

PROTOCOLS FOR FORENSIC MITOCHONDRIAL DNA ANALYSIS

Sequence Nomenclature and Alignment		
Status:Published		Document ID: 1168
DATE EFFECTIVE 09/15/2017	APPROVED BY mtDNA Technial Leader	PAGE 1 OF 3

Sequence Nomenclature and Alignment

- 1.1 Nucleotide positions are designated according to the standard one-letter code based on the nomenclature system adopted by the International Union of Pure and Applied Chemistry (IUPAC; see table below). Note that an “N” is used to denote unresolved sequence ambiguities where N can be any one of the four bases. **IUPAC codes that designate two possible bases should only be used in instances of sequence heteroplasmy.**

IUPAC code	Base designation	IUPAC code	Base designation
G	Guanine	R	A or G
A	Adenine	Y	C or T
T	Thymine	K	G or T
C	Cytosine	M	A or C
N	G, A, T, or C	S	C or G
		W	A or T

2 Using Sequencher 4.9

- 2.1 Sequence differences between the questioned sample and the revised Cambridge Reference Sequence (rCRS) are generated and printed out from the Comparison Report file in Sequencher. These differences are organized by hypervariable region (eg., one difference review file is generated for each HVI and HVII region). The differences are listed in order of occurrence on the mtDNA molecule.
- 2.2 In most cases, the alignment of a given mtDNA sequence with that of rCRS is straightforward. However, care must be taken in the placement of insertions and deletions in reference to that of rCRS. The following rules are taken from the SWGDAM Interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories (SWGDAM, 2013). They represent a blended approach of rule based and phylogenetic methods. It is recommended that the forensic analyst use the EDNAP mtDNA Population Database (EMPOP at <http://empop.org>) to help determine a consistent mtDNA haplotype.

PROTOCOLS FOR FORENSIC MITOCHONDRIAL DNA ANALYSIS

Sequence Nomenclature and Alignment		
Status:Published		Document ID: 1168
DATE EFFECTIVE 09/15/2017	APPROVED BY mtDNA Technial Leader	PAGE 2 OF 3

- 2.3 Variants from the rCRS should be aligned to the rCRS according to the following standard nomenclature:
- 2.3.1 **Rule 1** – Maintain known patterns of polymorphisms (e.g. 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.
- 2.3.2 **Rule 2** – Use nomenclature with the least number of differences unless it violates known patterns of polymorphisms.
- 2.3.3 **Rule 3a** – 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 variation in the short A-tract preceding 16184 should be noted as transversions.
- 2.3.4 **Rule 3b** – 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.
- 2.3.5 **Rule 4** – Maintain the AC Repeat Motif in the HVIII region from np 515-525.
- 2.3.6 **Rule 5** – Prefer substitutions to insertions/deletions (indels).
- 2.3.7 **Rule 6** – Prefer transitions to transversions unless this is in conflict with Rule 1.
- 2.3.8 **Rule 7** – Place indels contiguously when possible.
- 2.3.9 **Rule 8** – Place indels on the 3' end of the light strand.
- 2.4 Insertions (INS) should be listed to the right of a particular nucleotide position. Insertions are documented by first noting the site immediately 5' to the insertion followed by a point and a "1" for the first insertion, a "2" if there is a second insertion, and so on.
- 2.5 Deletions (DEL) should be listed exactly where the known base in the reference sequence is missing in the sample sequence to minimize the number of differences between the questioned sample and the rCRS reference sequence. Deletions are noted by a "." on the Sequencer printout in the consensus sequence.

PROTOCOLS FOR FORENSIC MITOCHONDRIAL DNA ANALYSIS

Sequence Nomenclature and Alignment		
Status:Published		Document ID: 1168
DATE EFFECTIVE 09/15/2017	APPROVED BY mtDNA Technial Leader	PAGE 3 OF 3

- 2.6 **Sequence heteroplasmy** (also known as point or site heteroplasmy) occurs when a single sample contains at least two mtDNA sequences that differ at one or two nucleotide positions. The appropriate one-letter IUPAC code will be used during the editing of a given site that shows sequence heteroplasmy. This designation will be reflected in the Sequencher Comparison Report. In addition, the presence of sequence heteroplasmy at the given nucleotide position for the respective heteroplasmic bases will be recorded on the sequence editing documentation.
- 2.7 **Length heteroplasmy** occurs in regions that contain many tandem C nucleotides. These regions are commonly referred to as polycytosine or C-stretch regions. Length heteroplasmy refers to a sample that has at least two types, each one differing by the total number of C nucleotides at a given C-stretch.
- 2.7.1 **It will be noted if a given casework sample has length heteroplasmy in HVI.** The number of C residues, however, in the area with HVI length heteroplasmy will not be recorded. Length heteroplasmy in HVI most commonly arises when there is a substitution of a C for a T at position 16,189. The reference type in HVI is C₅TC₄. Sequences showing length heteroplasmy in HVI will be truncated to fit the C₅TC₄ format including the T to C change at position 16,189.
- 2.7.2 **It will be noted if a given casework sample has length heteroplasmy in HVII.** Length variants in HVII are commonly observed in the number of C residues preceding a T residue at position 310. It is often possible to determine unambiguously the dominant length variant in this region. The profile used for further analysis in Sequencher should be composed of only the major type as determined by the analyst.