

# PROTOCOLS FOR FORENSIC MITOCHONDRIAL DNA ANALYSIS

## SEQUENCE NOMENCLATURE AND ALIGNMENT

DATE EFFECTIVE  
06-20-2016

APPROVED BY  
MITOCHONDRIAL DNA TECHNICAL LEADER

PAGE  
1 OF 5

## Sequence Nomenclature and Alignment

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           |

### A. Using Sequencher 4.9

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. 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 according to the following standard nomenclature:
  - a. Characterize profiles using the least number of differences from the reference sequence. Align the 310 T base in the rCRS with a T whenever possible.
  - a. If there is more than one way to maintain the same number of differences with respect to the reference sequence, differences should be prioritized in the

Controlled versions of Department of Forensic Biology Manuals only exist in the Forensic Biology Qualtrax software. All printed versions are non-controlled copies.

© NYC OFFICE OF CHIEF MEDICAL EXAMINER

# PROTOCOLS FOR FORENSIC MITOCHONDRIAL DNA ANALYSIS

| SEQUENCE NOMENCLATURE AND ALIGNMENT |   |                |
|-------------------------------------|---|----------------|
| DATE EFFECTIVE<br>06-20-2016        | APPROVED BY<br>MITOCHONDRIAL DNA TECHNICAL LEADER | PAGE<br>2 OF 5 |

following order: (i) substitutions: transitions are favored over transversions, (ii) insertions/deletions (indels).

- c. Insertions and deletions should be placed furthestmost 3' to a homopolymeric region, with respect to the light strand of rCRS. Insertions and deletions should be combined in situations where the same number of differences to the reference sequence is maintained. In situations involving the "AC" motif, treat this motif as a homopolymeric region with respect to indels in the AC repeat region. Alignment rules a, b, and c are described in Budowle, et al, 2007. For casework samples where alternative alignment following the hierarchy of Wilson, et al, 2002a,b, is also possible, the alternative alignment does not need to be included in the case file.
3. 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.
4. 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 Sequencher printout in the consensus sequence.
5. **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.

# PROTOCOLS FOR FORENSIC MITOCHONDRIAL DNA ANALYSIS

| SEQUENCE NOMENCLATURE AND ALIGNMENT |   |                |
|-------------------------------------|---|----------------|
| DATE EFFECTIVE<br>06-20-2016        | APPROVED BY<br>MITOCHONDRIAL DNA TECHNICAL LEADER | PAGE<br>3 OF 5 |

6. **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.
  - a. **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 C<sub>5</sub>TC<sub>4</sub>. Sequences showing length heteroplasmy in HVI will be truncated to fit the C<sub>5</sub>TC<sub>4</sub> format including the T to C change at position 16,189.
  - b. **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.

## **B. Using Sequencher 4.1.4**

1. Sequence differences between the questioned sample and the revised Cambridge Reference Sequence (rCRS) are generated and printed out from the Difference Review 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. “Ref” (reference) and “Con” (consensus) indicate what bases are present in the rCRS and the questioned sample, respectively, at the designated mtDNA sequence positions.

# PROTOCOLS FOR FORENSIC MITOCHONDRIAL DNA ANALYSIS

| SEQUENCE NOMENCLATURE AND ALIGNMENT |   |                |
|-------------------------------------|---|----------------|
| DATE EFFECTIVE<br>06-20-2016        | APPROVED BY<br>MITOCHONDRIAL DNA TECHNICAL LEADER | PAGE<br>4 OF 5 |

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 according to the following standard nomenclature:
  - a. Characterize profiles using the least number of differences from the reference sequence. Align the 310 T base in the rCRS with a T whenever possible.
  - b. If there is more than one way to maintain the same number of differences with respect to the reference sequence, differences should be prioritized in the following order: (i) substitutions: transitions are favored over transversions, (ii) insertions/deletions (indels).
  - c. Insertions and deletions should be placed furthestmost 3' to a homopolymeric region, with respect to the light strand of rCRS. Insertions and deletions should be combined in situations where the same number of differences to the reference sequence is maintained. In situations involving the "AC" motif, treat this motif as a homopolymeric region with respect to indels in the AC repeat region. Alignment rules a, b, and c are described in Budowle, et al, 2007. For casework samples where alternative alignment following the hierarchy of Wilson, et al, 2002a,b, is also possible, the alternative alignment does not need to be included in the case file.
3. 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.
4. 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 |   |                |
|-------------------------------------|---|----------------|
| DATE EFFECTIVE<br>06-20-2016        | APPROVED BY<br>MITOCHONDRIAL DNA TECHNICAL LEADER | PAGE<br>5 OF 5 |

5. **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 Difference Review. In addition, the presence of sequence heteroplasmy at the given nucleotide position for the respective heteroplasmic bases will be documented on the editing sheet.
6. **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.
- a. **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 C<sub>5</sub>TC<sub>4</sub>. Sequences showing length heteroplasmy in HVI will be truncated to fit the C<sub>5</sub>TC<sub>4</sub> format including the T to C change at position 16,189.
- b. **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 Sequencer should be composed of only the major type as determined by the analyst.