Forensic Science International: Genetics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

The GHEP–EMPOP collaboration on mtDNA population data—A new resource for forensic casework

L. Prieto [a], B. Zimmermann [b], A. Goios [c], A. Rodriguez-Monge [a], G.G. Paneto [d], C. Alves [c], A. Alonso [e], C. Fridman [f], S. Cardoso [g], G. Lima [h], M.J. Anjos [i], M.R. Whittle [j], M. Montesino [a], R.M.B. Cicarelli [d], A.M. Rocha [c], C. Albarrán [e], M.M. de Pancorbo [g], M.F. Pinheiro [h], M. Carvalho [i], D.R. Sumita [j], W. Parson [b,*]

[a] Comisaría General de Policía Científica, University Institute of Research in Forensic Sciences (IUICP), Madrid, Spain
[b] Institute of Legal Medicine, Innsbruck Medical University, Austria
[c] Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal
[d] Laboratory of Paternity, UNESP, Univ. Estadual Paulista, São Paulo, Brazil
[e] National Institute of Toxicology and Forensic Sciences (INTCF), Madrid, Spain
[f] Department of Legal Medicine, Bioethics and Occupational Health, Medical School, University of São Paulo, Brazil
[g] BIOMICs Research Group, Centro de Investigación y Estudios Avanzados "Lucio Lascaray", University of the Basque Country, Vitoria-Gasteiz, Spain
[h] National Institute of Legal Medicine, North Branch, Porto, Portugal
[i] National Institute of Legal Medicine, Centre Branch, Coimbra, Portugal
[j] Genomic Engenharia Molecular, São Paulo, Brazil

ARTICLE INFO

Article history:
Available online xxx

Keywords:
mtDNA
Mitochondrial DNA
GHEP-ISFG
EMPOP
Population analyses
Forensic science

ABSTRACT

Mitochondrial DNA (mtDNA) population data for forensic purposes are still scarce for some populations, which may limit the evaluation of forensic evidence especially when the rarity of a haplotype needs to be determined in a database search. In order to improve the collection of mtDNA lineages from the Iberian and South American subcontinents, we here report the results of a collaborative study involving nine laboratories from the Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) and EMPOP. The individual laboratories contributed population data that were generated throughout the past 10 years, but in the majority of cases have not been made available to the scientific community. A total of 1019 haplotypes from Iberia (Basque Country, 2 general Spanish populations, 2 North and 1 Central Portugal populations), and Latin America (3 populations from São Paulo) were collected, reviewed and harmonized according to defined EMPOP criteria. The majority of data ambiguities that were found during the reviewing process (41 in total) were transcription errors confirming that the documentation process is still the most error-prone stage in reporting mtDNA population data, especially when performed manually. This GHEP–EMPOP collaboration has significantly improved the quality of the individual mtDNA datasets and adds mtDNA population data as valuable resource to the EMPOP database (www.empop.org).

1. Introduction

The importance of mitochondrial DNA (mtDNA) analysis is still growing and nowadays it has become an essential technique in dedicated forensic laboratories [1]. It is usually investigated in forensic case work when not enough nuclear DNA is available in a questioned sample or when it is necessary to evaluate maternal relationships between individuals. When two mtDNA haplotypes cannot be excluded as originating from the same source mtDNA databases are queried to determine the rarity of that profile. Laboratories performing forensic mtDNA testing usually have data sets of their local population(s) at hand to aid frequency searches. Unfortunately, these data sets are usually not available to the general forensic community and therefore of limited use. Also, some of these data may contain errors or ambiguities as they only rarely – if at all – undergo independent data quality review [2]. However, they constitute a valuable source of information, as mtDNA population data for forensic purposes are generally still in demand. In order to make those data accessible, the individual data sets need to be collected, reviewed and harmonized in a number of aspects, including the systematic performance of plausibility checks, the minimization of error, the adaptation of the sequencing ranges and the standardized presentation (alignment and annotation) of the mtDNA haplotypes.

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

1872-4973/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2010.10.013

Table 1
List of participating laboratories in the collaborative GHEP-ISFG–EMPOP study.

Laboratory	Samples	Year of data generation	Range	Publication
Comisaría General de Policía Científica (Madrid, Spain)	249	2000–2010	Variable, but at least 16024–16365 and 72–340	This publication
National Institute of Toxicology and Forensic Sciences, INTCF (Madrid, Spain)	154	1995–2000	16024–16365 and 73–340	This publication
Laboratory of Paternity, UNESP, Univ. Estadual Paulista (São Paulo, Brazil)	142	2006–2010	16024–576	This publication
Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP (Porto, Portugal)	132	2008–2009	16024–576	This publication
Department of Legal Medicine, Bioethics and Occupational Health, Medical School, University of São Paulo, Brazil	102	2006–2009	16024–576	ONLY EMPOP + future publication
BIOMICs Research Group. Centro de Investigación y Estudios Avanzados "Lucio Lascaray". University of the Basque Country (Vitoria-Gasteiz, Spain)	84	2003–2007	16024–16383 and 66–370	29 New haplotypes this publication; 55 haplotypes already published in Ref. [5]
National Institute of Legal Medicine. North Branch (Porto, Portugal)	55	2005–2008	16024–16391 and 30–408; 10 codR SNPs + 1 non-coding region SNP	This publication
National Institute of Legal Medicine, Centre Branch (Coimbra, Portugal)	53	2000–2005	16024–16365 and 72–340	This publication
Genomic Engenharia Molecular (São Paulo, Brazil)	48	2002–2007	16024–16365 and 73–340	This publication
Total	1019			

The EDNAP Mitochondrial DNA Population Database (EMPOP) is a collaborative project among forensic and population genetic laboratories worldwide with the aim to increase the amount of reliable mtDNA population data in a searchable format via the internet (www.empop.org) [3]. The currently available version (Release 2) contains 10,970 haplotypes that have undergone meticulous revision using software-based format and plausibility control and inspection of the data with phylogenetic methods. Although populations of west Eurasian origin are the most well represented in EMPOP, it is necessary to continue their collection especially for underrepresented populations at the regional level, which is the case for Iberian and also South American lineages. In addition, the phenomenon of migration is influencing the dynamics of populations and new studies are necessary for a more accurate evaluation of the frequency and distribution of mtDNA lineages.

The current study follows a similar initiative driven by the Italian Ge.F.I-Group [4] which collected a total of 395 mtDNA haplotypes from Italy generated by 8 forensic laboratories. Those data were assembled and scrutinized with respect to EMPOP quality criteria and uploaded onto the database, thus making them available to the forensic community. In the current study, the Spanish and Portuguese-speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) has carried out a collaborative exercise by collecting and reviewing a total of 1019 haplotypes from different Iberian and Latin American populations that have been generated in the respective laboratories throughout the past 10 years. The current paper demonstrates the organization of the collaboration and the methods of data review. Observed ambiguities and questionable base calls were communicated to the authors who inspected raw data for review and clarification. Finally, comparative analysis of the Iberian populations is presented to support the data with forensically relevant information.

2. Materials and methods

2.1. Participants, samples and requirements

Participating laboratories and the number of contributed samples are shown in Table 1. This collaborative exercise was open to all the GHEP labs, which met the following requirements: (a) successful participation at the 2008 GHEP mtDNA proficiency test control excercise; (b) supply of mtDNA haplotypes of about 50 unrelated individuals (as far as could possibly be determined) with (c) established geographical origin (region/city/population); (d) minimum sequencing coverage of HVS-I (16,024–16,365) and HVS-II (73–340) and (e) retention of raw data, if available both forward and reverse sequence information. All data included herein have not been published elsewhere except for 55 samples from the Basque Country (total of 84) that were previously presented in [5]. Therefore, we add 29 new haplotypes from the Basque Country to the pool of data in the course of this study. We also note that a subset of 102 lineages from Brazil was part of the evaluation process described herein but the individual haplotypes will be published in a different context later (Table 1).

2.2. Summary of methods

The mtDNA sequences were generated between the years of 1995 and 2010. Therefore, a huge variety of methods in terms of DNA extraction, amplification, sequencing and electrophoresis were used. Therefore, we aimed at taking specific details into account that have a known effect on data interpretation, such as the older version of the Taq polymerase that left specific footprints in sequence electropherograms and was thus prone to introduce phantom mutations [6]. Details are summarized in Table 2.

2.3. EMPOP revision process

The analysis of mtDNA is usually more challenging for a forensic laboratory than Short Tandem Repeat typing. This is because of its biological characteristics that may lead to difficulties for interpretation, such as heteroplasmy and potential uncertainty of exclusion/non-exclusion scenarios as well as technical peculiarities, e.g. the lack of standardized commercial support to aid the laboratory process (manufacturing kits), the elevated risk of contamination and sequencing artifacts. In addition, there is a lack of automation of numerous steps in the entire laboratory process. Thus, the separate amplification of HVS-I and HVS-II, which harbors an increased risk of mixing up samples (artificial recombination) or the manual transfer of tabular data are some of the critical issues. Previous publications have aptly demonstrated these problems by example [9]. Therefore, a careful revision of

Please cite this article in press as: L. Prieto, et al., The GHEP–EMPOP collaboration on mtDNA population data—A new resource for forensic casework, Forensic Sci. Int. Genet. (2010), doi:10.1016/j.fsigen.2010.10.013

Table 2
Analysis methods employed to generate the mtDNA population data.

Laboratory	DNA extraction	Amplification primers	Sequencing primers	Sequencing chemistry	Sequencing machine
Comisaría General de Policía Científica (Madrid, Spain)	P/C/I-Centricon	L15997/H16395 or H17 L48/H408 L350/H619 or L16555/H619	L15997, H16395, L16555, L16209, H16164, L48, H17, H408, H285, L318, L350, H619	BigDye Terminator v2.0, v3.0 and v3.1	ABI 377/310/3130
National Institute of Toxicology and Forensic Sciences, INTCF (Madrid, Spain)	P/C/I-Centricon	L15997/H16391 L48/H408	L15997, H16391, L16209, H16164, L48, H408	dRhodamine Terminator	ABI 377
Laboratory of Paternity, UNESP, Univ. Estadual Paulista (São Paulo, Brazil)	FTA Reagent (Whatman)	L15997/H639	L15997, H16401, L16209, H16164, L29, H408, H159, H285, L314, H599, H639	Big Dye Terminator v3.1	ABI 3130
Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP (Porto, Portugal)	Chelex	L15997/H639 L15900/H599	L15900, L15997, H16, H159, L16268, L16555, L314, H599, H639	Big Dye Terminator v3.1	ABI 3130/3100
Department of Legal Medicine, Bioethics and Occupational Health, Medical School, University of São Paulo, Brazil	Salting out [7]	L15978/H16420 L29/H306 L153/H429 L256/H653	L15978, H16420, L29, H306, L153, H429, L256, H653	BigDye Terminator v3.1	ABI 3100/3130
BIOMICs Research Group. Centro de Investigación y Estudios Avanzados "Lucio Lascaray". University of the Basque Country (Vitoria-Gasteiz, Spain)	Organic	L15996/H16401 L29/H408	L15996, L29, H16401, H408	dRhodamine Terminator and Big Dye Terminator v3.1	ABI 310/3130
National Institute of Legal Medicine. North Branch (Porto, Portugal)	Chelex or P/C/I	L15996/H16401 L29/H408 SNPs: [8]	M13 Forward, M13 Reverse	BigDye Terminator v1.1	ABI 310/3100
National Institute of Legal Medicine, Centre Branch (Coimbra, Portugal)	Chelex	L15997/H16401/L16209/H16164 L48/H408/L314/H285	L15997/H16401/L16209/H16164 L48/H408/L314/H285	BigDye Terminator v1.1	ABI 3130
Genomic Engenharia Molecular (São Paulo, Brazil)	FTA Reagent (Whatman)	L15990/H16391 L34/H370	L15990/H16391/L16190/H16187 L34/H370/L313/H306	BigDyeTerminator v3.1	ABI 377/3130xl

the mtDNA haplotypes is crucial before they can be used for forensic interpretation in mtDNA databases. We performed IT-based evaluation of the data using formal and phylogenetic methods, such as NETWORK [3,10] to evaluate the following sources of error:

(a) Reference bias.
(b) Phantom mutations.
(c) Base mis-scoring.
(d) Nomenclature issues.
(e) Alignment violation.
(f) Clerical errors.

We further aimed at achieving uniformity regarding the following aspects:

(g) Haplogroup assignment, following [11; phylotree, build 10].
(h) Alignment and annotation in length variant regions.
(i) Confirmation of point heteroplasmy.
(j) Revision of sample affiliation (metadata).
(k) Achieving best possible uniformity of sequence ranges.

Compilation and revision processes were carried out at the Comisaría General de Policía Científica (Madrid) and reviewed by the EMPOP group at the Institute of Legal Medicine, Innsbruck Medical University. All polymorphisms were finally cross-referenced against commonly observed phantom mutations [12] and apparent "new polymorphisms" were evaluated using mtDNA literature data and direct Internet queries [13]. When necessary, contributing authors were asked to support their findings with raw data (electropherograms) to evaluate specific polymorphisms.

2.4. Population studies

Molecular diversity indices, pairwise differences between and within populations and an analysis of molecular variance (AMOVA) were calculated using ARLEQUIN (Version 3.5) [14]. The random match probability was calculated as the sum of squared haplotype frequencies based on mtDNA control region sequences. All sequences were aligned and trimmed to a greatest common range of ntps 16024–16365 and ntps 73–340, length variation around ntps 16193 and 309 was disregarded.

3. Results and discussion

3.1. Results of the revision process

A total of 1019 mtDNA haplotypes from 9 populations were examined in the present study (Table 1 and Table S1) of which 154 (from Spain) were already contributed and evaluated earlier. Another 249 haplotypes came from the organizing laboratory (Madrid) and 132 (North Portugal) were generated *de novo* in the course of this project. Therefore the total number of yet unreviewed haplotypes was 484. The communication with the authors of the sequences allowed the correction of questionable polymorphisms in 41 haplotypes (8.5%). The following sections list those according to their source (see also Section 2 and Table 3).

3.1.1. Reference bias

Reference bias is one of the most abundant forms of clerical error which is manifest in a failure to report a polymorphism relative to the rCRS. Note that in some cases (not observed here) also other "Anderson sequences" are mistakenly used as reference sequence to which the consensus sequences are reported, which can then result in a similar problem. Reference bias is more frequently observed at the beginning and at the end of sequencing strands, due to decreased quality of the electropherograms there. If reverse sequencing reactions are missing or of low quality, reference biases are more frequent. In the present study we noted 8 instances 3 of which were located at the beginning and 3 at the end of the sequences (Table 3a).

Table 3
Classification of ambiguities after revision and confirmation by the raw lane data.

Polymorphism		Times	
(a) Reference bias			
72C		1	
73G		2	
210G		1	
315.1C		1	
16355T		1	
16360T		1	
16390A		1	Total = 8
Position		**Times**	
(b) Phantom mutation			
16293M		1	
527G		1	Total = 2
Mistaken	**Correct**	**Times**	
(c) Base mis-scoring			
114G	114A	1	
146T	146C	1	
150C	150T	2	
150G	150T	1	
152T	152C	2	
195T	195C	1	
16278G	16278T	1	
16356T	16356C	3	Total = 12
Position		**Times**	
(d) Nomenclature			
309.2C without 309.1C		8	Total = 8
Position		**Times**	
(e) Alignment violation			
523.1C 524.1A instead of 524.1A 524.2C		3	Total = 3
Mistaken	**Correct**	**Times**	
(f) Clerical errors			
163G	263G	1	
315C	315.1C	2	
1620G	16207G	1	
16218C	16182C	1	
16223	16223T	1	
16278C	16288C	1	
19294T	16294T	1	Total = 8

3.1.2. Phantom mutations

Artificial signals in the sequencing electropherograms (e.g. dye blobs, unincorporated dye terminators, inadequate migration conditions leading to shoulder peaks, secondary structures, polymerase footprints, etc.) are referred to as phantom mutations, as they are designated by some analysis software as genuine base calls. This emphasizes the need of manual data review, especially when sequence quality is low. Phantom mutations are usually also located at sequence beginnings and ends, as the quality of the electropherograms is lower there. We observed two instances in this study (Table 3b), where one (527G) is a well-known phantom hot spot [12].

3.1.3. Base mis-scoring

Base mis-scoring was found to be the most frequent error in the present study (Table 3c). It originates from manual data transfer and insufficient results review. The majority of these could be identified by applying stringent scrutiny when checking the data

Please cite this article in press as: L. Prieto, et al., The GHEP–EMPOP collaboration on mtDNA population data—A new resource for forensic casework, Forensic Sci. Int. Genet. (2010), doi:10.1016/j.fsigen.2010.10.013

Table 4
Descriptive statistics for six populations from the Iberian Peninsula. Analyzed range: ntps 16024–16356, 73–340.

Population statistics	Basque [$n=84$]	Central Portugal [$n=53$]	North Portugal [$n=55$]	North Portugal [$n=132$]	Spain [$n=249$]	Spain [$n=154$]
Number of haplotypes	47	44	50	105	193	124
Number of unique haplotypes	31	40	47	88	167	114
Random match probability	0.043	0.033	0.023	0.014	0.014	0.016
Genetic diversity	0.957	0.967	0.977	0.986	0.986	0.984

Table 5
AMOVA results for the six investigated Iberian populations.

Source of variation	d.f.	Sum of squares	Variance components	Percent of variation
(a) Design and results (d.f. stands for degrees of freedom)				
Among populations	5	22.162	0.00839 Va	0.24
Within populations	721	2508.999	3.47989 Vb	99.76
Total	726	2531.161	3.48828	

	Basque [$n=84$]	Central Portugal [$n=53$]	North Portugal [$n=55$]	North Portugal [$n=132$]	Spain [$n=249$]	Spain [$n=154$]
(b) FST comparison among the regional populations						
Basque [$N=84$]	*	0.1290	0.0342	0.0049	0.0049	0.2432
Central Portugal [$N=53$]	0.0053	*	0.8731	0.43848	0.2002	0.3516
North Portugal [$N=55$]	0.0100	0.0000	*	0.77051	0.2891	0.4502
North Portugal [$N=132$]	0.0113	0.0000	0.0000	*	0.0986	0.4102
Spain [$N=249$]	0.0079	0.0025	0.0013	0.0022	*	0.1807
Spain [$N=154$]	0.0016	0.0006	0.0000	0.0002	0.0013	*

	Basque [$n=84$]	Central Portugal [$n=53$]	North Portugal [$n=55$]	North Portugal [$n=132$]	Spain [$n=249$]	Spain [$n=154$]
(c) Population average pairwise differences						
Basque [$N=84$]	5.82	6.40	7.11	6.85	6.16	6.62
Central Portugal [$N=53$]	0.03	6.92	7.56	7.31	6.67	7.16
North Portugal [$N=55$]	0.06	0.04	8.28	7.97	7.34	7.83
North Portugal [$N=132$]	0.08	0.00	0.02	7.71	7.07	7.55
Spain [$N=249$]	0.05	0.01	0.00	0.01	6.40	6.90
Spain [$N=154$]	0.02	0.01	0.01	0.00	0.01	7.40

F_{ST} values are below the diagonal and the p-values (1023 permutations, significance level = 0.05) above the diagonal.
Above diagonal: average number of pairwise differences between populations ($PiXY$); diagonal elements: average number of pairwise differences within population (PiX); below diagonal: corrected average pairwise difference ($PiXY - (PiX + PiY)/2$).

tables or by using automated plausibility checks, such as provided by the *emp-tool* (www.empop.org/modules/emptool/).

3.1.4. Nomenclature issues

One participating laboratory called 8 instances of 2 C-insertions between positions 303 and 310 as only 309.2C instead of the commonly used term 309.1C 309.2C. While this constitutes a minor issue the explicit documentation of 309.1C makes clear that there is no other base inserted here (Table 3d).

3.1.5. Alignment violation

The dinucleotide repeat region between ntps 514 and 524 has earlier been referred to as CA-repeat [15] and was later changed to an AC-repeat-based nomenclature in order to better accommodate a commonly observed transition at ntp 513 [16]. Since then AC-insertions relative to the rCRS (five repeats) are reported as 524.1A 524.2C (in contrast to the earlier formulated 523.1C 523.2A). In the present study we observed the designation of 523.1C 524.1A, which is incompatible with both alignment schemes (Table 3e). In general, the phylogenetically meaningful alignment is recommended [17].

3.1.6. Clerical error

While some of the above mentioned issues can also be regarded as clerical errors, we list only those here that are undoubtedly introduced by manual data transfer (Table 3f). Again, those would be captured by some electronic evaluation of the data table, such as the *emp-tool*.

3.2. Results of the Iberian population comparisons

A total of 727 mtDNA control region haplotypes from 6 Iberian populations (Basque, Central Portugal, 2 North Portugal and 2 mixed Spain; Tables S1 and 4) were analyzed and AMOVA was used to test for significant variation in the genetic structure (Table 5). Most of the observed genetic variation was attributable to differences within populations (99.76%). Variance among populations accounted for 0.24% (Table 5a). The Basque population

Table 6
Observed haplogroup frequencies in the Iberian populations.

Haplogroup	Basque [$n=84$]	Central Portugal [$n=53$]	North Portugal [$n=55$]	North Portugal [$n=132$]	Spain [$n=249$]	Spain [$n=154$]
R0	67.9%	49.1%	49.1%	45.5%	56.6%	51.3%
JT	15.5%	26.4%	20.0%	17.4%	14.9%	13.0%
UK	9.5%	15.1%	16.5%	19.7%	22.9%	20.1%
R*	2.4%	1.9%	3.6%	1.5%	0.4%	4.6%
N*	4.7%	7.5%	3.6%	9.8%	2.4%	7.8%
M	0.0%	0.0%	3.6%	1.5%	0.8%	0.6%
L	0.0%	0.0%	3.6%	4.6%	2.0%	2.6%

Please cite this article in press as: L. Prieto, et al., The GHEP–EMPOP collaboration on mtDNA population data—A new resource for forensic casework, Forensic Sci. Int. Genet. (2010), doi:10.1016/j.fsigen.2010.10.013

differed significantly in its composition of mtDNA lineages from both North Portuguese populations and one mixed Spanish population (Table 5b). This result may be explained by the relative overrepresentation of hg R0 lineages in the Basque population sample and the lack of hg L lineages that are present, albeit at low frequencies, in the other populations (Table 6). We note here that the different sample sizes may also have an effect on these results.

All Iberian populations shared (common) haplotypes to relatively great extent (Table S2). The Basque shared approximately half of their haplotypes (46.81%) with other Iberian populations from Spain and Portugal. All six Iberian populations included the same most common haplotype 263G 315.1C that represents the most common HVS-I/II haplotype in west Eurasia (here grouped under hg R0).

4. Conclusions

One of the most important issues in the forensic use of mtDNA analyses is the difficulty of accurately transmitting the significance of a match (non-exclusion) between unknown and reference samples to court. Non-DNA experts may not immediately be aware of the difference between nDNA and mtDNA evidence, which can then lead to overestimation of the mtDNA match (or underestimation of its significance when only statistical numbers are compared). Also reliable mtDNA population data in forensics are still scarce although many studies have been published. A sometimes unacceptable rate of error makes some of these studies unfortunately unusable. This is one of the main reasons why forensic mtDNA database projects need to be expanded. Due to the wide variability of populations that are presented in the GHEP-ISFG group and in order to join forces and make individual datasets available to the forensic community, we have carried out the present project in collaboration with the EMPOP database. The remittance of our data has been very useful since some of our populations are not represented in EMPOP (Release 2) yet.

Our data reviewing process confirmed earlier findings [2,18] that the majority of errors occur due to manual documentation processes without rigorous scrutiny. This study demonstrates that *a posteriori* plausibility and phylogenetic evaluations help to uncover data idiosyncrasies and obvious errors. By inspection of the raw data we were then able to solve ambiguities.

Acknowledgements

Antonio Amorim is greatly acknowledged for hosting the inauguration of the GHEP–EMPOP collaboration and for useful discussion. We are grateful to the contribution of Alexander Röck (Innsbruck), who provided software tools to handle the data. We would like to thank Theresa Harm (Innsbruck) for careful analysis of raw data. This study received support by the Austrian Science Fund (FWF): TR397.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2010.10.013.

References

[1] M.M. Holland, T.J. Parsons, Mitochondrial DNA sequence analysis-validations and use for forensic casework, Forensic Sci. Rev. 11 (1999) 21–50.
[2] W. Parson, H.-J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (2007) 13–19.
[3] W. Parson, A. Dür, EMPOP—a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88–92.
[4] C. Turchi, L. Buscemi, C. Previderè, P. Grignani, A. Brandstätter, A. Achilli, W. Parson, A. Tagliabracci, Ge.F.I. Group, Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing, Int. J. Legal Med. 122 (2008) 199–204.
[5] M.A. Alfonso-Sánchez, S. Cardoso, C. Martínez-Bouzas, J.A. Peña, R.J. Herrera, A. Castro, I. Fernández-Fernández, M.M. De Pancorbo, Mitochondrial DNA haplogroup diversity in Basques: a reassessment based on HVI and HVII polymorphisms, Am. J. Hum. Biol. 20 (2008) 154–164.
[6] W. Parson, The art of reading sequence electropherograms, Ann. Hum. Gen. 71 (2007) 276–278.
[7] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, Nucleic Acids Res. 16 (3) (1988) 1215.
[8] P.M. Vallone, R.S. Just, M.D. Coble, J.M. Butler, T.J. Parsons, A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome, Int. J. Legal Med. 118 (2004) 147–157.
[9] H.-J. Bandelt, P. Lahermo, M. Richards, V. Macaulay, Detecting errors in mtDNA data by phylogenetic analyses, Int. J. Legal Med. 115 (2001) 64–69.
[10] 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.
[11] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2009) E386–E394.
[12] A. Brandstätter, T. Sänger, S. Lutz-Bonengel, W. Parson, E. Béraud-Colomb, B. Wen, Q.-P. Kong, C.M. Bravi, H.-J. Bandelt, Phantom mutation hotspots in human mitochondrial DNA, Electrophoresis 26 (2005) 3414–3429.
[13] H.-J. Bandelt, A. Salas, C.M. Bravi, What is a 'novel' mtDNA mutation – and does 'novelty' really matter? J. Hum. Genet. 51 (2006) 1073–1082.
[14] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.
[15] S. Lutz, H.J. Weisser, J. Heizmann, S. Pollak, A third hypervariable region in the human mitochondrial D-loop, Hum. Genet. 101 (1997) 384.
[16] M.R. Wilson, M.W. Allard, K. Monson, K.W. Miller, B. Budowle, Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region, Forensic Sci. Int. 10 (2002) 35–42.
[17] 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.
[18] 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. Sørensen, 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.