Scientific Working Group on DNA Analysis Methods

Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories

Contents

SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories

The Scientific Working Group on DNA Analysis Methods, better known by its acronym of SWGDAM, is a group of approximately 50 scientists representing Federal, State, and Local forensic DNA laboratories in the United States and Canada. During meetings, which are held twice a year, Committees discuss topics of interest to the forensic DNA community and often develop documents to provide direction and guidance for the community. This document was presented to SWGDAM and received approval on July 18, 2013.

This document provides guidelines for the interpretation of mitochondrial DNA typing results and supersedes the Scientific Working Group on DNA Analysis Methods (SWGDAM) Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequencing Interpretation (2003). The revised guidelines are not intended to be applied retroactively. Laboratories are encouraged to

review their standard operating procedures and validation data in light of these guidelines and to update their procedures as needed. It should be noted that the recommendations pertaining to sequence nomenclature and frequency calculations have undergone significant modification. These guidelines are applicable to mtDNA data generated by current Sanger-based sequencing methods. It is anticipated that these guidelines will evolve further as future technologies emerge.

Introduction

Mitochondrial DNA (mtDNA) sequence interpretation in forensic casework is an essential stage of analysis that 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),* newly implemented 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). Additionally, over the past several years, as mtDNA forensic casework and research has progressed, laboratory experience and knowledge has continued to grow. Therefore, SWGDAM has revised the interpretation guidelines to more accurately reflect the current state of the field and to better serve forensic mtDNA practitioners. Development and implementation of written guidelines for the interpretation of analytical results for each individual laboratory still remains necessary and must be based on validation studies conducted pursuant to the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories, with consideration given to the SWGDAM Validation Guidelines.

1. Evaluation of controls

For data to be of requisite quality for interpretation, measures should be established to demonstrate that the testing performed as expected. The use of controls is among the most important quality measures for DNA testing. Controls shall include, at a minimum, a positive amplification control, a negative amplification control, and a reagent blank control. Evaluation

criteria must be established for each control and all controls shall be treated the same as, and parallel to the forensic and/or casework sample(s) being analyzed, as outlined in the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories (QAS).

Reagent blanks and negative amplification controls are used to monitor levels of contamination and also assist in identifying at which step of the process contamination may have been introduced. Contamination is the unintentional introduction of exogenous DNA into a DNA sample or PCR reaction. Reagent blanks monitor contamination from extraction to final sequence analysis. Negative amplification controls monitor contamination from amplification to final sequence analysis. Reagent blanks and negative amplification controls that contain DNA must be assessed to ensure that any results for the corresponding sample(s) are legitimately from the samples and not due to contamination. If contamination in the reagent blank and/or negative amplification control is present above laboratory established acceptance parameters, then the data cannot be used for interpretative purposes.

A positive amplification control is a sample of known mtDNA sequence used to monitor the success of the analysis. The positive amplification control shall be processed starting at amplification (DNA purified from the HL60 cell line is currently required as a positive amplification control for inclusion of mtDNA forensic data into the National DNA Index System [NDIS]).

To minimize the introduction of contamination during testing, a laboratory shall implement sample handling procedures and quality control practices designed for this purpose. Methods shall be in place to monitor contamination within the laboratory. A laboratory shall verify that all control results meet the laboratory's interpretation guidelines for all reported results. In addition, a laboratory shall have and follow policies and/or procedures that are supported by validation studies for interpreting data potentially affected by contamination.

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, a haplotype(s) obtained from a contaminant observed in a reagent blank and/or negative amplification control must not be concordant with a haplotype obtained from a corresponding sample(s). In addition to this minimum criterion, a laboratory may choose to assign a maximum threshold based on the amount of contaminant DNA observed (e.g. *Wilson et al. 1995b*). All such thresholds shall be supported by internal validation studies. Depending on the extent of the observed contamination, reanalysis may be initiated, starting with the appropriate step of the process. If reanalysis is not possible, the reliability of the results from the affected sample(s) should be critically assessed for use. Furthermore, access to the DNA profiles of laboratory personnel is helpful when attempting to trace potential sources of contamination.

2. Sequence Analysis and Nomenclature

2.1 Sequencing Overview

The regions of the mtDNA genome that are typically targeted for evidentiary testing are hypervariable region 1 (HV1- positions 16024-16365) and hypervariable region 2 (HV2 - positions 73-340) located within the control region. Due to expected differences in the mtDNA quality (i.e. the state of degradation and the quantity of the mtDNA present) between evidence samples and known samples, amplification strategies specific to the sample type should be employed. Forward and reverse strands of the amplified product are sequenced to reduce ambiguities in base determination. When overlapping ranges exist between separate amplifications of the same sample, these regions should be examined carefully for sequence consistency. Sequence differences in overlapping regions, as well as discrepancies in expected amplicon quantities, could indicate primer binding issues and should be interpreted with caution. It is important for each laboratory to determine sample sequence coverage and review requirements for interpretation. Additional sequencing approaches should be implemented for samples containing common homopolymeric regions such as those that may occur in the HV1 Cytosine-stretch (C-stretch) region (between positions 16183-16194) and the HV2 C-stretch region (between positions 302-310) as these motifs can prove to be challenging. Suggested primers and processing strategies are described in the

literature (for example, *Wilson et al. 1995a, Wilson et al. 1995b, Gabriel et al. 2001, Edson et al. 2004, Berger et al. 2008,* and *Eichmann et al. 2008).*

2.2 Sequence Analysis

2.2.1 Criteria

A laboratory shall have and follow written guidelines for the interpretation of data that are supported through its validation. The laboratory should establish criteria to assign nucleotide base calls to appropriate peaks and to determine whether the results are of sufficient quality for interpretation purposes. To accomplish this, the overall quality of the electropherogram data should be first assessed by importing all relevant sequence data into a software program for viewing. An alignment of overlapping sequences is then performed using software programs specifically designed for this purpose. The heavy strand sequences are reverse-complemented so that the bases are aligned in the light strand orientation. Strands are compared and bases designated.

2.2.2 DNA Base Call Designations

DNA bases are designated by the nomenclature system set forth by the International Union of Pure and Applied Chemistry (IUPAC). At confirmed positions of ambiguity, the following IUPAC codes should be used:

G/T = K	A/C = M
A/G = R	A/G/T = D
G/C = S	A/C/T = H
A/T = W	A/C/G = V

C/T = Y C/T/G = B

A/C/G/T = N

Base call designations for insertions and deletions are described in section 2.3.3.

2.2.3 Heteroplasmy

Heteroplasmy is defined as more than one mtDNA sequence present in an individual. Detectable heteroplasmy can be observed as point heteroplasmy where two DNA bases are observed at the same nucleotide position. Heteroplasmy can also be seen as length heteroplasmy, which typically is observed as overlapping sequences with a variation in the number of bases in a homopolymeric stretch of bases (i.e., C-stretch). Each laboratory should define heteroplasmy within the operational limits of the system used for sequencing. When the specimens under consideration differ by a single nucleotide, additional samples may be run in order to attempt to resolve the difference.

2.2.4 Mixtures

Mitochondrial DNA mixed sequences are not commonly interpreted. Regardless, a laboratory shall have and follow documented procedures for mixture interpretation supported by its validation.

2.3 Sequence Nomenclature

Since the resulting mtDNA sequence is a long string of letters (representing the DNA bases) that can theoretically differ at any position along this sequence, a shorthand method of naming the sequences is commonly used. The use of standardized nomenclature principles to determine the mtDNA sequence allows for the consistent representation of a sample's haplotype. However, experience has demonstrated that the same nomenclature principles have not always been employed by laboratories. In addition, for some sequences, consistent

application of standardized nomenclature principles has proven difficult to achieve manually. In such situations, differences in the representation of the same sequence string could result in a false exclusion in a direct comparison or database search. Modification of search algorithms used for sequence comparisons to include string-based capability can resolve this potential issue.

2.3.1 Use of a Sequence Reference Standard

A consensus sequence obtained from the sample is compared to the revised Cambridge Reference Sequence (rCRS) described by *Andrews et al. (1999)*. Differences between the rCRS and the sample sequence will be recorded as polymorphisms with both the nucleotide position and the DNA base difference from the reference noted (e.g., 16089 C). This process derives the mtDNA shorthand used to record a sample's haplotype.

2.3.2 Applied Nomenclature (i.e. the use of Nomenclature Rules)

There are different mtDNA nomenclature approaches to derive a sample's haplotype (*Wilson et al. 2002a, Wilson et al. 2002b, Bandelt and Parson 2008,* and *Budowle et al. 2010*). These methods employed either a hierarchical series of rules or a phylogenetic approach. A comparison of the rule based approach and the phylogenetic approach showed that generally both systems code the haplotypes in the same manner even though they use different strategies (*Polanskey et al. 2010*). However, due to the inherent differences between the approaches, the potential exists for the same sequence to be annotated differently between laboratories, particularly when mtDNA types have atypical insertions and deletions (see the SWGDAM mtDNA Nomenclature Examples document). Furthermore, there are mtDNA sequences whose base compositions (sequence strings) truly differ by only one base. Yet when these sequences are evaluated using a rule based nomenclature, the results may yield mtDNA haplotypes which appear to differ by more than the one true base (Table I, Example 1).

Nomenclature differences like these may not be a problem with direct one-to-one comparison of samples within the same laboratory. However, it is problematic when performing forensic database searches for missing person cases using mtDNA data. In these situations, the database comparison between these samples would result in a missed association since mtDNA database searches using mtDNA sequences for missing person cases account for the possibility of *only a single mutational event* between generations (Table I, Example 2). Ideally, full sequence strings would be aligned for these database searches, making any subtle differences in the coded nomenclature irrelevant. However, the infrastructure for database string searches is not yet in place for forensic (Missing Person) databases in the United States. Until string searches are possible, a nomenclature system that is easy to apply and for which readily-available tools exist is vital.

It is important that no matter which rules are applied, efforts are made to maintain known patterns of polymorphisms in mtDNA analysis. When rules alter known patterns (i.e. established phylogenetic patterns of polymorphisms), it is possible that two mtDNA haplotypes will appear to differ at two or more sites when they actually only differ at one. For example, a rather common deletion at nucleotide position 249 is present in existing populations. When a polymorphism at position 247 is coupled with the 249 deletion, a rule based approach would code this area as a 247DEL instead of 247A, 249DEL. By failing to maintain the phylogenetically established 249DEL, sequences coded as a 247DEL are now 2 differences away from a sequence containing only 249DEL. On the contrary, if the known pattern of 249DEL is maintained, a sequence coded as 247A, 249DEL is only one difference away from a sequence containing only a 249DEL (Table I, Example 3).

Table I – Examples of Nomenclature Issues

Example 1: nucleotide position (np) 16024-16365, 73-340

Rule Based	SWGDAM Guidelines	EMPOP release 8, California Admixed Sequence
16182DEL 16183C 16193.1 C 16217C 73G 263G 309.1C 315.1C	16182C 16183C 16188T 16189C 16217C 73G 263G 309.1C 315.1C	16182C 16183C 16188T 16189C 16217C 73G 263G 309.1C 309.2C 315.1C

In this example the rule based and SWGDAM haplotypes represent the same nucleotide sequence string, yet are coded 3 differences away from each other (ignoring the 16193.1C). When queried in EMPOP (string-based) – an exact match is found. The rule based haplotype compared to the EMPOP haplotype results in the same 3 differences, where the SWGDAM haplotype compared to the EMPOP haplotype results in 0 differences (ignoring length differences at 309).

Example 2: np 16024-16365, 73-340

Rule Based	SWGDAM Guidelines	EMPOP release 8, Texas Native American Sequence with one difference
16111T 16189C 16192.1T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 315.1C	16111T 16189C 16191.1C 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 315.1C	16111T 16189C 16192T 16223T 16233G 16290T 16319A 16331G 73G 146C 153G 235G 263G 315.1C

In this example the rule based and SWGDAM haplotypes represent the same nucleotide sequence string, yet are coded 3 differences away from each other. When queried in EMPOP (string-based) – an exact match is not found, but a sequence is returned with one base difference. The rule based haplotype compared to the EMPOP haplotype results in two differences, where the SWGDAM haplotype compared to the EMPOP haplotype results in the appropriate one difference. In a missing person database search, the rule based haplotype and the EMPOP haplotype would not be considered as a potential match (tolerating a single mutation between the haplotype). Please note that the EMPOP haplotype would be coded the same using either the rule based approach (*Budowle et al. 2010*) or the SWGDAM approach.

Example 3: np 16024-16365, 73-340		
Rule Based	SWGDAM Guidelines	EMPOP release 8, Colorado African American Sequence with one difference
16129A 16172C 16184T 16187T 16189C 16223T 16261T 16278T 16290T 16293G 16311C 16360T 16519C 73G 150T 151T 152C 182T 186A 189C 195C 247DEL 263G 297G 315.1C 316A	16129A 16172C 16184T 16187T 16189C 16223T 16261T 16278T 16290T 16293G 16311C 16360T 16519C 73G 150T 151T 152C 182T 186A 189C 195C 247A 249DEL 263G 297G 315.1C 316A	16129A 16172C 16174C 16184T 16187T 16189C 16223T 16261T 16278T 16290T 16293G 16311C 16360T 16519C 73G 150T 151T 152C 182T 186A 189C 195C 247A 249DEL 263G 297G 315.1C 316A
In this example the rule based and SWGDAM haplotypes represent the same nucleotide sequence string, yet are 3 differences away from each other. When queried in EMPOP (string-based) – an exact match is not found, but a sequence is returned with one base difference (16174C). The rule based haplotype compared to the EMPOP haplotype results in four differences, where the SWGDAM haplotype compared to the EMPOP haplotype results in one difference at 16174. In a missing person database search, the rule based haplotype and the EMPOP haplotype would not be considered as a potential match (tolerating a single mutation between the haplotypes).		

Given the issues highlighted above, it is recommended that a blended application of rule based and phylogenetic approaches be used. This blended approach allows for the forensic scientist to recognize and maintain known patterns of polymorphisms. For those rare sequences containing peculiar insertions and deletions, it is recommended that the forensic analyst use the EDNAP mtDNA Population Database (EMPOP) (http://empop.org/) to help determine a consistent mtDNA haplotype for entry into forensic databases. It is noted that these new rules (noted below) do not affect previous forensic case comparisons where a forensic sample was directly compared to a submitted known reference sample. However, caution should be exercised when comparing newly interpreted samples to those interpreted using former approaches.

2.3.3 SWGDAM Nomenclature Rules

Variants from the rCRS should be coded in accordance with the following nomenclature rules:

Rule 1 – Maintain known patterns of polymorphisms (a.k.a. known phylogenetic alignments). Most violations to known patterns of polymorphisms involve insertions and deletions.

Example: Maintain deletions at positions 249, 290 and/or 291 when present. See other examples in the SWGDAM mtDNA Nomenclature Examples document.

Rule 2 - Use nomenclature with the least number of differences unless it violates known patterns of polymorphisms.

Rule 3a – Homopolymeric C-Stretches in Hypervariable Region I (HVI): C-stretches in HV1 should be interpreted with a 16189C when the otherwise anchored T at position 16189 is not present. Length variation in the short A-tract preceding 16184 should be noted as transversions.

Rule 3b - Homopolymeric C-Stretches in Hypervariable Region II (HVII): C-stretches in HV2 should be interpreted with a 310 C when the otherwise anchored T at position 310 is not present.

Rule 4 – Maintain the AC Repeat Motif in the HVIII region from np 515-525.

Rule 5 – Prefer substitutions to insertions/deletions (indels).

Rule 6 – Prefer transitions to transversions unless this is in conflict with Rule 1.

Rule 7 – Place indels contiguously when possible.

Rule 8 - Place indels on the 3' end of the light strand.

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. Deletions are described by noting the deleted site followed by either a dash '-' or 'del', depending on the preference of the laboratory or the requirements of the target database.

EMPOP queries are performed with alignment-free nucleotide sequence strings. In the output, however, the haplotype is presented in difference coded format using the phylogenetic alignment. Thus the mtDNA analyst can use the presented sequences that are zero differences away from the searched sequence to confirm that they have properly called the haplotype using the rules stated above. If there are no consistent (exact) sequences in EMPOP, analysts can evaluate those sequences that differ by one or more bases to determine the proper sequence notation. Analysts should also use the maximum likelihood calculation in EMPOP to help determine the most appropriate haplotype nomenclature for the sample.

Limited sequences obtained from degraded samples may not yield enough sequential information to identify and maintain known patterns of polymorphisms when applying the nomenclature rules. With a limited sequence and the inability to apply Rule 1, the analyst should apply the remaining rules to best determine the proper nomenclature for the observed sequence. The analysts can still use EMPOP to assess the proper sequence determination. However if the sequence information is so limited in range that the software provides multiple nomenclature options for the same string of bases, the analyst should defer to the haplotype which is consistent with the remaining rules (Rule 2 through Rule 8).

It is noted that the mtDNA region between nucleotide positions 50 – 70 is known to be highly variable. For insertions and deletions within this region, EMPOP should be consulted to determine the haplotype notation. If no example exists within EMPOP, the analyst should defer to Rule 2 to strive for the least number of differences.

2.3.4 Length Variants

Homopolymeric tracts are prone to exhibiting length heteroplasmy. Homopolymeric tracts can differ in length within the same individual and/or maternal lineage. In most cases, no attempt will be made to determine the exact number of bases in an HV1 C-stretch; however, laboratories must develop their own interpretation guidelines for HV2 length variants. A length variant alone cannot be used to support an interpretation of exclusion (*Stewart et al. 2001*).

2.3.5 Database Searching

Whether the developed mtDNA haplotype is being used to conduct a forensic database search for an association (e.g. missing person investigation) or a population database search to determine the haplotype frequency (see Section 4. Weight of Evidence), the search is performed on the basis of the determined haplotype. Regardless of the nomenclature rules applied, certain unusual mtDNA types that generally involve atypical insertions and deletions may be difficult to represent consistently. By converting mtDNA haplotype results to alignment-free nucleotide sequence strings, samples can be compared within forensic or population databases without the concern of interpretation differences resulting in missed associations. The publication by *Röck et al. (2010)* demonstrates that the application of a string-based search algorithm ensures that identical sequences are associated in a database query. Future versions of forensic and population databases used for the comparison of mtDNA profiles should incorporate this functionality.

3. Sequence Comparison and Reporting Results

Each laboratory must define conditions under which the sequence data obtained would lead to the conclusion that an individual can or cannot be eliminated as a possible source of the mtDNA evidence in a case. This may be accomplished by an examination of the number, position, and nucleotide composition of polymorphic sites.

3.1 Sequence Comparisons

Generally, the following guidelines are used:

3.1.1 Exclusion: If samples differ at two or more nucleotide positions (excluding length heteroplasmy), they can be excluded as coming from the same source or maternal lineage.

3.1.2 Inconclusive: The comparison should be reported as inconclusive if samples differ:

a. at a single position only (whether or not they share a common length variant between positions 302-310)
b. only by not sharing a common length variant between positions 302-310 (all other positions are concordant)

3.1.3 Cannot Exclude: If samples have the same sequence, or are concordant (sharing a common DNA base at every nucleotide position, including common length variants), they cannot be excluded as coming from the same source or maternal lineage.

3.2 Additions and/or Modifications

Laboratories should develop guidelines for the evaluation of cases that involve heteroplasmy. The guidelines stated above may need to be modified by a laboratory to allow for increased mutational events in cases involving a closed population (e.g. a plane crash) or where the reference samples are from distant maternal relatives of the individual

of interest. The guidelines may also need to be modified when the sequences compared extend beyond the current standard range of HV1 and HV2 as intra-individual variation has not yet been fully established for these regions (e.g. AC repeat region).

4. Weight of Evidence

The mtDNA profile of a reference sample and an evidence sample that cannot be excluded as potentially originating from the same source should be searched in a population database in order to provide a statistical weight to a reported result of "cannot exclude". By searching an appropriate and established population database, a point estimate of the true population frequency for a haplotype can be obtained (i.e., the counting method). Scientifically, it is also valuable to provide a confidence interval for the population frequency estimate.

4.1 Population Databases

Currently, the SWGDAM Mitochondrial DNA Population Database available to CODIS users is an appropriate database for assessing the relative frequency of mtDNA haplotypes within the United States using rule based nomenclature practices.[1] The EMPOP database is available to both public and private forensic laboratories, utilizes the applied nomenclature of Section 2.3.3 and contains representative populations for the United States. In either case, laboratories should ensure that any population database used for casework is representative of the appropriate population(s), of an appropriate size, and employs quality measures to assess the data entered into the database (e.g. limited ambiguities, minimum range of HV1 and HV2, phylogenetic assessment).

4.2 Frequency Calculations

[1] The forthcoming CODIS 7.0 Service Pack 3 upgrade will contain the SWGDAM Mitochondrial DNA Population Data that has been developed to incorporate the updated nomenclature rules noted in Section 2.3.3. This data set contains Whole Control Region (WCR) data for over 10,000 sequences, and was created in collaboration between the Armed Forces DNA Identification Laboratory (AFDIL), EMPOP and the FBI Laboratory.

4.2.1. 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.

4.2.2. The mtDNA haplotype sample frequency (p) is calculated using the p = x/N formula, where x is equal to the number of times the haplotype is observed in a database containing N number of haplotypes. For example, if a haplotype has been observed twice in a database of N = 2000, the frequency of that haplotype will be: 2/2000 = 0.001.

4.2.3. Reporting an mtDNA haplotype sample frequency without a confidence interval is acceptable as a factual statement regarding observations in the database.

4.2.4. In order to provide an upper confidence limit for the frequency of the mtDNA haplotype in the population, the method described by *Clopper and Pearson (1934)* should be used. This uses the binomial distribution for the probabilities of counts, including zero or other small numbers that are found for mtDNA haplotypes. If the database has n haplotypes and *x* of the haplotype of interest are found, then the required upper confidence limit p_0 is the solution to the equation

Eq. 1

$$\sum_{k=0}^{x} \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha$$

Here α gives the level of confidence: α = 0.05 gives a 95% confidence limit. The equation finds the value p_0 of the population proportion p for which the cumulative probability 0, 1, . . . *x* copies of the profile is equal to α. This equation will require a computer to solve. A special case of the result in equation 1 is when the haplotype of interest is not seen in the database, and *x* = 0. The equation

now has only one term in the sum on the left hand side: $(1 - p_0)^n = \alpha$. The solution is

Eq. 2

$$p_0 = 1 - \alpha^{1/n}$$

and for a 95% confidence limit this is very close to 3/n. (If n = 2000 the exact 95% upper confidence limit is 0.0014967, whereas 3/2000 is 0.0015.)

4.2.5. Typical Clopper and Pearson upper confidence interval p_0 values at $\alpha = 0.05$ for generic *n* and *x* values utilizing Equation 1 are provided below as examples.

x	*n* = 500	*n* = 1000	*n* = 1,500	*n* = 2,000	*n* = 3000
0	0.00597	0.00299	0.002	0.0015	0.001
1	0.00945	0.00473	0.00316	0.00237	0.00158
2	0.01254	0.00628	0.00419	0.00314	0.0021
3	0.01543	0.00774	0.00516	0.00387	0.00258
4	0.01821	0.00913	0.00609	0.00457	0.00305
5	0.02091	0.01048	0.007	0.00525	0.0035

4.3. Population Substructure

It is recognized that population substructure exists for mtDNA haplotypes. However, determination of an appropriate theta (θ) value is complicated by the variety of primer sets, covering different portions of HV1 and/or HV2, which may be applied to forensic casework. SWGDAM has not yet reached consensus on the appropriate statistical approach to estimating θ for mtDNA comparisons.

4.4. Combining Statistics for mtDNA and Autosomal Results

The frequency estimates for autosomal and mtDNA typing results obtained for a given sample may be combined. There are examples of dependencies between autosomal and mtDNA profiles, the extent of which has been demonstrated to be small (*Roby et al. 2009*).

4.4.1. A haplotype frequency estimation for the mtDNA profile is to be multiplied by the autosomal profile frequency as defined by National Research Council (1996) Recommendation 4.1.

4.5 Combining Statistics for mtDNA and Y-STR Results

The CODIS software generates a combined likelihood ratio for autosomal, mtDNA and Y-STR results for missing person searches to rank potential candidates. *Ge et al. (2010)* and *Roby et al. (2009)* provide evidence of statistical independence between mtDNA and Y-STR profiles in U.S. and Chilean populations. Prior to reporting combined statistics for mtDNA and Y-STR results, the laboratory issuing the report should determine that each population used demonstrates independence between the mtDNA and Y-STR results. If independence cannot be determined between the mtDNA and Y-STR results for the referenced population(s), combining these systems is not recommended.

5. References and Literature

Amorim A. A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. Forensic Science International: Genetics (2008) 2:376–378.

Anderson, S., Bankier, A.T., Barrell, B.G., deBruijin, 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., and Young, I.G. Sequence and organization of the human mitochondrial genome. Nature (1981) 290:457-465.

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

Ayadi, I., Mahfoudh-Lahiani, N., Makni, H., Ammar-Keskes, L., and Reba, A. Combining autosomal and Y-Chromosomal Short Tandem Repeat Data in paternity Testing with Male Child: Methods and Application. Journal of Forensic Science (2007) 52(5):1068-72.

Bandelt, H.J. and Parson, W. Consistent treatment of length variants in the human mtDNA control region: a reappraisal. International Journal of Legal Medicine (2008) 122:11-21.

Berger, C. and Parson W. Mini-midi-mito: Adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. Forensic Science International: Genetics (2009) 3:149–153.

Buckleton, J.S., Krawczak, M., and Weir, B.S. The Interpretation of Lineage Markers in Forensic DNA Testing. Forensic Science International: Genetics (2011) 5(2):78-83.

Budowle, B., Ge, J., Aranda, X.G., Planz, J.V., Eisenberg, A.J., Chakraborty, R. Texas Population substructure and Its Impact on Estimating the Rarity of Y STR Haplotypes from DNA Evidence. Journal of Forensic Science (2009) 54(5): 1016-21.

Budowle, B., Polanskey, D., Fisher, C., Den Hartog, B., Kepler, R., and Elling, J. Automated Alignment and Nomenclature for Consistent Treatment of Polymorphisms in the Human Mitochondrial DNA Control Region, Journal of Forensic Science (2010) 55(5):1190-1195.

Budowle, B., Wilson, M.R., DiZinno, J.A., Stauffer, C., Fasano, M.A., Holland, M.M., and Monson, K.L. Mitochondrial DNA regions HV1 and HV2 population data. Forensic Science International (1999) 103:23-35.

Carracedo, A., Bar, W., Lincoln, P., Mayr, W., Morling, N., Olaisen, B., Schneider, P., Budowle, B., Brinkman, B., Gill, P., Holland, M., Tully, G., and Wilson, M. DNA Commission of the International Society for Forensic Genetics: Guidelines for mitochondrial DNA typing. Forensic Science International (2000) 110:79-85.

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

Eichmann, C. and Parson, W., 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. International Journal of Legal Medicine (2008) 122(5):385-8.

Edson, S., Ross, J., Coble, M., Parson, T., Barritt, S. Naming the dead – confronting the realities of rapid identification of degraded skeletal remains. Forensic Science Review (2004)16(1):64-90

Gabriel, M., Huffine, E., Ryan, J., Holland, M., Parsons, T. Improved mtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. Journal of Forensic Science (2001) 46(2): 247-253.

Ge, J., Chakraborty, R., and Budowle, B. Test of independence in contingency tables of large dimension with ordered categories and its application in population genetics. Abstract Volume of the American Society of Human Genetics 60th Annual Meeting, Washington DC (November 2-6, 2010; Abstract # 3045/F).

Gjertson, D.W., Brenner, C.H., Baur, M.P., Carracedo, A., Guidet, F., Lugue, J.A., Lessig, R., Mayr, W.R., Pascali, V.L., Prinz, M., Schneider, P.M., Morling, N. ISFG: Recommendations on biostatistics in paternity testing. Forensic Science International: Genetics (2007) 1(3-4):223-31.

Holland, M.M. and Parsons, T.J. Mitochondrial DNA sequence analysis: Validation and use for forensic casework. Forensic Science Review (1999) 11:22-50.

Miller, K.W.P. and Budowle, B. A compendium of human mitochondrial DNA control region sequences: Development of an international standard forensic database. Croatian Medical Journal (2001) 42:315-327.

Monson, K.L., Miller, K.W.P., Wilson, M.R., DiZinno, J.A., and Budowle, B. The mtDNA population database: An integrated software and database resource. Forensic Science Communications (2002) [Online]. Available: http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/april2002/miller1.htm

Parson, W. and Dur, A. EMPOP- a forensic mtDNA database. Forensic Sci Int Genet (2007) 1(2):88-92.

Polanskey, D., Den Hartog, B.K., Elling, J.W., Fisher, C.L., Kepler, R.B., Budowle, B. Comparison of Mitotyper Rules and Phylogenetic-based mtDNA Nomenclature Systems. Journal of Forensic Science (2010) 55(5):1184-1189.

Roby, R.K., Gonzalez, S.D., Phillips, N.R., Planz, J.V., Thomas, J.L., Pantoza-Astudillo, J.A., Ge, J., Morales, E.A., Eisenberg, A.J., Chakraborty, R., Bustos, P., and Budowle, B. Autosomal STR allele frequencies and Y-STR and mtDNA haplotypes in Chilean sample populations. Forensic Science International Genetics, Supplemental Series (2009) 2:533-534.

Röck, A., Irwin, J., Dür, A., Parsons, T., Parson, W. SAM: String-based sequence search algorithm for mitochondrial DNA database queries. Forensic Science International: Genetics (2011) 5(2):126-132.

Sanger, F. and Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol (1975) 94(3):441-8.

Scientific Working Group on DNA Analysis Methods (SWGDAM). Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequence Interpretation. Forensic Science Communications (2003) [Online]. Available: http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/april2003/swgdammitodna.htm/

Scientific Working Group on DNA Analysis Methods (SWGDAM). Revised Validation Guidelines, Forensic Science Communications (2004) [Online]. Available: http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2004/standards/2004_03_standards02.htm/

Stewart, J.E.B., Fisher, C.L., Aagaard, P.J., Wilson, M.R., Isenberg, A.R., Polanskey, D., Pokorak, E., DiZinno, J.A., and Budowle, B. Length variation in HV2 of the human mitochondrial DNA control region. Journal of Forensic Sciences (2001) 46:862-870.

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

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

Wilson, M.R., DiZinno, J.A., Polanskey, D., Replogle, J., and Budowle, B. Validation of mitochondrial DNA sequencing for forensic casework analysis. International Journal of Legal Medicine (1995a) 108:68-74.

Wilson, M.R., Polanskey, D., Butler, J., DiZinno, J.A., Replogle, J., and Budowle, B. Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. Biotechniques (1995b) 18:662-669.

Wilson, M.R., Allard, M.W., Monson, K., Miller, K.W.P., and Budowle, B. Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. Forensic Science International (2002a) 129:35-42.

Wilson, M.R., Allard, M.W., Monson, K., Miller, K.W.P., and Budowle, B. Further discussion of the consistent treatment of length variants in the human mitochondrial DNA control region, Forensic Science Communications (2002b) [Online]. Available: www.fbi.gov/about-us/lab/forensic-science-communications/fsc/oct2002/wilson.htm.

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); available at http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/april2003/index.htm/swgdammitodna.htm
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.