J Forensic Sci, Nov. 2004, Vol. 49, No. 6
Paper ID JFS2004044
Available online at: www.astm.org

TECHNICAL NOTE

Bruce Budowle,[1] Ph.D.; Deborah Polanskey,[1] B.S.; Marc W. Allard,[2] Ph.D.; and Ranajit Chakraborty,[3] Ph.D.

Addressing the Use of Phylogenetics for Identification of Sequences in Error in the SWGDAM Mitochondrial DNA Database*

ABSTRACT: The SWGDAM mtDNA database is a publicly available reference source that is used for estimating the rarity of an evidence mtDNA profile. Because of the current processes for generating population data, it is unlikely that population databases are error free. The majority of the errors are due to human error and are transcriptional in nature. Phylogenetic analysis of data sets can identify some potential errors, and coupled with a review of the sequence data or alignment sheets can be a very useful tool. Seven sequences with errors have been identified by phylogenetic analysis. In addition, two samples were inadvertently modified when placed in the SWGDAM database. The corrected sequences are provided so that users can modify appropriately the current iteration of the SWGDAM database. From a practical perspective, upper bound estimates of the percentage of matching profiles obtained from a database search containing an incorrect sequence and those of a database containing the corrected sequence are not substantially different. Community wide access and review has enabled identification of errors in the SWGDAM data set and will continue to do so. The result of public accessibility is that the quality of the SWGDAM forensic dataset is always improving.

KEYWORDS: forensic science, mitochondrial DNA, phylogenetics, sequence error, SWGDAM database

Mitochondrial DNA (mtDNA) sequencing is a well-accepted methodology for analyzing forensic biological samples, particularly those materials that contain very little intact nuclear DNA suitable for reliably typing the current nuclear forensic markers (1–10). In a forensic analysis, the mtDNA types between a known exemplar(s) and an evidence sample(s) are compared. When the interpretation is a failure to exclude the known and evidence samples as possibly having the same origin (or the same maternal lineage), information is provided about the rarity of the mtDNA profile. The current practice is to count the number of times a particular sequence is observed in a database(s) and place an upper bound on this to correct for sampling (11). Thus, the quality of the reference database and how this is to be accommodated for assessing the weight of the evidence needs to be addressed.

Recently, there have been a number of reports suggesting and documenting that errors can and do exist in some sequences in a number of reference population databases (12–16). We laud those who are attempting to identify such errors and make them known. Due to such efforts, the forensic databases that are used will continue to improve. Because of current technical limitations, however, it is unlikely that any forensic mtDNA reference database will be completely error free (12–14). But we can strive to have databases that contain a minimal number of errors. While the effect of many of the previously identified errors on an upper bound estimate of the mtDNA profile frequency typically is marginal (12), identification and rectification of errors should be carried out whenever possible. For forensic mtDNA databases, quality improvement should be a primary practice. Indeed, in 2001 the SWGDAM mtDNA database was reviewed manually for transcriptional errors. Even with this scouring of the data, it is likely that a few transcriptional errors remain.

Identifying Errors

Identification of reasonable sources of error can focus efforts and enable better review mechanisms for continuously improving the SWGDAM mtDNA database. Bandelt et al. (12) suggested that sequence errors in the database may arise due to biochemistry, contamination, merging of sequences from two different sources (creating an "odd concatenation"), and transcriptional errors.

Errors due to biochemistry are unlikely. If errors were chemistry based, then systematic errors, not the more likely random errors, would be observed. Some suggest that the sequencing error rate is about 1% (17). This is a grossly misplaced value. This estimate was based on older chemistry, earlier version electrophoretic instruments, and software less developed than current programs (18,19), and thus does not apply stringently to current methodologies. In addition, automated and manual reviews of the sequence reads are done routinely in many forensic laboratories. Moreover, the 1%

[1] FBI Laboratory, 2501 Investigation Parkway, Quantico, VA.
[2] Biology Department of Biological Science, George Washington University, Washington, DC.
[3] Center for Genome Information, Dept. of Environmental Health, University of Cincinnati, Cincinnati, OH.
* This is publication number 04-04 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.
Received 8 Feb. 2004; and in revised form 18 May 2004; accepted 30 May 2004; published 5 Oct. 2004.

Copyright © 2004 by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.

error rate combines sequence mismatches, ambiguities, insertions and deletions. In current practices, mismatches identified between aligned light and heavy strand sequences of a mtDNA amplicon and ambiguities are both classified as ambiguities. Ambiguities are designated as "N" and are not assigned a base call. Therefore, ambiguities are not real errors. The 1% error rate is an overestimate. Another practice that impacts on the error rate is that the typical practice is to sequence both the heavy and light strands and align and compare both sequences before rendering a sequence call. Thus, an error would have to occur in both strands at exactly the same position. So even if one were to assume (incorrectly) that the 1% error rate was a realistic value, the operational error rate would be substantially less than 0.001%. It is also worth noting that the nucleotide misincorporation rate of the polymerase is included as part of the 1% value. But, the misincorporation rate of *Taq* polymerase could only have an impact on obtaining an error from the sequence output when very few template molecules are copied during PCR or cycle sequencing. For most, if not all, database samples, there are ample numbers of template molecules for amplification. Thus, analytical technology and biochemistry do not appear to be major contributors to the observed errors.

A second proposed source of error in the databases is contamination. Because of the sensitivity of the mtDNA assay, the potential for contamination and its effects must be considered and monitored. However, because of the quantity and quality of samples that are typed for reference databases, errors due to contamination would be unlikely. In most instances, contamination that could affect an interpretation would result in a mixture which would readily be identified.

Most errors in our SWGDAM database are due to human involvement in data processing (e.g., transcriptional error). One might imply that these errors can arise equally in both casework analyses and in database generation (14). While errors, particularly of human origin, can arise, these two sets of data are not equivalent with respect to the chance of containing errors. For casework, analysis involves low throughput scrutiny of individual data profiles both automatically and manually, followed by a technical and administrative review. In contrast, database sequences tend to be generated using high throughput protocols, and while two readers review profile data manually, upon entrance into the database these are more prone to transcriptional error. To date (since 1996) we have not found a sequence error in our casework (although such could occur for casework), but some have been observed in the SWGDAM reference population data sets (see below).

Phylogenetics and Error Detection

Because most errors tend to be transcriptional in nature, mechanisms can be put in place to reduce their occurrence. Phylogenetic network analyses (12) have been recommended to assist in identifying incongruent haplogroup data and thus possible errors. The accumulation of incorrect combinations and/or unusual linkage of certain variable sites can indicate possible transcriptional errors. As an example, Bandelt (personal communication) identified eight samples that belong to Haplogroup A that could be in error in the SWGDAM Hispanic data set. The premise for this supposition is the absence of the 235G (and presence of 235A or the presence of 253G) polymorphism in these particular samples. These samples were: USA.HIS.000093, USA.HIS.000100, USA.HIS.000110, USA.HIS.000204, USA.HIS.000267, USA.HIS.000274, USA.HIS.000552, and USA.HIS.000770. Indeed, three of these samples appear to have transcriptional errors (Samples 1–3 in Table 1). Thus, phylogenetic analysis was helpful in identifying these transcriptional errors. However, the other five identified by Bandelt are not errors, but are true reversals. The data were reviewed by assessing alignment data, raw sequence data, and/or resequencing (Fig. 1). Therefore, while we endorse phylogenetic network analyses for reviewing data as part of the quality control assessment of any new data added to the SWGDAM database, we urge caution. While phylogenetic analysis is a useful tool to identify obvious and potential errors (12,13), it should not be used as an absolute rule for asserting error (14). If we had changed the sequence on these five samples to coincide with the character state 235G, even more formidable errors would have been introduced into the data set. The maintained SWGDAM data set enables us by direct review of data to confirm or refute a potential error identified by phylogenetic analysis. This would be more difficult to accomplish with other published sequence data in public domain databases, or in the literature.

To further investigate errors in the database using phylogenetics, we looked for other sequences that might be considered transcriptional errors or "odd" reversals. Table 2 displays the haplogroup defining sites and number of individuals per population group that were investigated. These samples, a subset of possible samples, were those that could be readily reviewed. Only one transcriptional error was found (Sample 7 in Table 1).

If phylogenetic analysis is to be used, one should be circumspect on assertions of some errors as real until they can be confirmed. For example, during the review process of a population study by Budowle et al. (20), a reviewer insisted that sample #75 (belonging to haplogroup T) in the Apache sample population was in error, because the site 16126 was displayed as a T instead of C. Although, we conveyed the sequence had been re-checked and was correct, the reviewer continued to insist it was an error because of the inconsistency with phylogenetic analysis. The reviewer did not consider the possibility of a reversal until the raw data profile was sent to the Editor. If one relies solely on phylogenetic analysis and does not accept that some character changes may be true reversals, the amount of error may be overstated.

Another aspect of phylogenetics-based scrutiny of mtDNA sequence data to consider is that many currently employed phylogenetic methods rely on the assumption that sequence site-specific mutations are unique, and no back mutations are involved. In the mtDNA context (as the mtDNA molecules lack an efficient repair system during replication), this assumption (technically called the Infinite Site Model of mutations) is not strictly correct. In the HV1 and HV2 regions of mtDNA, direct as well as indirect signatures of recurrent mutations are seen in many databases. For example, 8.6% of the HV1 region nucleotide sites and 4.9% of those in the HV2 region included in the SWGDAM mtDNA database have three or four segregating alleles. In addition, nearly 68% of the pairs of bi-allelic polymorphic HV1 sites and almost 88% of the pairs in HV2 region fail the 4-gamete test of perfect linkage disequilibrium (Chakraborty and Budowle, unpublished data). This suggests that even among the bi-allelic polymorphic nucleotide sites of both HV1 and HV2 regions of the mtDNA genome, there are signatures of back mutations. When such events are not accounted for, phylogenetic network-based inference may ascribe the observed discordance with expectations either to an error or to mechanisms that are not biologically correct (e.g., recombination). As a consequence, mtDNA sequences carrying signatures of such recurrent mutations, not investigated, may be incorrectly viewed as containing errors. Again, we advocate the use of phylogenetic analyses to identify potential errors, but do not support reliance solely on such analyses for verification of sequence errors in mtDNA.

TABLE 1—*Identified sequences in error in SWGDAM mtDNA database and corrected sequences.**

Sample	Sequence Identified	Corrected Sequence
1. USA.HIS.000100	16111T	16111T
	16223T	16223T
	16290T	16290T
	16319A	16319A
	16362C	16362C
	73G	73G
	146C	146C
	153G	153G
	253G	**235G**
	263G	263G
	315.1C	315.1C
2. USA.HIS.000110	16111T	16111T
	16223T	16223T
	16290T	16290T
	16319A	16319A
	16335G	16335G
	73G	73G
	146C	146C
	153G	153G
	236G	**235G**
	263G	263G
	309.1C	309.1C
	315.1C	315.1C
3. USA.HIS.000204	16111T	16111T
	16182C	16182C
	16183C	16183C
	16189C	16189C
	16223T	16223T
	16290T	16290T
	16319A	16319A
	16362C	16362C
	73G	73G
	146C	146C
	153C	**153G**
	235C	**235G**
	263G	263G
	310C	310C
	315–	315–
	316–	316–
	317–	317–
4. USA.AFR.000063	16067T	**16189C**
	16126C	**16223T**
	16187T	**16278T**
	16189C	**16294T**
	16204A	**16309G**
	16223T	
	16264T	
	16270T	
	16278T	
	16311C	
	73G	73G
	143A	143A
	146C	146C
	152C	152C
	195C	195C
	263G	263G
	264T	264T
	315.1C	315.1C
5. USA.AFR.000074	16126C	16126C
	16187T	16187T
	16189C	16189C
	16223T	16223T
	16264T	16264T
	16270T	16270T
	16278T	16278T
	16311C	16311C
	73G	**73G**
	143A	**152C**
	146C	**182T**
	152C	**185T**
	195C	**189G**
	263G	**195C**
	264T	**247A**
	315.1C	**263G**
		315.1C
		357G
6. USA.AFR.000942	16126C	16126C
	16187T	16187T
	16189C	16189C
	16223T	16223T
	16264T	16264T
	16270T	16270T
	16278T	16278T
	16293G	16293G
	16311C	16311C
	16519C	16519C
	73G	**73G**
	249–	**152C**
	263G	**182T**
	290–	**185T**
	291–	**189G**
	309.1C	**195C**
	315.1C	**247A**
	489C	**263G**
		315.1C
		357G
		523–
		524–
7. USA.AFR.001201	16051C	**16051G**
	16223T	16223T
	16264T	16264T
	16519C	16519C
	73G	73G
	150T	150T
	263G	263G
	315.1C	315.1C
	493G	493G
	523–	523–
	524–	524–
8.[†] USA.CAU.000001	16222T	16222T
		16224C
	16311C	16311C
	73G	73G
	146C	146C
	263G	263G
	297G	297G
	309.1C	309.1C
	315.1C	315.1C
9.[†] USA.CAU.000002	16091G	**16051G**
	16192T	16192T
	16256T	16256T
	16270T	16270T
	16316G	16316G
	73G	73G
	263G	263G
	315.1C	315.1C

* Sites in bold are those that were identified and corrected.
† These samples were inadvertently modified when placed in the SWGDAM database and thus are provided here as well.

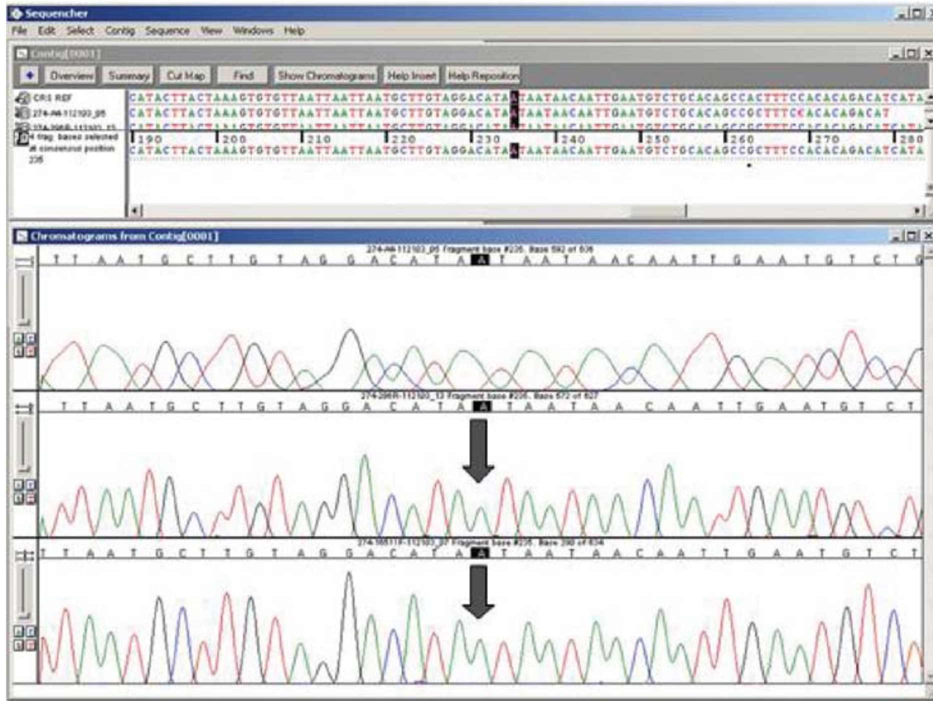

FIG. 1—*Sequence electropherogram of sample USA.HIS.000274 demonstrating at position 235 the presence of an A. The putative state at the site was confirmed by review of original alignment data. Then, the sample was re-sequenced and displayed here. Thus, the sequence in the SWGDAM database is correct. The arrow points to site 235.*

TABLE 2—*Populations, number of individuals, and sites searched to identify potential transcriptional errors.*

African American $N = 196$	Asian $N = 24$	Caucasian $N = 11$	Hispanic $N = 33$	Native American $N = 48$
189 A/G	489 C/T	217 C/T	235 A/G	146 C/T
195 C/T	16136 C/T	489 C/T	499 A/G	150C/T
200 A/G	16162 A/G	16162 A/G	16189 C/T	152 C/T
204 C/T	16164 C/T	16270 C/T	16223 C/T	153 A/G
207 A/T	16183 C/T	16294 C/T	16298 C/T	235 A/G
236 C/T	16217 C/T	16296 C/T	16290 C/T	272 G/T
297 A/G	16223 C/T	16304 C/T	16319 A/G	16086 C/T
316 A/G	16261 C/T	16356 C/T	16362 C/T	16092 C/T
325 C/T	16298 C/T			16111 C/T
489 C/T	16304 C/T			16126 C/T
499 A/G				16147 T/C
16051 C/T				16183 C/T
16114 C/T				16189 C/T
16124 C/T				16192 C/T
16126 C/T				16217 T/C
16148 C/T				16223 T/C
16172 C/T				16274 A/G
16189 C/T				16294 C/T
16223 C/T				16325 C/T
16230 A/G				16331 A/G
16327 C/T				16362 C/T
16257 C/T				
16264 C/T				
16265 A/G				
16278 C/T				
16286 C/T				
16292 C/T				
16294 C/T				
16309 A/G				
16320 C/T				
16327 C/T				

In other circumstances, phylogenetic methods can identify obvious errors. We have identified three samples in the African American data set that appear to be "odd concatenations." They are samples 4–6 displayed in Table 1. For these samples, regions HV1 and HV2 were typed separately and then the two separate profiles were merged to create the composite profile of each individual. In other words, incorrect HV1 and HV2 sequences were merged. Current procedures utilizing phylogenetic analysis will reduce such human error.

Bandelt (12) also has called for accessibility of the forensic mtDNA databases. We concur. The current iteration of the SWGDAM database has been on the web and is readily accessible (21). The SWGDAM database was made publicly accessible, so many forensic analysts could use the data for casework analyses and others could use the data for research purposes. Community wide access and review will and have identified errors in our data set. The result is that the quality of our forensic data set is always improving. Therefore, we advocate the public accessibility of forensic reference databases because of the benefit of improving quality. But, because of this continuous review process, we cannot agree that our forensic SWGDAM database is of the same quality as published literature. Primarily, most published data are static, and dynamic processes for corrections do not occur.

We also are hesitant to globally merge the SWGDAM forensic database with other published anthropological and evolutionary studies (12). Ideally, the more data available for assessing upper bound frequency estimates, the narrower will be the confidence interval for estimates. However, there are practical issues to consider. First, not all of the publicly available data cover the specific regions of nucleotide positions (i.e., HV1 and HV2) sequenced in our routine forensic analyses. So only a subset can be applied. Second, some of the reports do not present the entire sequence within the specified region per sample but instead provide only haplogroup specific sites; so again only a subset of the data will be available.

TABLE 3—*Upper bound estimate (95% CI) of percentage profiles in a SWGDAM mtDNA reference population* matching corrected individual sequences.*

Sample	Population	Prior Correction	After Correction
USA.HIS.000100	Apache	12.37	13.05
	Navajo	20.87	21.66
	Hispanic	5.18	5.34
	Caucasian	0.16	0.26
	African American	0.26	0.26
USA.HIS.000110	Apache	1.65	1.65
	Navajo	3.26	4.36
	Hispanic	0.39	0.39
	Caucasian	0.17	0.17
	African American	0.26	0.26
USA.HIS.000204	Apache	1.65	1.65
	Navajo	2.03	2.03
	Hispanic	0.39	0.39
	Caucasian	0.17	0.17
	African American	0.26	0.26
USA.AFR.000063	Apache	1.65	1.65
	Navajo	2.03	2.03
	Hispanic	0.39	0.39
	Caucasian	0.17	0.17
	African American	0.26	0.26
USA.AFR.000074	Apache	1.65	1.65
	Navajo	2.03	2.03
	Hispanic	0.39	0.39
	Caucasian	0.17	0.17
	African American	0.26	0.26
	Sierra Leone	5.82	7.20
USA.AFR.000942	Apache	1.65	1.65
	Navajo	2.03	2.03
	Hispanic	0.39	0.63
	Caucasian	0.17	0.17
	African American	3.43	3.54
USA.AFR.001201	Apache	1.65	1.65
	Navajo	2.03	2.03
	Hispanic	0.39	0.39
	Caucasian	0.17	0.17
	African American	0.69	0.82
USA.CAU.000001	Apache	1.65	1.65
	Navajo	2.03	2.03
	Hispanic	0.39	0.39
	Caucasian	0.17	0.17
	African American	0.26	0.26
USA.CAU.000002	Apache	1.65	1.65
	Navajo	2.03	2.03
	Hispanic	0.39	0.39
	Caucasian	0.17	0.17
	African American	0.26	0.26

* Sample population sizes are: Apache ($N = 180$); Navajo ($N = 146$); Hispanic ($N = 759$); Caucasian ($N = 1814$); African American ($N = 1148$); Sierra Leone ($N = 109$).

Third and most importantly, there is not a routine active quality control evaluation of the published data. The stringency of collection of SWGDAM sequence data is obviously different from that of much of the other publicly available data (for which stringency criteria often are not even mentioned). Supplementary data from other published sources typically are not available to confirm or refute possible errors, so as stated above, making changes based on phylogentic analyses solely could introduce errors in the public data portion of a database. However, evolutionary and anthropological data should be evaluated for confirming the relevance and representativeness of forensic databases, as well as for providing a framework of an established worldwide phylogeny (12).

Conclusion

A few errors have been identified in the SWGDAM mtDNA database, and the correct sequences are described so that users can correct their data sets (until an updated version is posted on the web). From a forensic perspective, the impact on upper bound frequency estimates is at worst marginal. The most likely scenario is a correct mtDNA sequence obtained from an evidence sample is searched against the SWGDAM reference database for matching sequences for assessing the rarity of the profile (11). But, because a few sequences in the database may harbor transcriptional errors, the number of true matching sequences may be less than they would be if no errors had occurred. One can gain an appreciation of the impact on the upper bound estimates of the percentage of matching database profiles, by using real samples where a sequence was entered into the database incorrectly (i.e., those displayed in Table 1) and comparing the estimates obtained from a database containing an incorrect sequence with that of a database containing the corrected sequence (Table 3). They are not substantially different, and many do not differ at all. Moreover, both before and after correction estimates are upper bound values and thus in most cases are conservative estimates. Some errors are still likely to exist in the SWGDAM data set, and they are sporadic. Thus, the impact on forensic estimates of the rarity of an evidentiary profile is likely to be similar to those displayed in Table 3, which is nominal or no difference.

Phylogenetic network analysis can be useful for identifying potential transcriptional errors and wrongly merged sequences. Verification of such potential errors must be carried out or at times errors can be introduced in to the database. We are continuing our phylogenetic review of the SWGDAM data and welcome input on any additional errors that one may find. If additional errors are found, they will be documented and disclosed. Future additions to the database will be screened with a phylogenetic approach so that the highest quality data can be presented. Because of this dynamic interactive approach for quality improvement, the SWGDAM database can be considered a reliable source of data. Lastly, we strongly recommend that the sequence profiles of other forensic databases be made publicly available, as have the SWGDAM data.

References

1. Allen M, Engstrom AS, Myers S, Handt O, Saldeen T, Von Haeseler A, et al. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J Forensic Sci 1998;43:453–64. [PubMed]
2. Ginther C, Issel-Tarver L, King MC. Identifying individuals by sequencing mitochondrial DNA from teeth. Nat Gen 1992;2:135–8.
3. Holland MM, Fisher DL, Mitchell LG, Rodriguez WC, Canik JJ, Merril CR, et al. Mitochondrial DNA sequence analysis of human skeletal remains: Identification of remains from the Vietnam War. J Forensic Sci 1993;38:542–53. [PubMed]
4. Hopgood R, Sullivan KM, Gill P. Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. Biotechniques 1992;13:82–92. [PubMed]
5. Pfeiffer H, Huhne J, Ortmann C, Waterkamp K, Brinkmann B. Mitochondrial DNA typing from human axillary, pubic and head hair shafts—success rates and sequence comparisons. Int J Leg Med 1999;112: 287–90.
6. Schneider PM, Seo Y, Rittner C. Forensic mtDNA hair analysis excludes a dog from having caused a traffic accident. Int J Leg Med 1999;112: 315–6.
7. Sullivan KM, Hopgood R, Lang B, Gill P. Automated amplification and sequencing of human mitochondrial DNA. Electrophoresis 1991;12: 17–21. [PubMed]

8. Sullivan KM, Hopgood R, Gill P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Leg Med 1992;105:83–6.
9. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Leg Med 1995;108:68–74.
10. Wilson MR, Polanskey D, Butler J, DiZinno JA, Replogle J, Budowle B. Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. BioTechniques 1995;18:662–9. [PubMed]
11. Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, et al. Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 1999;103:23–35. [PubMed]
12. Bandelt H-J, Lahermo P, Richards M, Macaulay V. Detecting errors in mtDNA data by phylogenetic analysis. Int J Leg Med 2001;115: 64–69.
13. Bandelt H-J. The fallible mtDNA databases. Preceedings of the International Symposium on Forensic DNA Technologies. Munster, Germany, September 2003.
14. Forster P. To err is human. Ann Hum Genet 2003;67:2–4. [PubMed]
15. Helgason A, Stefansson K. Erroneous claims about the impact of mitochondrial DNA sequence database errors. Am J Hum Genet 2003;73:974–5. [PubMed]
16. Herrnstadt C, Preston G, Howell N. Errors, phantom and otherwise, in human mtDNA sequences. Am J Hum Genet 2003;72:1585–6. [PubMed]
17. Koop BF, Rowan L, Chen W-Q, Deshpande P, Lee H, Hood L. Sequence length and error analysis of sequenase and automated *Taq* cycle sequencing methods. BioTechniques 1993;14(3):442–7. [PubMed]
18. Ewing B, Hiller L, Wendl MC, Green P. Base-calling of automated sequencer traces Using Phred. I. Accuracy assessment. Genome Res 1998;8:175–85. [PubMed]
19. Ewing B, Green P. Base-calling of automated sequencer traces Using Phred. II. Error probabilities. Genome Res 1998;8:186–94. [PubMed]
20. Budowle B, Allard M, Fisher CL, Isenberg AR, Monson KL, Stewart JEB, Wilson MR, Miller KWP. HVI and HVII Mitochondrial DNA population data in Apaches and Navajos. Int J Leg Med 2003;116(4):212–5.
21. Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B. The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Science Communications 4(2) April 2002. Available: http://www.fbi.gov/hq/lab/fsc/current/index.htm.

Additional information and reprint requests:
Bruce Budowle, Ph.D.
FBI Laboratory
2501 Investigation Parkway
Quantico, VA 22135
Tel: 703-632-8386
Fax: 703-632-7817
E-mail: bbudowle@fbi.gov