SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS[1]

Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories

Short Title: *MtDNA Interpretation Guidelines*

Approved and Effective: September 9, 2024

Scope

The SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories provides guidelines for the interpretation of mitochondrial DNA typing results developed by either Sanger or Next Generation Sequencing (NGS), also commonly referred to as Massively Parallel Sequencing (MPS) and supersedes the Scientific Working Group on DNA Analysis Methods (SWGDAM) Interpretation Guidelines for Mitochondrial DNA (mtDNA) Analysis by Forensic DNA Testing Laboratories (2019). Laboratories are encouraged to review their standard operating procedures and validation data in light of these guidelines and to update their procedures as needed. Refer to the Supplemental Information for the SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories (2024) for further explanation of these guidelines and background information.

---

[1] The Scientific Working Group on DNA Analysis (SWGDAM; see SWGDAM.org) is comprised of forensic science practitioners and other experts who represent government laboratories within the U.S and Canada, as well as intra- and international professional groups, and academia. SWGDAM recommends to the FBI Director revisions to the *Quality Assurance Standards for Forensic DNA Testing Laboratories* and the *Quality Assurance Standards for DNA Databasing Laboratories (QAS)*. SWGDAM also provides a forum for its members and invited guests to discuss research, technologies, techniques, and training; and conduct or recommend studies to develop, test, and validate methods for use by forensic laboratories. SWGDAM's Guidelines and Recommendations represent best practices within the discipline. The term "should" is used herein to indicate those best practices identified by SWGDAM. "Shall" distinguishes mandatory elements, which may be specified in the *Quality Assurance Standards for Forensic DNA Testing Laboratories* and/or *Quality Assurance Standards for DNA Databasing Laboratories*.

SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories
Approval/Effective Date: September 9, 2024

Table of Contents

1. Introduction…………………………..….3

2. Evaluation of Controls………………….....3

3. Sequence Analysis and Interpretation …4

4. Sequence Nomenclature………..…...….6

5. Haplotype Comparison and Reporting Results ……………………..………...….8

6. Weight of Evidence, Population Databases, and Statistics/Frequency Calculations………………………………9

7. Glossary……..……………….……….…..11

8. References..……………………..…… 13

Key Concepts:

In this document, you will find:

- Specific considerations described for the interpretation of mitochondrial sequence data developed through traditional (Sanger) or NGS processes.
- Guidelines promoting the sole use of a phylogenetic approach to mtDNA alignments through the use of EMPOP. A rule-based approach should only be employed in specific circumstances.
- Statistical considerations described for the reporting of mtDNA haplotype frequency estimations for whole genome sequence data.

1. Introduction

Mitochondrial DNA (mtDNA) sequence interpretation in forensic casework results in the generation of mtDNA profiles typically used for comparisons, frequency estimates, and entry into pertinent databases. Since the release of the initial version of the SWGDAM Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequence Interpretation (Forensic Science Communications, April 2003), quality control methods and additional resources have become more widely available that provide assistance in the evaluation of mtDNA profiles and population databases (e.g., phylogenetic assessments). Information on these resources and methods were originally detailed in the 2013 version of this document and are now maintained in this document or in the Supplemental Information document. NGS-specific guidelines were introduced in the previous revision of these guidelines (2019) and whole mtDNA genome sequencing topics are expanded upon in this revision. While both Sanger and NGS are accepted methods capable of generating mtDNA profiles, maintaining consistent mtDNA haplotype nomenclature within and among laboratories remains a challenge and is emphasized in this revision. As a result, laboratories may need to revisit reported mtDNA alignments to ensure they are consistent with these guidelines, particularly for unassociated mtDNA haplotypes residing in local, state, or national databases and included in current searches. The laboratory's interpretation guidelines and thresholds shall be based on and supported by applicable internal validation studies pursuant to the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories and as described in the SWGDAM Validation Guidelines.

2. Evaluation of Controls

2.1 For data to be of requisite quality for interpretation, quality control measures (positive, negative, and reagent blanks) shall be established to demonstrate that the testing

performed as expected. A laboratory shall verify that all control results meet the laboratory's interpretation guidelines for all reported results.

2.1.1 Evaluation criteria must be established for each control and all controls shall be processed as outlined in the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories.

2.1.2 In addition, the laboratory should have and follow procedures for evaluating the success of each step of the process.

2.1.2.1 The expected performance and/or sequence of the positive control(s) used for these purposes should be a known sample of high quality and detailed in the validation studies or other documentation.

2.2 To minimize the introduction of contamination during testing, a laboratory shall implement sample handling procedures and quality control practices designed for this purpose (see SWGDAM Contamination Prevention and Detection Guidelines for Forensic DNA Laboratories).

2.2.1 Methods shall be in place to monitor contamination within the laboratory.

2.2.2 A laboratory shall have and follow policies and/or procedures that are supported by validation studies for interpreting data potentially affected by contamination.

2.2.3 For interpretation to proceed in the event of a contaminated control, the existence of such contamination must not render the results of the corresponding sample(s) unreliable. At a minimum, haplotype(s) obtained from a contaminant observed in a reagent blank and/or negative amplification control above a laboratory's established qualitative or quantitative threshold must not be concordant with a haplotype obtained from a corresponding sample(s).

3. Sequence Analysis and Interpretation

3.1 Regions targeted for testing should be determined by individual laboratories and be based on laboratory policy (e.g., mtDNA control region, portions of mtDNA control region, entire mtDNA genome).

3.2 Due to expected differences in the mtDNA quality (e.g., the state of degradation and the quantity of the mtDNA present) between evidence samples and known samples, amplification and/or sequencing strategies specific to the sample type should be employed.

3.3 Amplicons should be sequenced in forward and reverse directions to reduce ambiguities in base determination. When overlapping ranges or reads exist between separate sequences and/or separate amplifications of the same sample, these regions should be examined carefully for sequence consistency.

3.4 The laboratory shall have and follow written guidelines for interpretation of data that are supported by validation studies. The laboratory shall establish criteria to assign nucleotide base calls and to determine whether the results are of sufficient quality for interpretation.

3.4.1 For Sanger data, the overall quality of the electropherogram data should be first assessed. Criteria that may be considered in the designation of bases include relative fluorescence units, sequence background noise, peak resolution and shape, and amplification and/or sequencing artifacts (Pont-Kingdon et al. 2012, Ellard et al. 2016).

3.4.2 For NGS data, the criteria used for filtering or trimming reads should be established. Metrics including, but not limited to, read depth, read quality scores (Q-scores), and strand bias should be considered (Pont-Kingdon et al. 2012, Gargis et al. 2012, Rehm et al. 2013, Aziz et al. 2015, Ellard et al. 2016). Additional criteria that may be considered specifically for the designation of bases include variant frequency, variant count, and variant quality.

3.5 The laboratory shall define heteroplasmy within the operational limits of the system used for sequencing and across the range of the region sequenced.

3.6 If the laboratory intends to perform mtDNA mixture interpretation, validation studies shall be performed according to the SWGDAM Validation Guidelines which fully characterize the limitations of analysis and interpretation.

3.6.1 These studies should establish the laboratory's protocol for mtDNA mixture interpretation and include experiments to address the criteria for distinguishing among mixture, heteroplasmy, damage, and noise.

3.7 The laboratory must develop interpretation guidelines to confirm and/or report length variants. These should be done with consideration of the sequencing methodology, platform, and analysis software used.

3.8 Validation studies should characterize the performance of the analysis software and its handling of repeats to establish interpretation guidelines for homopolymeric/repetitive regions.

4. Sequence Nomenclature

4.1 Until database searching is string-based, and not difference-based, mtDNA sequences should be aligned to the Revised Cambridge Reference Sequence (rCRS) using a phylogenetic approach.

4.2 The following rules should be applied:

4.2.1 Maintain phylogenetic alignments (also known as patterns of polymorphisms). Most violations to known patterns of polymorphisms involve insertions and deletions. Refer to the Supplemental Information for the SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories for specific examples.

4.2.2 Homopolymeric C-Stretches in Hypervariable Region I (HVI): C-stretches in HVI should be interpreted with a 16189C when the otherwise anchored T at position 16189 is not present. Length variations in the short A-tract preceding 16184 should be noted as transversions.

4.2.3 Homopolymeric C-Stretches in Hypervariable Region II (HVII): C-stretches in HVII should be interpreted with a 310C when the otherwise anchored T at position 310 is not present. C-stretches should be interpreted with a 311T when the anchored T at position 310 is followed by a second T.

4.2.4 Maintain the AC Repeat Motif in the HVIII region from np 515-525.

4.2.5 The 3107 nucleotide should not be reported in sample data. As 3107 in the rCRS is simply a placeholder intended to maintain historical nomenclature (Andrews et al. 1999), differences from the rCRS (e.g., deletions) at this position are not biologically meaningful.

4.3 The phylogenetic alignment should be verified using the EDNAP (European DNA Profiling Group) mtDNA Population Database (EMPOP) (https://empop.online). A laboratory should determine exceptions to verifying phylogenetic alignment in EMPOP.

4.4 If EMPOP is not available to assist with a phylogenetic alignment, and/or the software provides multiple nomenclature options for the same string of bases, the analyst should apply the rules below in numerical order to best determine the proper nomenclature for the observed sequence.

4.4.1 **Rule 1**—Use nomenclature with the least number of differences unless it violates the phylogenetic alignment.

4.4.2 **Rule 2**—Prefer substitutions to insertions/deletions.

4.4.3 **Rule 3**—Prefer transitions to transversions unless this is in conflict with the phylogenetic alignment.

4.4.4 **Rule 4**—Place indels contiguously when possible.

4.4.5 **Rule 5**—Place indels on the 3' end of the light strand.

4.5 Insertions are described by noting the site immediately prior to the insertion with respect to the light strand of the rCRS followed by a point and a '1' for the first inserted base, with sequential numbering for each inserted base thereafter (e.g., 315.1C). Insertions should not alter subsequent numbering of the sequence.

4.6 Deletions are described by noting the deleted site followed by either a dash '-' or 'del' or 'DEL,' depending on the preference of the laboratory or the requirements of the target database (e.g., 249-, 249del, or 249DEL).

4.7 The laboratory should have procedures for when haplotypes aligned using previous rules will need to be realigned using current rules. .

4.7.1 Realignment of a list of sequence differences entered into EMPOP is not considered a reinterpretation. However, reassessment of data within laboratory sequence assembly software would be considered reinterpretation and the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories standards for reinterpretation of legacy data shall be followed.

4.7.2 If applicable, haplotype frequency statistics should be recalculated.

5. Haplotype Comparison and Reporting Results

5.1 The laboratory shall define conditions under which a comparison would lead to a conclusion of inclusion, exclusion, or inconclusive result.

5.1.1 The laboratory's policy should address the use of common length variants (e.g., following positions 16193, 309, and 573) for comparisons and searches. SWGDAM does not recommend the use of such length variants for comparisons or searches at these common locations.

5.1.2 The laboratory should establish the number of differences used to make an exclusion based on published and/or internal studies. Considerations may include the sequence range of the mtDNA data reported, the variant detection threshold used, and the genetic distance between individuals (direct versus indirect).

5.1.3 The laboratory shall develop guidelines for the comparison of sequences that involve heteroplasmy.

5.1.4 The laboratory shall develop guidelines for the comparison of mixtures when applicable.

6. Weight of Evidence, Population Databases, and Statistics/Frequency Calculations

6.1 The mtDNA haplotype of a reference sample and an evidence sample that cannot be excluded as potentially originating from the same source or lineage shall be searched in a population database in order to provide a statistical weight to a reported inclusion.

6.2 The laboratory should ensure that any population database used for casework is representative of the appropriate population(s), of an appropriate size, relevant sequence range(s), and employs quality measures to assess the data entered into the database.

6.3 The laboratory should establish guidelines for the amount or range of sequence data used for searches of the population databases, particularly regarding whole mtDNA

genomes. In general, the search results that provide the highest discrimination potential are the most informative and are those typically reported. It is acceptable to perform additional searches of the population database using reduced ranges of the mtDNA sequence in an attempt to obtain the maximal discrimination potential for that combination of evidence and population database profiles. However, laboratories performing such reduced sequence range searching must ensure that matches based on a search of the shorter range are not counted if they would not be matches when searched against the longer range.

6.4 The basis for the mtDNA haplotype frequency estimation is the counting method. The application of a confidence interval accounts for database size and sampling variation.

6.4.1 The mtDNA haplotype sample frequency (p) is calculated using $p = x/n$ formula, where x is equal to the number of times the haplotype is observed in the database containing n number of haplotypes in the database.

6.4.2 A mtDNA haplotype upper bound profile probability estimate should be calculated from the observed haplotype frequencies by including an upper confidence interval (generally 95% or greater) to capture the effect of database size (Clopper and Pearson 1934).

6.4.3 A likelihood ratio may be calculated from the haplotype frequency.

6.5 If there is reasonable expectation of genetic independence, match probabilities from any combination of mtDNA, Y-STR, and/or autosomal STRs may be combined. Such an expectation could arise from large scale independence testing or strong population genetic models (Walsh et al. 2008, Buckleton and Myers 2014).

7. Glossary (for use with these Guidelines only)

Also see *Addendum to the SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories to Address Next Generation Sequencing (approved April 23, 2019).*

Damage: An alteration in the chemical structure of DNA, such as a break in a strand of DNA, a base missing from the backbone of DNA, or a chemically changed base, that can occur naturally or via environmental factors.

Heteroplasmy: The presence of more than one type of mtDNA genome within a cell or individual.

Indel: The abbreviation for insertion/deletion used to describe a location in an alignment of two sequences where an insertion or deletion of one or more bases is required to maintain alignment.

Next Generation Sequencing or NGS: The simultaneous sequencing of millions of DNA molecules that are localized onto solid substrates such as particles or flow cells. NGS is also known as massively parallel sequencing, deep sequencing, high throughput sequencing, and second-generation sequencing.

Phylogenetic alignment: The process of arranging a mtDNA haplotype in an evolutionary relationship to other haplotypes.

Quality scores (Q-scores): A metric that is used to indicate whether a base has been called correctly. Specifically, it is the probability that a given base has been miscalled. Mathematically, it is defined as $-10\log10(e)$, where e is the estimated probability of the base call being incorrect. Higher Q scores indicate a lower probability of base-calling error, while lower Q scores indicate a higher probability of error.

rCRS (revised Cambridge Reference Sequence): A corrected version of the first human mtDNA genome sequenced and published. See Anderson (1981) and Andrews (1999). Sequence data are aligned to the rCRS sequence and the collection of differences as compared to the rCRS constitute a sample's mtDNA haplotype.

Reads: The nucleotide sequence generated from a single clonally amplified DNA fragment.

Read depth: The number of reads that align at a given DNA location.

Strand bias: When performing paired-end sequencing, strand bias refers to any directional bias. Ideally, every nucleotide would have the same number of forward and reverse reads. While strand bias can, under certain circumstances, indicate reduced support for the affected nucleotide calls, in some assays, and in particular genomic regions, only one strand is routinely sequenced.

Variant count: The number of reads that contain a given variant.

Variant frequency: The percent of reads that contain a given variant.

Variant quality: Some analysis software will provide a quality score for a variant call. Some of the factors considered for this score can include frequency, strand bias, known artifact positions, read length, and number of variants within the same read.

8. References

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G. (1981) *Sequence and organization of the human mitochondrial genome*. Nature. 290: 457-465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999) *Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA*. Nature Genetics 23:147.

Aziz, N., Zhao, Q., Bry, L., Driscoll, D., Funke, B., Gibson, J., Grody, W., Hegde, M., Hoeltge, G., Leonard, D., Merker, J., Nagarajan, R., Palicki, L., Robetorye, R., Schrijver, I., Weck, K., Voelkerding, K. (2015) *College of American Pathologists' laboratory standards for next-generation sequencing clinical tests*. Arch Pathol Lab Med. 139:481-493.

Buckleton, J., Myers, S. (2014) *Combining autosomal and Y chromosome match probabilities using coalescent theory*. FSI Genetics. 11:52-55.

Clopper, C.J. and Pearson, E.S. (1934) *The use of confidence or fiducial limits illustrated in the case of the Binomial*. Biometrika. 26:404-413.

Ellard, S., Charlton, R., Yau, S., Gokhale, D., Taylor, G., Wallace, A., Ramsden, S., Berry, I. (2016) *Practice guidelines for Sanger sequencing analysis and interpretation*. Association for Clinical Genetic Science; available at https://www.researchgate.net/publication/237629350_Practice_Guidelines_for_Sanger_Sequencing_Analysis_and_Interpretation.

Gargis, A.S., Kalman, L., Berry, M.W., Bick, D.P., Dimmock, D.P., Hambuch, T., et al. (2012) *Assuring the quality of next-generation sequencing in clinical laboratory practice*. Nature Biotechnology. 30 (11):1033-1036.

Pont-Kingdon, G., Gedge, F., Wooderchak-Donahue, W., Schrijver, I., Weck, K., Kant, J., Oglesbee, D., Bayrak-Toydemir, P., Lyon, E. (2012) *Design and analytical validation of clinical DNA sequencing assays*. Arch Pathol Lab Med. 136:41–46.

Rehm, H., Bale, S., Bayrak-Toydemir, P., Berg, J., Brown, K., Deignan, J., Friez, M., Funke, B., Hegde, M., Lyon, E. for the Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. (2013) *ACMG clinical laboratory standards for next-generation sequencing*. Genet Med. 15:733-747.

Scientific Working Group on DNA Analysis Methods (SWGDAM). (2003*) Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequence Interpretation*. Forensic Science Communications [Online]; available at https://www.swgdam.org/publications

Scientific Working Group on DNA Analysis Methods (SWGDAM). (2017) *Contamination Prevention and Detection Guidelines for Forensic DNA Laboratories.* [Online]. Available at https://www.swgdam.org/publications

Scientific Working Group on DNA Analysis Methods (SWGDAM). (2024) *Supplemental Information for the SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories.* [Online]. Available at ahttps://www.swgdam.org/publications

Walsh, B., Redd, A.J., Hammer, M.F. (2008) *Joint match probabilities for Y chromosomal and autosomal markers*. Forensic Sci. Int. 174:234-238.

Document Version	Revision History
2003	Original entitled Scientific Working Group on DNA Analysis Methods (SWGDAM) Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequencing Interpretation. (Published in Forensic Science Communications in April 2003 (Vol. 5, No. 2).
July, 2013	The document was restructured and substantially revised for consistency with current best practices in mtDNA interpretation, particularly with respect to sequence nomenclature and frequency calculations. The revisions were approved by the SWGDAM membership on July 18, 2013.
February, 2015	Example 3 on page 10 was revised by removing all references to 16519 and changing 16174C to 16174A (rev1). The revisions were approved by the SWGDAM Executive Board on March 6, 2015.
2018	Revisions were made to the following sections to address next generation sequencing: Introduction, Sections 1, 2.1, 2.2.1, 2.2.4, Rule 9, Sections 2.3.4, 3.1.2, 4.2.4.1, and 4.2.5. The revisions were approved by the SWGDAM membership on April 23, 2019.
2024	This document contains several revisions to the 2018 guidelines and is intended to reflect the current state of forensic mitochondrial DNA typing, including the use of next generation sequencing (NGS), whole mitochondrial DNA genome sequencing (mtDNA), and statistical analysis. In addition, some information was removed from this document and transferred to the Supplemental Information for the SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories document.