibility to apply the necessary and internationally demanded recommendations on good mtDNA laboratory practice [12–15,18], including complete analysis and correct interpretation of control samples, double-strand sequencing, independent double analysis/interpretation of the results by experienced scientists and retention of a portion of the casework sample for potential confirmation by an independent laboratory. As a worst case scenario a false exclusion of the suspect due to erroneous mtDNA typing can be expected, whereas a false inclusion would certainly be more unlikely.

The establishment of mtDNA population databases has not yet been subject to such detailed control mechanisms. It is very likely that population database compilations have not undergone the same procedure of validation and confirmation by direct comparison as casework profiles. It is evident that mtDNA databases require additional control mechanisms.

- Based on the results obtained in the course of the collaborative exercises, the avoidance of any manual transcription at any stage of data handling is a requirement to prevent clerical errors.
- Further, the double analysis of the results in another laboratory using the standard nomenclature [15,18] helps reveal interpretation errors and phantom mutations.
- The retention of the raw data serves as a prerequisite for the ability to trace back the origin of the results, whose authenticity may be inquired at a later stage.
- Still, artificial recombinants remain errors that would not be detected by the above mentioned control mechanisms. Alternatively, phylogenetic analyses will help to identify these errors taking advantage of the known patterns of mutations with respect to the matrilinear origin of the sample in question. Therefore, it may become essential to obtain bio- and phylogeographic information upfront on the population samples prior to the sequence analysis.
allowing for an independent analysis if necessary, a basic
would be the retention of a portion of the original sample,
—

A very useful precaution to demonstrate quality for
mtDNA databasing—similar to the casework scenario—
would be the retention of a portion of the original sample,
allowing for an independent analysis if necessary, a basic
forensic principle.

References