GHEP-ISFG proficiency test 2011: Paper challenge on evaluation of mitochondrial DNA results

L. Prieto [a,*], C. Alves [b], B. Zimmermann [c], A. Tagliabracci [d], V. Prieto [e], M. Montesino [a], M.R. Whittle [f], M.J. Anjos [g], S. Cardoso [h,i], B. Heinrichs [j], A. Hernandez [k], A.M. López-Parra [l], A. Sala [m], V.G. Saragoni [n], G. Burgos [o], M. Marino [p], M. Paredes [q], C.A. Mora-Torres [r], R. Angulo [s], G. Chemale [t], C. Vullo [u], M. Sánchez-Simón [v], D. Comas [w], J. Puente [x], C.M. López-Cubría [y], N. Modesti [z], M. Aler [A], S. Merigioli [B], E. Betancor [C], S. Pedrosa [D], G. Plaza [E], M.V. Masciovecchio [F], P.M. Schneider [G], W. Parson [c,**]

[a] Comisaría General de Policía Científica, University Institute of Research in Forensic Sciences (IUICP), Madrid, Spain
[b] Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal
[c] Institute of Legal Medicine, Innsbruck Medical University, Austria
[d] Dipartimento Scienze Biomediche e Sanità Pubblica, Medicina Legale, Università Politecnica delle Marche, Ancona, Italy
[e] National Institute of Toxicology and Forensic Sciences (INTCF), Sevilla, Spain
[f] Genomic Engenharia Molecular, Sao Paulo, Brazil
[g] National Institute of Legal Medicine, Centre Branch, Coimbra, Portugal
[h] Servicio General de Investigación Genómica: Banco de ADN (SGIker), UPV/EHU, Vitoria-Gasteiz, Spain
[i] BIOMICs Research Group, Centro de Investigación "Lucio Lascaray", University of the Basque Country, Vitoria-Gasteiz, Spain
[j] National Institute of Toxicology and Forensic Sciences (INTCF), Madrid, Spain
[k] National Institute of Toxicology and Forensic Sciences (INTCF), Canarias, Spain
[l] Lab. Population and Forensic Genetics, Dep. Toxicology and Health Legislation, Faculty of Medicine, Complutense University, Madrid, Spain
[m] Servicio de Huellas Digitales Genéticas, University of Buenos Aires, Argentina
[n] Forensic Genetics Unit, Legal Medicine Service, Santiago, Chile
[o] Molecular Genetics Laboratory, Red Cross, Ecuador
[p] Laboratorio de Análisis de ADN, School of Medical Sciences, National University of Cuyo, Mendoza, Argentina
[q] National Institute of Toxicology and Forensic Sciences (INTCF), Barcelona, Spain
[r] Instituto Nacional de Medicina Legal y Ciencias Forenses, Bogotá, Colombia
[s] Forensic Sciences Department, Poder Judicial, Costa Rica
[t] Forensic Genetics Laboratory, National Institute of Criminalistics, Federal Police, Brazil
[u] LIDMO-EAAF, Equipo Argentino de Antropología Forense, Córdoba, Argentina
[v] Centro de Análisis Genéticos, Citogen S.L., Zaragoza, Spain
[w] Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain
[x] LabGenetics SL., Madrid, Spain
[y] Servicio de Criminalística, Dpto. de Biología, Guardia Civil, Spain
[z] Instituto de Genética Forense, Poder Judicial de Córdoba, Argentina
[A] Institute of Legal Medicine of Valencia, Spain
[B] Laboratory of Forensic Genetics and Legal Medicine, Universitá Cattolica del Sacro Cuore, Roma, Italy
[C] Forensic Genetics Department, University of Las Palmas de Gran Canaria, Spain
[D] NASERSA SA, Navarra, Spain
[E] NEODIAGNOSTICA SL., Lleida, Spain
[F] IACA Laboratories, Bahía Blanca, Argentina
[G] Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Germany

ARTICLE INFO

ABSTRACT

The GHEP-ISFG Working Group performed a collaborative exercise to monitor the current practice of mitochondrial (mt)DNA reporting. The participating laboratories were invited to evaluate a hypothetical case example and assess the statistical significance of a match between the haplotypes of a case (hair) sample and a suspect. A total of 31 forensic laboratories participated of which all but one used the EMPOP database. Nevertheless, we observed a tenfold range of reported LR values (32–333.4), which was mainly

* Corresponding author.
** Corresponding author. Tel.: +43 512 9003 70640; fax: +43 512 9003 73640.
E-mail addresses: lourditasmt@gmail.com (L. Prieto), walther.parson@i-med.ac.at (W. Parson).

due to the selection of different reference datasets in EMPOP but also due to different applied formulae. The results suggest the need for more standardization as well as additional research to harmonize the reporting of mtDNA evidence.

1. Introduction

The interpretation of matching DNA profiles involves a statistical evaluation to assess the strength of the evidence. Likelihood ratios (LRs) or Random Match Probabilities (RMPs) are two of some possible ways of expression [1,2]. There are different procedures with nuclear and mitochondrial markers based on their genetic properties. The variation of autosomal Short Tandem Repeat loci (aSTRs) can be mainly attributed to slippage mutation events and recombination albeit to a much smaller extent. The distribution of allele frequencies has traditionally been estimated on the basis of relatively small sizes of randomly sampled individuals (e.g. few hundreds), as the number of observed alleles for an STR locus is usually not very large. Further, Mendelian inheritance allows for the assumption that populations are more or less in equilibrium (that is, the allelic frequencies are roughly maintained throughout different generations). Thus, generally accepted criteria have been established for estimating allele and genotype frequencies in aSTRs [3].

On the contrary, the level of polymorphism in mtDNA is restricted to single mutational base exchanges with an apparent lack of recombination. The spatial distribution of contemporary mtDNA haplotypes is based on the history of the radiating lineages that left their footprints during human migration. Numerous population studies established and confirmed that the number of unique (novel) mtDNA haplotypes is usually well exceeding 50% of the total sample size and is only reduced in populations with restricted gene flow, genetic bottleneck and/or strong founder effects with further limited genetic exchange (maternal). The rarity of an mtDNA profile is traditionally estimated by querying datasets (databases) of randomly sampled individuals. The high number of unobserved "alleles" requires much larger sample sizes to be established compared to aSTRs. In fact, the estimates of haplotype frequencies when using small mtDNA population databases are often biased, and even in larger databases, the estimation of rare haplotypes poses a major challenge [4–6].

Between 1990 and 2000, numerous individual laboratories started to establish local mtDNA datasets for frequency estimations that were particularly small and – as found out later – often fraught with error. The Europe DNA Profiling Group (EDNAP, www.ednap.org) initiated collaborative exercises on mtDNA typing to understand sources and reasons for error [7]. As a result, human clerical error and phantom mutations (artificial signals due to suboptimal experimental conditions) were found to be responsible for the majority of problems in the generated data. As a consequence, improved experimental protocols and specific guidelines were elaborated to improve the quality of mtDNA datasets [8–12]. Sequence data generated under these and similar protocols form the *forensic* data source in EMPOP, an internet-based resource presenting mtDNA haplotypes from all over the world (www.empop.org; [13,14]). By means of continued collaborative initiatives, the total sample size meanwhile exceeded 16,000 (Release 6, March 2012), resulting in improved estimation of haplotype frequencies [e.g. 15–17].

The interpretation and statistical evaluation of mtDNA typing results is not always straight forward. It is not uncommon for differences to be observed between haplotypes of maternal relatives [18] or even between tissues from the same individual

[19]. In addition, the distribution of haplotypes in worldwide populations can differ significantly based on the region-specific migration history [20,21]. The allelic frequency distribution of aSTR markers is known to differ between populations, however, the effect on the final LR is not as pronounced as with mtDNA haplotypes, where it is more crucial to select the relevant population database [21]. Also, correction for subpopulation effects is more difficult with mtDNA and may lead to higher correction factors than for aSTRs. Other issues that are typically encountered when analyzing mtDNA are the edition range of the query haplotype, inclusion/exclusion of point and/or length heteroplasmies, the selection of the appropriate reference population and the applied methodology to correct for sampling bias.

Historically, the Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) has shown major interest in the use and standardization of mtDNA analysis for forensic purposes since many years [22–27]. So far, a lot of information has been made available on protocols and different mtDNA analysis strategies that forensic laboratories use. Little however is documented about how member laboratories report mtDNA results. In order to obtain an overview of the communication of mtDNA evidence, the GHEP-ISFG proposed an exercise regarding the interpretation of mtDNA results in its last proficiency test performed in 2011.

2. Hypothetical case example

Participating laboratories were asked to evaluate a hypothetical forensic case (homicide in Barcelona) in which the mtDNA haplotype of a hair shaft found in the hand of the victim matched the suspect's haplotype (see Table 1). The laboratories were also asked to consider the evidence under the following two hypotheses:

(a) Prosecutor: H_p: the hair shaft originates from the suspect or from a maternal relative of the suspect.
(b) Defense: H_d: the hair shaft originates from an unknown individual of the European population not related to the suspect.

Two additional questions had to be answered by the laboratories in order to better evaluate their results: (i) which mtDNA database they have used and (ii) the number of matches that were found in that database.

A total 31 laboratories participated in this GHEP-ISFG exercise (not including the laboratory managing the EMPOP database in Innsbruck).

Table 1
MtDNA haplotypes from a hypothetical case example. The hair sample was recovered from a hand of the victim and matched the suspect's haplotype.

Sample	Edition range	Haplotype
Victim	16024–576	263G 315.1C 16519C
Suspect	16024–576	72C 263G 315.1C 16298C
Hair shaft	16024–576	72C 263G 315.1C 16298C

Table 2
Overview of database query settings, search results and LR values provided by 31 participating laboratories.

Lab	Database, population	Data source	Match type	Matches (x/N)	Frequency calculation	LR
1	EMPOP, Europe	F	Pattern	8/2636[a]	$x+1/N+1$	1/freq = 293.0
2	EMPOP, all	F + L	Pattern	35/8410	x/N	1/freq = 240.3[b]
3	EMPOP, Europe	F + L	Pattern	26/2667	$x+2/N+2$	1/freq = 95.3[c]
4	EMPOP, Europe	F + L	Pattern	26/2667	x/N	1/freq = 102.6[c]
5	EMPOP, Europe	F + L	Pattern	26/2667	$x+2/N+2$	1/freq = 95.3[c]
6	EMPOP, all	F + L	Pattern	35/7973	x/N	1/freq = 227.8[b]
7	EMPOP, all	F + L	Pattern	43/8558	$x+1/N+1$	1/freq = 194.5
8	EMPOP, all	F + L	Pattern	35/8410	$x+1/N+1$	1/freq = 233.6[b]
9	EMPOP, Europe	F	Pattern	26/2636	$x+2/N+2$	1/freq = 94.2[c]
	EMPOP, all	F	Pattern	35/7015	$x+2/N+2$	1/freq = 189.6
10	EMPOP, Europe	F + L	Pattern	26/2667	x/N	1/freq = 102.6[c]
11	EMPOP, Europe	F + L	Literal	24/2667	$x+1/N+1$	1/freq = 106.7[c]
12	EMPOP, all	F	Pattern	35/7015	x/N	1/freq = 200.4[b]
13	EMPOP, all	F + L	Pattern	35/8410	x/N	1/freq = 240.3[b]
14	EMPOP, Europe	F	Pattern	26/2636	$x+1/N+1$	1/freq = 97.7[c]
15	EMPOP, Spain	F + L	Pattern	0/31	$x+1/N+1$	1/freq = 32.0
16	EMPOP, Europe	F	Pattern	26/2636	$x+1/N+1$	1/freq = 97.7[c]
17	EMPOP, Europe	F + L	Pattern	8/2667[a]	x/N	1/freq = 333.4
18	EMPOP, S-Europe	F	Pattern	6/742	x/N	1/freq = 123.7[c]
19	EMPOP, all	F + L	Pattern	35/8410	$x+2/N+2$	1/freq = 227.3[b]
20	EMPOP, Europe	F + L	Pattern	26/2667	$x+2/N+2$	1/freq = 95.3[c]
21	EMPOP, all	F + L	Pattern	35/8410	x/N	1/freq = 240.3[b]
22	EMPOP, all	F + L	Pattern	35/8410	$x+1/N+1$	1/freq = 233.6[b]
23	EMPOP, Europe	F + L	Pattern	26/2667	$x+1/N+1$ and Wilson interval	1/freq = 68.0
24	EMPOP, Europe	F + L	Pattern	26/2667	x/N	1/freq = 102.6[c]
25	OWN (Chile)	n.a.	Pattern	2/1040	$(x/N)+1.96\sqrt{(pq/N)}$	1/freq = 218.0[b]
26	EMPOP, all	F + L	Pattern	35/8410	$x+1/N+1$	$e^{-g\mu}$/freq = 233.6[b]
27	EMPOP, Europe	F + L	Pattern	26/2667	x/N	1/freq = 102.6[c]
28	EMPOP, Europe	F + L	Literal	24/2667	$x+1/N+1$	1/freq = 106.7[c]
29	EMPOP, Europe	F	Pattern	26/2636	$x+1/N+1$	1/freq = 97.7[c]
30	EMPOP, Europe	F + L	Pattern	26/2667	$x+1/N+1$	1/freq = 98.8[c]
31	EMPOP, Spain	F + L	Pattern	0/31	$x+1/N+1$	1/freq = 32.0

F = forensic data, L = literature data, details see text.
Match type indicates query settings with respect to heteroplasmy, details see text.
x = matches in the database; N = database size; $p = x/N$; $q = 1 - p$.
g = number of generations (in this case 2 generations were considered, see details in the text); μ = mutation rate per base and per generation (3.6×10^{-6}).

[a] This laboratory performed the search disregarding insertions/deletions in the homopolymeric C-tract around position 309.
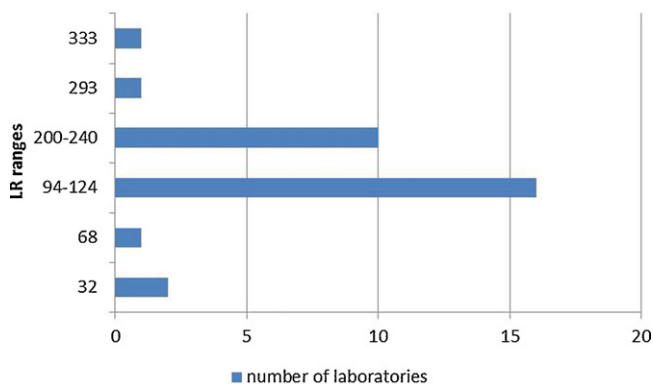[b] LR within 200–241.
[c] LR within 94–124.

3. Results

Almost all participating laboratories (30 of 31) used the EMPOP database to infer the rarity of the given control region mtDNA haplotype (Table 2). The exercise took place in spring 2011 when EMPOP Release 4 was online with a total of 12,785 haplotypes. Of these, 2667 were full control region profiles from Europe (3201 full control region profiles with west Eurasian metapopulation affiliation). Under standard query settings (pattern search, disregarding differences in C-stretches; see below) the mtDNA haplotype in question was observed 26 times in Europe (29 times

Fig. 1. Summary of reported LR results (N = 31 labs).

in "west Eurasia"). The EMPOP database has provided the uncorrected frequency estimates for the queries (9.749e−3 and 9.060e−3, respectively) along with the lower and upper bound 95% confidence intervals (C.I.) ([6.661e−3; 1.425e−2] and [6.315e−3; 1.298e−2], respectively) estimated according to Wilson [28].

The reported LR values by the participating laboratories ranged from 32.0 to 333.4 (Table 2 and Fig. 1). This can in part be explained by the varying selection of the reference dataset including (i) the entire database content (N = 10, LRs between 200 and 241), (ii) the European portion of EMPOP (N = 17, LRs from 94 to 124), (iii) the South European portion of EMPOP (N = 1, LR = 123.7) and (iv) the Spanish population as reference dataset (N = 2, LR = 32). One laboratory (#25) reported a different result based on its own database. Varying LR values within the first two groups (i and ii) are further due to the following:

(a) Applied formulae: Eleven laboratories (35%) reported the uncorrected frequency as the haplotype in question was observed more than once in the datasets (with the exception of those two labs that used the Spanish dataset; Table 2). Thirteen laboratories (42%) added the suspect's haplotype to the database [$(x + 1)/(N + 1)$; x = number of matches in the database, N = size of the database]. Five laboratories (16%) added both haplotypes (evidence and suspect) to the database, [$(x + 2)/(N + 2)$]; one laboratory (#25) used the upper confidence interval (C.I.) as described in [20]: [$p + 1.96\sqrt{(pq/N)}$, $p = x/N$, $q = 1 - p$] with p = matches in the local database, and

one laboratory (#23) used the $(x + 1)/(N + 1)$ correction plus the upper bound of the C.I. following [28].

(b) The combined use of *forensic* and *literature* data or *forensic* data only: the EMPOP database Vs. 2.1 R4 contained two types of mtDNA data regarding its source, *forensic* and *literature*. While the former are supported by high-quality raw sequencing data, the latter are based on lower quality sequence data only. Seven laboratories (23%) selected only the *forensic* data source of EMPOP, while the remaining laboratories queried both data sources (except one laboratory that used a different database).

(c) Point heteroplasmy (PHP): All but one laboratory used *pattern* search mode (standard settings), which regards two sequences differing by only heteroplasmic positions as non-exclusive (e.g. 152Y matches 152C). In contrast the *literal* search mode would result in exclusion for the given example.

(d) Length variants: All but two laboratories (#1 and 17) performed the EMPOP query under standard settings with respect to the treatment of differences at the HVS-2 homopolymeric C-tract, i.e. differences were ignored. Similar to PHP it is common practice not to formulate exclusion on differences there [29,30]. When using the query settings of laboratories #1 and 17 only 8 exact matches (instead of 26) to the haplotype in question were found, as all database haplotypes with 309.1C were excluded from the search. These two laboratories consequently reported lower estimates of the relative haplotype frequencies and therefore higher LRs.

4. Discussion

To the best of our knowledge this is the first report on a collaborative exercise on statistical evaluation of mtDNA results. A total of 31 laboratories participated in the study, where a hypothetical case example was described and the resulting mtDNA haplotypes provided (Tables 1 and 2). Although all but one used the same database (EMPOP Vs. 2.1 R4), we observed considerable variation in reported LR values. We briefly discuss why.

4.1. The selected source dataset

The suspect's haplotype (that matched the one of the hair) belongs to haplogroup HV0 [31]. This haplogroup is typically found in west Eurasia, particularly in Europe and North America. Some HV0 lineages are also present in northern parts of Africa and extend over the Middle East to Central Asia and the Indian subcontinent, albeit at much lower frequencies.

Under the defense hypothesis that the hair shaft originated from an unknown individual of the European population unrelated to the suspect, it would be appropriate to use the European portion of EMPOP (or any other database) as reference dataset. Alternatively, the west Eurasian metapopulation in EMPOP could have also been selected that includes additional "European" lineages sampled outside the borders of contemporary Europe (e.g. US "Caucasians"). One laboratory selected the Southern European dataset, which is another interpretation of the defense hypothesis. While it is not fully in agreement with the H_d it may represent a more cautious way of evaluating the evidence.

The restriction of the query dataset to the Kingdom of Spain may seem logical (at least to some layperson) as the crime happened there, but violates the defense hypothesis, as this would exclude the possibility that the suspect derives from other European populations. We see that the resulting LR is smaller (Table 2), as the haplotype has not been observed in this very small sample of full CR haplotypes that are available for Spain.

The selection of the entire database as reference dataset resulted in the highest LR values but violates the defense

hypothesis and – even more problematic – leads to underestimation of the haplotype's frequency to the disadvantage of the suspect, e.g. by including all East Asian populations that are characterized by a different phylogenetic background and usually do not harbor HV0 lineages.

The variability of the reported answers demonstrates that more research, education, and harmonization is needed to improve consistency in mtDNA interpretation, albeit some flexibility will (need to) remain in response to different possible formulations of the defense hypothesis.

4.2. Applied formulae

Some of the observed variation in reported LR values can be attributed to differences in the applied formulae. Roughly one third of the laboratories used the uncorrected haplotype frequency, since the haplotype was observed more than once (except for those two laboratories that used Spain as reference population). The recommendations of the International Society for Forensic Genetics and from the EDNAP group state the observations in the database should be corrected for sampling bias [29,30], which was state of the art with the smaller mtDNA databases available at that time. Further research is needed to evaluate the requirement of sampling correction with today's database sizes. Other laboratories added the suspect's or the suspects and the crime scene sample's haplotypes to the database, which resulted in increasingly conservative estimates. Apparently there is different understanding and practice among laboratories, which highlights the need for harmonization. Diverse approaches can be found in the literature [e.g. 6,32–35], that do not only differ by their mathematical interpretation but more importantly by the underlying hypotheses.

4.3. LR calculation

As no differences between evidence and suspect haplotypes were present in this exercise, the LR $(p(E|H_p)/p(E|H_d))$ can be calculated as $1/f$ (f = haplotype frequency). Almost all laboratories performed this. One laboratory calculated the LR by using the formula $e^{-g\mu}/f$ [36], where g is the number of generations and μ the mutation rate of the control region. This laboratory considered $g = 2$ since "a maternal relative" was included in the prosecutor's hypothesis. The HVR mutation rate of 3.6×10^{-6} [37] was used. The resulting LR (233.6) however was in the range of the labs which used the same reference dataset (LRs between 200 and 241).

Earlier guidelines stimulated the estimation of mutation rates not only with respect to the number of generations or years, but also the tissue type of an individual [30]. The somatic mutation rate depends on several factors, including the number of cell divisions (and therefore the age of an individual) or the metabolic activity in the tissue [38]. In fact, it was previously described that active tissues such as hair follicles or muscle show a higher rate of mutation than other tissues with little cell division or low energy requirement [39,40]. The body of data however, is not large and consistent enough to allow for the deduction of reliable values for mutation rates. More research is needed here.

4.4. Quality of EMPOP reference data

Regarding the source of reference data for their queries seven laboratories have based their conclusions on EMPOP *forensic* data only. Both *forensic* and *literature* data presented in EMPOP undergo rigorous quality control, and great scrutiny is applied to unveil possible errors (e.g. reference bias, phantom mutation, base misscoring, nomenclature issues, alignment violation, clerical errors, sample mix-up, phylogenetic sense; see [15] as an example). As

Table 3
Summary of haplotype queries under *pattern* and *literal* search modes in EMPOP (v2.1, Release 6; Europe). LRs = 1/*f*.

Haplotype	Pattern matches	Pattern LR	Literal matches	Literal LR	Excluded haplotypes in literal searches
72C 263G 315.1C 16298C	34/2834	83	31/2834	91	72C 204Y 263G 309.1C 315.1C 16126Y 16298C 72C 152Y 263G 315.1C 16298C 72C 263G 309.1C 309.2C 315.1C 16144W 16298C
73G 152Y 263G 315.1C 16126C 16294T 16304C 16519C	10/2834	283	1/2834	2834	73G 263G 315.1C 16126C 16294T 16304C 16519C 73G 152C 263G 309.1C 315.1C 16093Y 16126C 16294T 16304C 16519C

indicated in the description of the EMPOP website the term *forensic* refers to the collection of haplotypes for which high quality sequencing raw data are available, so that questionable positions can be evaluated any time. The term *literature* data refers to haplotypes that were either extracted from reliable publications or sent to EMPOP for review without accompanying raw data. The evaluation of *literature* follows the same stringent procedure of reviewing *forensic* data and authors/submitters have been contacted for raw data of questionable observations. However, raw lane data may not always be available to full extent and quality and therefore *literature* data not harbor the same quality as *forensic* data.

Apparently some forensic users are confused with respect to which data sources to query in a forensic case. The classification *forensic* and *literature* data was established for the first EMPOP Release (2006) when still little was known about the efficiency of detecting errors in data tables. We believe that both data sources are of very high quality – especially with respect to the uncorrected literature – and therefore represent valuable resources for forensic applications. We understand that the classification scheme is confusing and we will perform further research to establish if *forensic* and *literature* data can be pooled in the future to avoid additional complexity for database searches.

4.5. Heteroplasmy

Point heteroplasmy (PHP) is observed in approx. 6% of blood and saliva samples (in the entire control region [41]), and more frequently in hair and metabolically active tissues (e.g. muscle). In the forensic context two sequences are usually not excluded as originating from the same lineage/individual when they differ at PHP positions only (assuming that the homoplasmic variant is included in the PHP). Database queries should reflect this convention, i.e. a haplotype including PHP at 152 (152Y) should not be considered different from the same haplotype with the exception of a 152C (or T). While this can easily be addressed in a database search when the query haplotype shows the PHP (e.g. by querying both variants in separate searches) a user cannot know how many haplotypes in the database carry PHP and which positions are affected. It is therefore important that the query engine of a database executes this convention, which is the case in EMPOP when *pattern* search is used. Under the *literal* search mode such sequences would be considered one-step neighbors as only exact matches are found identical. This function was implemented to enable specific searches for PHP positions and their relative frequencies in EMPOP.

Therefore, pattern searches are always preferred when evidence has to be statistically evaluated. In this exercise the effect of the two modi was only small. In fact, two haplotypes went missing under the *literal* search mode (laboratories 11 and 28): 72C 152Y A263G 315.1C 16298C and 72C 263G 309.1C 309.2C 315.1C 16144W 16298C (24 matches instead of 26). Nevertheless, greater differences can be observed in the case of other haplotypes (see Table 3).

5. Conclusions

The GHEP-ISFG is known as an active working group on forensic mtDNA matters. With this study it has pioneered the field by performing a collaborative exercise on statistical evaluation of mtDNA evidence. The results derived from the interpretation of a hypothetical case example show that albeit using the same database laboratories arrive at discordant statistical interpretation with respect to the reported LR values in a range of one order of magnitude.

The reasons for this can be summarized in few categories: there is an apparent lack of harmonization with respect to the query settings, which can have a considerable effect on the result. Although the usage of EMPOP is described on the website in the help section, the lack of harmonization can be addressed by adapting the database settings, more and easier understandable directions of use and by continued education. Apparently, some participants selected a defense hypothesis that did not adequately consider the circumstances of the case. That became evident by the selection of different reference datasets for the query. This problem is difficult to be solved by the database but lies in the responsibility of the reporting person.

The observed variability does not come as surprise. The field of mtDNA analysis has always been a niche application in forensic science and therefore suffers the evident lack of guidance and education that is necessary to comply with the high demands of stable and comparable procedures in the international forensic community. This study demonstrates the need for further research and database development to strengthen the application in the field.

References

[1] J. Buckleton, A framework for interpreting evidence, in: J. Buckleton, C.M. Triggs, S.J. Walsh (Eds.), Forensic DNA Evidence Interpretation, CRC Press, London, 2005, pp. 27–63.
[2] D.J. Balding, Weight-of-Evidence for Forensic DNA Profiles, Wiley, 2005.
[3] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci. Int. 165 (2007) 238–243.
[4] T. Egeland, H.M. Bøvelstad, G.O. Storvik, A. Salas, Inferring the most likely geographical origin of mtDNA sequence profiles, Ann. Hum. Genet. 68 (Pt. 5) (2004) 461–471.
[5] B. Budowle, J. Ge, X.G. Aranda, J.V. Planz, A.J. Eisenberg, R. Chakraborty, Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence, J. Forensic Sci. 54 (5) (2009) 1016–1021.
[6] C.H. Brenner, Fundamental problem of forensic mathematics – the evidential value of a rare haplotype, Forensic Sci. Int. Genet. 4 (2010) 281–291.
[7] W. Parson, A. Brandstätter, A. Alonso, N. Brandt, B. Brinkmann, A. Carracedo, D. Corach, O. Froment, I. Furach, T. Grzybowski, K. Hedberg, C. Keyser-Tracqui, T. Kupiec, S. Lutz-Bonengel, B. Mevag, R. Ploski, H. Schmitter, P.M. Schneider, D. Syndercombe-Court, E. Sörensen, 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.

[8] 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.

[9] W. Parson, H.-J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (2007) 13–19.

[10] 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.

[11] 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.

[12] L. Fendt, B. Zimmermann, M. Daniaux, W. Parson, Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences, BMC Genomics 10 (2009) 139.

[13] W. Parson, A. Dür, EMPOP – a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88–92.

[14] A. Röck, J.A. 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.

[15] 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.B. 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.

[16] 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.

[17] M. Bodner, J.A. Irwin, M.D. Coble, W. Parson, Inspecting close maternal relatedness: towards better mtDNA population samples in forensic databases, Forensic Sci. Int. Genet. 5 (2011) 138–141.

[18] 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.

[19] 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.

[20] M.M. Holland, T.J. Parsons, Mitochondrial DNA sequence analysis-validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 21–50.

[21] 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.

[22] A. Alonso, A. Salas, C. Albarrán, E. Arroyo, A. Castro, M. Crespillo, A.M. di Lonardo, M.V. Lareu, C.L. Cubría, M.L. Soto, J.A. Lorente, M.M. Semper, A. Palacio, M. Paredes, L. Pereira, A.P. Lezaun, J.P. Brito, A. Sala, M.C. Vide, M.R. Whittle, J.J. Yunis, J. Gómez, Results of the 1999-2000 collaborative testing program on mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor, Forensic Sci. Int. 125 (2002) 1–7.

[23] L. Prieto, M. Montesino, A. Salas, A. Alonso, C. Albarrán, S. Álvarez, M. Crespillo, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A.G. de la Vega, L. Gusmão, C.M. López, M. López-Soto, J.A. Lorente, M. Malaghini, C.A. Martínez, N.M. Modesti, A.M. Palacio, M. Paredes, S.D. Pena, A. Pérez-Lezaun, J.J. Pestano, J. Puente, A. Sala, M.C. Vide, M.R. Whittle, J.J. Yunis, J. Gómez, The 2000–2001 GEP-ISFG Collaborative Exercise on mtDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft samples, Forensic Sci. Int. 134 (2003) 46–53.

[24] A. Salas, L. Prieto, M. Montesino, C. Albarrán, E. Arroyo, M.R. Paredes-Herrera, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A.G. de la Vega, C. Alves, C.M. López, M. López-Soto, J.A. Lorente, A. Picornell, R.M. Espinheira, A. Hernández, A.M. Palacio, M. Espinoza, J.J. Yunis, A. Pérez-Lezaun, J.J. Pestano, J.C. Carril, D. Corach, M.C. Vide, V. Álvarez-Iglesias, M.F. Pinheiro, M.R. Whittle, A. Brehm, J. Gómez, Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02-03 proficiency testing trial, Forensic Sci. Int. 148 (2005) 191–198.

[25] M. Crespillo, M.R. Paredes, L. Prieto, M. Montesino, A. Salas, C. Albarran, V. Álvarez-Iglesias, A. Amorin, G. Berniell-Lee, A. Brehm, J.C. Carril, D. Corach, N. Cuevas, A.M. Di Lonardo, C. Doutremepuich, R.M. Espinheira, M. Espinoza, F. Gómez, A. González, A. Hernández, M. Hidalgo, M. Jimenez, F.P. Leite, A.M. López, M. López-Soto, J.A. Lorente, S. Pagano, A.M. Palacio, J.J. Pestano, M.F. Pinheiro, E. Raimondi, M.M. Ramón, F. Tovar, L. Vidal-Rioja, M.C. Vide, M.R. Whittle, J.J. Yunis, J. Garcia-Hirschfeld, Results of the 2003–2004 GEP-ISFG collaborative study on mitochondrial DNA: focus on the mtDNA profile of a mixed semen-saliva stain, Forensic Sci. Int. 160 (2006) 157–167.

[26] M. Montesino, A. Salas, M. Crespillo, C. Albarrán, A. Alonso, V. Álvarez-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.

[27] L. Prieto, A. Alonso, C. Alves, M. Crespillo, M. Montesino, A. Picornell, A. Brehm, J.L. Ramírez, M.R. Whittle, M.J. Anjos, I. Boschi, J. Buj, M. Cerezo, S. Cardoso, R. Cicarelli, D. Comas, D. Corach, C. Doutremepuich, R.M. Espinheira, I. Fernández-Fernández, S. Filippini, J. Garcia-Hirschfeld, A. González, B. Heinrichs, A. Hernández, F.P.N. Leite, R.P. Lizarazo, A.M. López-Parra, M. López-Soto, J.A. Lorente, B. Mechoso, I. Navarro, S. Pagano, J.J. Pestano, J. Puente, E. Raimondi, A. Rodríguez-Quesada, M.F. Terra-Pinheiro, L. Vidal-Rioja, C. Vullo, A. Salas, 2006 GEP-ISFG collaborative exercise on mtDNA: reflections about interpretation, artefacts, and DNA mixtures, Forensic Sci. Int. Genet. 2 (2008) 126–133.

[28] E.B. Wilson, Probable inference, the law of succession, and statistical inference, J. Am. Stat. Assoc. 22 (1927) 209–212.

[29] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P.M. Schneider, B. Budowle, B. Brinkmann, P.P. Gill, M. Holland, G. Tully, M. Wilson, DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.

[30] G. Tully, W. Bär, B. Brinkmann, A. Carracedo, P. Gill, N. Morling, W. Parson, P. Schneider, Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles, Forensic Sci. Int. 124 (2001) 83–91.

[31] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2009) E386–E394.

[32] T. Egeland, A. Salas, Estimating haplotype frequency and coverage of databases, PLoS One 3 (12) (2008) e3988.

[33] T. Egeland, A. Salas, Statistical evaluation of haploid genetic evidence, Open Forensic Sci. J. 1 (2008) 4–11.

[34] J.S. Buckleton, M. Krawczak, B.S. Weir, The interpretation of lineage markers in forensic DNA testing, Forensic Sci. Int. Genet. 5 (2011) 78–83.

[35] S.S. Willuweit, A. Caliebe, M.M. Andersen, L. Roewer, Y-STR frequency surveying method: a critical reappraisal, Forensic Sci. Int. Genet. 5 (2) (2011) 84–90.

[36] P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, I. Evett, E. Hagelberg, K. Sullivan, Identification of the remains of the Romanov family by DNA analysis, Nat. Genet. 6 (1994) 130–135.

[37] M. Richards, V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F. Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Gölge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Cali, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. Di Rienzo, A. Novelletto, A. Oppenheim, S. Nørby, N. Al-Zaheri, S. Santachiara-Benerecetti, R. Scozzari, A. Torroni, H.-J. Bandelt, Tracing European founder lineages in the Near Eastern mtDNA pool, Am. J. Hum. Genet. 67 (2000) 1251–1276.

[38] C.D. Calloway, R.L. Reynolds, G.L. Herrin, 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.

[39] K.M. Sullivan, R. Allison-Greiner, F.I.A. Archampong, R. Piercy, G. Tully, P. Gill, C. Lloyd-Davies, A single difference in mtDNA control region sequence observed between hair shaft and reference samples from a single donor, in: Proceedings of the Seventh International Symposium on Human Identification, Scottsdale (AZ), Promega Corporation, (1996), pp. 126–129.

[40] R. Del Bo, M. Crimi, M. Sciacco, G. Malferrari, A. Bordoni, L. Napoli, A. Prelle, I. Biunno, M. Moggio, N. Bresolin, G. Scarlato, G.P. Comi, High mutational burden in the mtDNA control region from aged muscles: a single-fiber study, Neurobiol. Aging 24 (2003) 829–838.

[41] J.A. Irwin, J.L. Saunier, H. Niederstätter, K.M. Strouss, K.A. Sturk, T.M. Diegoli, A. Brandstätter, W. Parson, T.J. Parsons, Investigation of heteroplasmy in the human mitochondrial DNA Control Region: a synthesis of observations from more than 5000 global population samples, J. Mol. Evol. 68 (2009) 516–527.