mtDNA Sequence Analysis, Interpretation, and Report Writing

Table of Contents

mtDNA Sequence Analysis, Interpretation, and Report Writing

1   INTRODUCTION

These procedures describe the use of the software packages for mitochondrial DNA (mtDNA) sequence analysis and assembly, as well as how to perform mtDNA comparisons, interpretations, weight assessments, and write reports.

2   SCOPE

These procedures apply to DNA personnel in the DNA Casework Unit (DCU) who analyze and interpret mtDNA sequence data and generate reports.

3   EQUIPMENT

- Sequencing Analysis Software (Applied Biosystems™, version 6.0 or higher)
- Sequencher™ Software (Gene Codes Corporation, version 5.0 or higher)
- PopStats (CODIS, version 9.0 or higher)
- EMPOP (Available at http://empop.online/, current version)

4   STANDARDS AND CONTROLS

The controls used throughout mtDNA analysis are the reagent blank (RB), negative control (NC), and positive control (HL60). Refer to Section 5.4.1 for the interpretation of these controls.

4.1   Reagent Blanks and Negative Controls

A. The RB monitors for the presence of exogenous DNA from the extraction steps through sequencing and is processed from the DNA extraction procedure through the sequencing procedure. If a sample is re-extracted, a new RB is generated in the repeated extraction.

B. The NC monitors for the presence of exogenous DNA from the amplification procedure through the sequencing procedure.

C. If a sample is re-amplified with the same primers, sample volume, and amplification parameters (cycle #), the associated RB does not need to be re-amplified, provided that an HL60 and NC are included with the re-amplification.

D. The HL60 and NC will be included with all re-cycle sequencing. The HL60 will be included with all re-injections. If only the HL60 requires re-injection, only the HL60 needs to be processed.

E. Regardless of the relative quantity of mtDNA in the RB, NC, and the sample, if the RB or NC sequence is in concordance with the sample, the results of the amplicon are not used for comparison purposes. The sample sequence can be used under the following conditions:
1. the RB or NC does not analyze or is not of requisite quality for comparison purposes;

2. the RB or NC analyzes but does not align to the revised Cambridge Reference Sequence (rCRS);

3. the RB or NC sequence analyzes, but the sequence is not in concordance with the sample

F. If analyzable and alignable mtDNA sequence is obtained from the RB or NC, an attempt is made to determine the possible source of this DNA.

4.2 Positive Control

A. The use of a positive control monitors the success of amplification, sequencing, and sequence analysis. The HL60 cell line DNA is used as the positive control for mtDNA analysis.

B. The mtDNA analysis of a particular sample extract is repeated, going back to the appropriate step as judged by the Examiner, if the positive control fails to amplify (e.g., is <1ng/µl) and/or sequence correctly.

C. Positive Control Sequences

HL60 DNA is used as a positive control. The whole control region (WCR) of the mtDNA from this cell line has been amplified and sequenced by the FBI Laboratory. The HL60 mtDNA WCR sequence, expressed as differences from the rCRS, is provided in the table below. The mtDNA HV1 and HV2 sequences for the cell lines 9947A and K562 are provided for information.

Cell Line	HL60 (WCR)	9947A	K562
HV1	16069 T 16193 T 16278 T 16362 C	16311 C	16126 C 16294 T 16296 T 16324 C
HV2	73 G 150 T 152 C 263 G 295 T 315.1 C 489 C (outside of HV2)	93 G 195 C 214 G 263 G 309.1 C 309.2 C 315.1 C	73 G 263 G 310-313

5 PROCEDURE

5.1 Sequencing Analysis

Raw sequencing data is first processed using Sequencing Analysis to trim data obtained by the sequencer before and after the actual sequence. This allows for maximum quality of the sequencing data.

The steps outlined below describe the general steps involved in analyzing the data from sequencing runs. Alternative pathways or shortcuts for accomplishing the same tasks may be used.

    A. Launch Sequencing Analysis software. A username and password are required.

    B. Add files to the Sample Manager window by selecting "Add Sample" from the File menu. Select desired run folder and choose desired analysis files. Click "Add Selected Samples" and "OK."

    C. To view data from a sample file, click to check the "Show" box or highlight all samples and click on the "Show" button. The electropherogram may be viewed by clicking on the "Electropherogram" tab.

    D. Display the raw data by clicking on the "Raw" tab. Note the beginning and ending scan numbers by moving the cursor to the beginning and end of data peaks (the scan numbers should be similar for all capillaries within a run). Signal intensity may also be reviewed in the "Raw" tab.

    E. Enter the beginning scan numbers in the "Peak 1" and "Start" columns. This ensures that the first sequenced base is included.

    F. Enter the ending scan numbers in the "Stop" column. This will decrease the number of "Ns" at the end of the run and focus the signal intensity calculations on the appropriate section. All runs [NC (Negative Control), RB (Reagent Blank), sample, and HL60 (positive control)] with a given primer must have the same "Peak 1" "Start" and "Stop" values. Note: A single "Peak 1" and "Start," as well as a single "Stop" number may be used for all files within a run, as long as these numbers encompass all sequence data in each file.

    G. Ensure the "BC" (BaseCaller) column is checked.

    H. Click on the green arrow button to start the analysis.

    I. Review the resulting electropherograms in the "Electropherogram" tab to ensure the applicable range of sequence data has been captured. Save the files. **Note: The saved files still retain all of the raw data.**

5.2 Assembly of Sequences Using Sequencher™

Trimmed sequence data is assembled and edited using Sequencher™. Two strands (one from the forward primer and one from the reverse primer) must be used to edit sequence whenever possible. Single-stranded sequence can be used, but can be only edited toward ambiguity (the most conservative editing call) when there is a surrounding context for making the editing decision. Artificial insertions, deletions, or ambiguities created by the Sequencing Analysis software may be edited in single-stranded sequence if supported by the surrounding context.

The steps outlined below describe the general steps involved in assembling the data from various primers to yield an mtDNA sequence. Alternative pathways or shortcuts for accomplishing the same tasks may be used.

A. Launch Sequencher™ software, using the settings represented in Appendix A.

B. Open "New Project" under the "File" pull-down menu.

C. Highlight the files you wish to import and drag them into the project window. Open each sequence individually and delete unanalyzable sequence from the ends. Note which samples are not analyzable (typically this includes RBs and NCs).

D. Highlight analyzable forward and reverse sequences for assembly and click on "Assemble Automatically." The default settings of 85 minimum match percentage and 20 minimum overlap are usually sufficient for assembly of samples and positive controls (see Appendix A, Figure 1). Generally, RBs and NCs that do not assemble under the default parameters should be considered not analyzable.

E. View the contig produced. Check the direction of the sequence files by looking at the primers in the file name. If they are not in the appropriate orientation, go to the view menu and select "Reverse Comp."

F. Click on the "Bases" box to observe the nucleotide sequence and click on "Show Chromatograms" to observe the electropherograms. Delete the amplification primer sequences.

G. Edit the sequences using information from all strands in the electropherogram window until the forward and reverse primer sequences are the same at all positions. Each base in the sequence must be manually reviewed for correct identification using the electropherogram windows. It is possible that ambiguous bases will remain after the sequence is fully edited. These ambiguities will be designated according to International Union of Pure and Applied Chemistry (IUPAC) nomenclature (see Section 5.3 for more details).

If the sequence is not of requisite quality (e.g., high background, poor peak resolution, or indicative of a mixture), the sequence will not be used. The sample

may be re-extracted, re-amplified, re-cycle sequenced, or re-injected, as appropriate.

A mixture is a sequence with multiple mixed base positions. Determination of a mixture is dependent on the length of sequence obtained, the contextual sequence information, and the site of the mixed base positions. If any region of a sample is determined to be a mixture, absent primer binding site mutations, the entire extract will be considered a mixture and not used. If available, the sample may be re-extracted.

H. When the editing is completed, import the rCRS into the project. Highlight the rCRS and the contig(s) relevant to the sample (e.g., HV1 and HV2) in the project window and click on "assemble automatically." A new contig will be created which will consist of the sample's consensus sequence and the rCRS.

Note: The rCRS may be imported and assembled with the sample sequences in Step 4 above. At a minimum, the contig is named to reflect the sample designation and the appropriate symbols/initials. Ensure base calls adhere to the sequence nomenclature set forth in Section 5.3.

I. Differences from the rCRS can be viewed in the contig window or by highlighting the contig and choosing "Compare Consensus to Reference" from the "Contig" pull-down menu.

J. The Difference file can be saved by exporting it as "Report Format: Individual Reports." Ensure the exported file and/or the Sequencer™ project view contain(s) the following information: Laboratory Number, Examiner symbols/initials, sample ID, date, and sequence range(s). On the Sequencer™ project view, sequences which are not analyzable are noted "NA", and sequences which are analyzable but do not align with the rCRS are noted "ABNA."

5.3 Sequence Nomenclature

A. Polymorphisms are noted on the Sequencher™ files and transferred to a sequence summary sheet by noting the nucleotide position followed by the code for the polymorphic base (e.g., 263 G). Nucleotide base positions are designated according to the standard nomenclature set by IUPAC and listed in the table below:

CODE	DESIGNATION
A	**Adenine**
B	C,G, or T
C	**Cytosine**
D	A, G, or T
G	**Guanine**
H	A, C, or T
K	G or T (Keto)
M	A or C (aMino)
N	aNy base or combination of bases
R	A or G (puRine)
S	G or C
T	**Thymine**
V	A, C, or G
W	A or T
Y	C or T (pYrimidine)
OTHER	DESIGNATION
-	Deletion
X.1(.2,...)	Insertion after position X

B. Variants from the rCRS will be coded in accordance with the hierarchical nomenclature rules from *SWGDAM Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories* (reproduced text below in italics):

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

    ○ Insertions are described by noting the site immediately prior to the insertion followed by a ".1" (for the first insertion), a ".2" (if there is a second insertion), and so on, and then by the nucleotide that is inserted (e.g., 315.1 C). Deletions are noted by a dash (-).

    ○ All sequences containing insertions or deletions outside the expected HV1, HV2, or HV3 C-stretch regions, as well as nucleotide positions 249, 290, 291, and between 514 - 525, will be compared to known phylogenetic alignments.

- For assistance with identifying known patterns of polymorphisms, the sequence may be compared to those listed in the *SWGDAM Mitochondrial DNA Nomenclature Examples Document*, queried using an approved phylogenetic alignment tool [such as the one available within the EMPOP database (http://www.empop.online)], or a subject matter expert (SME), as designated by the Technical Leader, may be consulted.

- **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 1 (HV1): 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 2 (HV2): C-stretches in HV2 will be named with a 310 C 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.

  - When a 310 C occurs, no insertions relative to the rCRS at 303-315 are noted.

- **Rule 4** – Maintain the AC repeat motif in the HVIII region, generally between np 514-525.

  - The AC repeat is treated as a single unit, with indels grouped immediately 5' to np 525. A single AC insertion is noted as 524.1 A, 524.2 C and a deletion of AC is noted as 523 -, 524 -. Length variants will not be recorded for this region, but the sequence will be trimmed, if needed, to include all bases which do not exhibit other underlying bases.

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

- **Rule 6** – Prefer transitions to transversions.

- **Rule 7** – Place indels contiguously when possible.

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

C. Point or sequence heteroplasmy is defined as more than one base at a given position (independent of overlap of peaks created by length variants), present in both sequencing directions and above the level of background noise. Typically, point heteroplasmy is detected at only one or two positions within the mtDNA control region for a given sample.

D. Length heteroplasmy is defined as more than one length variant in a given region and typically occurs in the homopolymeric C stretch region in HV1, HV2, and/or HV3.

1. Length heteroplasmy in HV1 most commonly arises when there is a substitution of a C for a T at position 16189. The reference type in HV1 is C_5TC_4. HV1 length variants will not be recorded in casework samples due to the pronounced out-of-phase sequence following this stretch of Cs. Generally, the sequences will be truncated to fit the format of the reference. This is accomplished by starting at position 16180 in the forward strand and truncating after the 14th base. The reverse strand is truncated by counting backwards starting at position 16193 and matching the number of Cs in the forward strand. Only non-C insertions will be recorded following position 16193 when length heteroplasmy is observed in this area.

2. Length variants in HV2 are commonly observed in the number of C residues preceding a T residue at position 310. The reference type in HV2 is C_7TC_5. Length variants of C residues between positions 302 - 310 will not be noted. Instead, the major length variant as called by Sequencher™ will be included in the Difference file and listed in a sequence summary table.

3. Length variants in HV3 will not be noted and typically will be based on single-stranded data. The sequence will be trimmed to include all bases which do not exhibit other underlying bases.

E. Typically, the range is defined as the span of double-stranded sequence obtained and denoted by the positions relative to the rCRS. In some instances, single-stranded data at the 5' or 3' end or surrounding C-stretch regions may be included in the range.

5.4 Procedure for Sequence Comparisons

5.4.1 *Evaluation of Control Sequences*

- The positive control must type correctly (see Section 4.2) and be of requisite quality in order to report the sequence for an associated sample.

- After RB and NC sequence data are analyzed, a determination is made as to whether any analyzable sequence is present. If analyzable sequence is present in any RB or NC, it must first be compared to the sequence obtained from the associated sample to determine whether the contamination seen in the RB or NC could be the source of the sequence seen in the sample. To eliminate this possibility, the sequence of the sample and the corresponding RB or NC must differ at one or more base positions within each amplified region. A difference in length variants is not considered as a base difference when comparing RB and NC sequences against the sample sequence. If the sequence of the RB or NC within a region is concordant with the sample, the sample should be either re-extracted, re-amplified, or re-cycle

sequenced as appropriate at the Examiner's discretion. If these methods fail to resolve the issue, the region may not be included in the reported sequence for that sample.

5.4.2   *Comparisons Among Samples*

- When comparing sequences obtained from samples, only the regions with a common range will be evaluated. For example, if a partial sequence (16024-16365 and 73-284) is obtained for a sample of unknown origin, and a full sequence is obtained for the sample of known origin, the comparison will be conducted on positions 16024-16365 and 73-284 only. Cytosine insertions at nucleotide position 309 will not be used for comparison purposes.

5.4.3   *Interpretation of Sequence Comparisons*

- The following conclusions are determined for sequence comparisons:

    - **Exclusion**: If the samples differ at two or more positions, they are excluded as coming from the same source.

    - **Inconclusive:** If the samples differ at a single position only the comparison is inconclusive.

- Because of the possibility of undetected heteroplasmy, additional samples of known origin may be analyzed when sequences from samples of unknown origin and sequences from samples of known origin differ at a single position. Known origin samples may include blood, buccal swabs, and hair. Hair fragments from a hair standard of known origin may be combined and processed as a single known sample.

    - **Cannot Exclude**: If the samples have the same sequence, or demonstrate sequence concordance, they cannot be excluded as potentially coming from the same source. Sequence concordance is defined as having a common base at each position at which sequence data were obtained in the sample.

5.4.4   *Sequence Confirmation*

- The sequences for all samples and associated controls (HL60, NC, and RB) must be confirmed by a second qualified individual. Confirmation of the sample, NC, and RB involves independently evaluating and assembling all of the same sequence data from the analysis data that were used by the reporting Examiner. Confirmed sequence range is defined as the shortest length of sequence upon which the Examiner and the confirmer have reached consensus. Separate Sequencher™ projects and Difference files are generated by each individual.

- The confirmer will evaluate the assembled and edited HL60 contig (i.e., set of overlapping DNA segments) for the sample and indicate approval by appending the Difference file name with their symbols/initials.

- Confirmation of mixtures or of samples of insufficient quality will be indicated by annotation of the confirmer's Sequencher™ project with "mixture" or "insufficient quality." Portions of samples not used in the final project due to insufficient data quality do not need to be reviewed by a confirmer.

- For cases in which the interpretation "Cannot Exclude" has been made, proceed to Section 5.5, "Procedure for Weight Assessment." For other cases, proceed to Section 5.6, "Procedure for Report Writing."

5.5 Procedure for Weight Assessment

5.5.1 *Searching Profiles in CODIS*

- Typically, a CODIS Popstats search is conducted when samples of unknown origin and samples of known origin are found to be concordant. CODIS 9.0 or higher is used to search casework profiles in the database using the settings represented in Appendix B.

- Those database samples whose range(s) are fully included within the casework profile range are used in the database search result. Presently, this includes nucleotide positions 16024-576.

- Ambiguous base positions are searched using the appropriate IUPAC symbol (see 5.3).

- The CODIS Popstats database search result provides the number of database profiles that match the casework profile, as well as the number of profiles in the database that differ by up to five positions. In addition, the search result details the database profiles that have 0 base differences from the search profile, which may be included in the case file. Special attention should be paid to database search results from queries with sequences with a 310 C, as well as with non-C insertions following position 16193.

5.5.2 *Special Considerations*

- All sequence polymorphisms in the sample searched are entered; however, all length variants at nucleotide positions 16193, 309, and 573 are ignored in database searches of relevant concordant sequences. Hence, length variability in these regions will not add any additional rarity to a database profile search.

5.5.3 _Upper Bound Frequency Estimate_

- The one-tailed Clopper and Pearson method for determining the upper bound frequency estimate of a mtDNA profile is used at a confidence level of 95% (α = 0.05).

- The following formulae are used:
  1. In cases where the profile has been observed in the database:

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

   This equation finds the value _po_ of the population proportion p for which the cumulative probability 0, 1, . . . _x_ copies of the profile is equal to α and n is the number of profiles in that population. This equation will require a computer to solve.

  2. In cases where the profile has <u>not</u> been observed in the database (x=0):

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

5.5.4 _Searching Profiles in EMPOP_

- In order to provide additional information regarding population groups relevant to a case, but not contained within the CODIS mtDNA population database, a search may be conducted using the EMPOP database (http://empop.online), with search parameters set to "pattern" match type, and disregarded insertions/deletions (indels) at positions 16193, 309, and 573 (see Appendix C). Note that EMPOP is not compatible with Internet Explorer.

- The EMPOP search result is provided for geographic origin affiliations: Africa, America, Asia, and Europe. Relevant subgroups within these populations can be selected in the "Find origin" box. The number of observations and number of profiles searched is entered into a table in the report for the EMPOP search (see Section 5.6.1.14). If a one-tailed Clopper and Pearson frequency estimate is not provided in the EMPOP search, it will be calculated using the formula in Section 5.5.3, and the value entered into the report table.

5.6   Procedure for Report Writing

The following are examples of wording for use in a _Laboratory Report._ These examples represent typical case situations. Determination of the exact wording is made by the reporting Examiner on a case-by-case basis.

5.6.1 _Results of Examinations Section_

- As appropriate, the final sentence in the Results of Examinations section will be "No other mtDNA examinations were conducted."

5.6.1.1 _Cannot Exclude (Full Sequence)_

- For concordant samples of unknown origin and known origin, the Report of Examination will include a table (see below) containing the CODIS 9.0 Popstats (or higher) search results of relevant population groups, to include at least the African-American, US Caucasian, and US Hispanic groups. Any group in which there is an observation of the sequence will also be included in the table. Population groups in which the searched profile was not observed are noted in a footnote to the table (see footnote below table, with population sample size indicated parenthetically). When multiple database searches have been conducted in a case and have the same results, a single table may be incorporated into the report.

  "Mitochondrial DNA (mtDNA) sequences were obtained from item 1 and DOE. The mtDNA sequences obtained from item 1 and DOE are the same within the sequence range obtained in common to both samples. Therefore, DOE cannot be excluded as the source of item 1.

  Searching the CODIS mtDNA population database (CODIS 9.0, containing 10,629 individuals, searching positions 16024-16390 and 49-408)[A], the mtDNA sequence obtained from item 1 and DOE has been observed as follows[B]:

Population Group	Number of Observations	Number of Profiles in Population Group	Upper Bound Frequency Estimate
African-American	0	2449	0.12%
US Caucasian	0	2609	0.11%
US Hispanic	0	2576	0.12%

In addition to the population groups listed in the table above, the population database search included a search of mtDNA sequences from individuals from China – unspecified region (168), China – Hong Kong (376), Japan (302), Korea (281), US Asian (645), Vietnam (187), US Native American (1036) with no observation of the queried sequence in these groups. The numbers in parentheses indicate the number of individuals in each population group.

[A] The range of sequence positions included in a database search is the one obtained in common for the samples.

[B] The population database table(s) above has been included to indicate how common or rare a sequence is expected to be in the general population. The upper bound frequency estimate is based on a 95% confidence interval and gives an estimate of the highest percentage of individuals in each population group expected to have the same profile as the referenced samples. Calculation of the upper bound frequency estimate is directly dependent upon the

number of profiles in the population group; larger population group sizes will provide more refined upper bound frequency estimates. Mitochondrial DNA profiles were placed into population groups within the database based on self-identification by the donor. A searched profile may or may not appear in the population database or it may be observed within multiple groups in the database. Therefore, mtDNA cannot be used to identify the population group to which an individual belongs."

5.6.1.2 Cannot Exclude (Partial Sequence)

"A mitochondrial DNA (mtDNA) sequence was obtained from DOE. A partial mitochondrial DNA (mtDNA) sequence was obtained from item 1 (nucleotide positions 16005-16236 and 49-408 only). The mtDNA sequences obtained from item 1 and DOE are the same across the sequence range obtained in common to both samples. Therefore, DOE cannot be excluded as the source of item 1.

Searching the CODIS mtDNA population database (CODIS 9.0, containing 10,629 individuals, searching positions 16024-16236 and 49-408)[A], the mtDNA sequence obtained from item 1 and DOE has been observed as follows[B]:"

- Include the database search results table and explanatory wording as in 5.6.1.1.

- Note: A limited sequence range will only be indicated in the report if it is less than the minimum range of HV1and HV2 (i.e., less than 16024-16365 and 73-340).

5.6.1.3 Concordance (Sequence Heteroplasmy)

"Mitochondrial DNA (mtDNA) sequences were obtained from item 1 and DOE. The mtDNA sequence obtained from item 1 is the same as the mtDNA sequence obtained from DOE, with the exception of position 16069. At this position, the presence of a thymine (T) was observed in item 1. In DOE, evidence of both a thymine (T) and a cytosine (C) was characterized at position 16069. Since the sequences obtained from item 1 and DOE are concordant, JOHN DOE cannot be excluded as the source of the item 1 hair.

Searching the CODIS mtDNA population database (CODIS 9.0, containing 10,629 individuals, searching positions 16024-16365 and 73-340)[A], the mtDNA sequence obtained from item 1 and DOE (including both nucleotides, C and T, at position 16069) has been observed as follows[B]:"

- Include the database search results table and explanatory wording as in 5.6.1.1.

5.6.1.4 Exclusion

"Mitochondrial DNA (mtDNA) sequences were obtained from item 1 and DOE. The mtDNA sequence obtained from item 1 is different than the sequence obtained from DOE. Therefore, DOE is excluded as the source of item 1."

5.6.1.5 Inconclusive (One Base Difference)

"Mitochondrial DNA (mtDNA) sequences were obtained from item 1 and DOE. The mtDNA sequence obtained from item 1 is the same as that obtained from DOE, with the exception of position 16069. At this position, the presence of a thymine (T) was observed in item 1. In DOE, a cytosine (C) was characterized at position 16069.

- Due to the one base pair difference observed between item 1 and DOE, no conclusion could be reached as to whether DOE can be excluded as the source of item 1."

5.6.1.6 Evidentiary Samples Concordant

"Mitochondrial DNA (mtDNA) sequences were obtained from items 1 and 2. The mtDNA sequences obtained from items 1 and 2 are the same. Therefore, items 1 and 2 cannot be excluded as originating from a common source."

- A CODIS Popstats search will be conducted in such cases.

5.6.1.7 Mixture

"The mitochondrial DNA (mtDNA) sequence obtained from item 1 indicates the presence of a mixture of mtDNA from more than one individual. Because mixtures of mtDNA are not interpretable, no comparisons can be performed using item 1."

5.6.1.8 No Sequence Obtained

"There was insufficient mitochondrial DNA (mtDNA) present for a DNA sequence to be obtained from items 1 or 2."

5.6.1.9 Not of Requisite Quality:

"The mitochondrial DNA (mtDNA) sequence obtained from item 1 was not of the requisite quality for comparisons or interpretations to be conducted."

5.6.1.10 Missing Person Reference Sample

"A mitochondrial DNA (mtDNA) sequence was obtained from MARY DOE, the mother of JANE DOE (NCIC No. MXXXXXXXXX). The mtDNA profile from MARY DOE will be entered into the Relatives of Missing Persons Index of the Combined DNA Index System (CODIS) and maintained by the FBI Laboratory for future comparisons."

5.6.1.11 Missing Person Unknown Origin Sample

"A mitochondrial DNA (mtDNA) sequence was obtained from item 1 (NCIC No. UXXXXXXXXX). The mtDNA profile from item 1 will be entered into the Unidentified Human (Remains) Index of the Combined DNA Index System (CODIS) and maintained by the FBI Laboratory for future comparisons."

5.6.1.12 Matching mtDNA Sequences from a CODIS search

"This report supplements the FBI Laboratory report dated September 15, 2008. Refer to that original report for the results of the previous mtDNA testing. This report contains the results of a CODIS (Combined DNA Index System) search."

- Results of the CODIS search:

"As a result of searching the FBI's CODIS database, the mtDNA profile obtained from item 1 (ID: AB/CD010101001Q1) has the same mtDNA profile as a biological sibling of JANE DOE (ODTL-02-0299) submitted by the [name of other DNA testing laboratory]. Therefore, JANE DOE cannot be excluded as the source of item 1 (bone)."

5.6.1.13 TEDAC Cases

"A mitochondrial DNA (mtDNA) sequence was obtained from item 1 (CEXC XX/XXX/XX) and will be entered and searched in the appropriate databases. This profile is retained for future comparisons."

5.6.1.14 Cannot Exclude using an EMPOP search

- Typically, a CODIS Popstats search will be conducted along with an EMPOP search. Therefore, the wording from Section 5.6.1.1 would be used in addition to the following (Note: The number of subpopulation groups, as well as the number of individuals searched, is case dependent):

"A search of the European DNA Profiling Group Mitochondrial DNA Population Database (EMPOP; currently available online at http://empop.online) was also performed to provide additional information regarding population groups relevant to this case.

Searching the EMPOP population database[C] (EMPOP version 11, containing 27,479 individuals within the searched sequence range of positions 16024-16390 and 49-408)[D] the mtDNA sequence obtained from items 1 and 2 has been observed as follows[E]:

Geographic Affiliation	Number of Observations	Number of Individuals in Population Group	Upper Bound Frequency Estimate

AFRICA	0	1900	0.16%
AMERICA	0	14813	0.02%
ASIA	1	6440	0.07%
South East Asia			
Philippines	1	153	3.06%
Indonesia	0	277	1.08%
Malaysia	0	123	2.41%
Thailand	0	190	1.56%
Vietnam	0	187	1.59%
East Asia			
China	0	544	0.55%
Japan	0	402	0.74%
Korea	0	692	0.43%
EUROPE	0	4326	0.07%

[C] The EMPOP search was performed with "pattern" match type search parameters and disregarded indels at positions 16193, 309, and 573.

[D] The range of sequence positions included in a database search is the one obtained in common for the samples.

[E] The CODIS and EMPOP population database tables(s) above have been included to indicate how common or rare a sequence is expected to be in the general population. The upper bound frequency estimate is based on a 95% confidence interval and gives an estimate of the highest percentage of individuals in each population group expected to have the same profile as items 1 and 2. Calculation of the upper bound frequency estimate is directly dependent upon the number of profiles in the population group; larger population group sizes will provide more refined upper bound frequency estimates. Mitochondrial DNA profiles were placed into population groups within the CODIS database (first table) based on self-identification by the donor, and within the EMPOP database (second table) based on geographic affiliation. Some population groups within the CODIS database contain a subset of profiles from similar groups contained within EMPOP. A searched profile may or may not appear in the population database or it may be observed within multiple groups in the database. Therefore, mtDNA cannot be used to identify the population or subpopulation group to which an individual belongs."

5.6.2   *Methodology and Interpretation Section*

5.6.2.1   *If a Comparison is Conducted*

"DNA is extracted from each sample, and portions of the control region of the mitochondrial DNA are amplified using the polymerase chain reaction (PCR). The amplified regions are sequenced using fluorescent dye-labeled chemistry. The

sequences obtained are aligned and compared to the revised Cambridge Reference Sequence (rCRS). Differences between the sample sequence and the rCRS are noted by nucleotide position and DNA base. The annotated profiles for all of the samples are then compared. Matching profiles may be searched against the CODIS mtDNA population database to provide an upper bound frequency estimate.

Mitochondrial DNA cannot be used to conclusively identify an individual because mtDNA is maternally inherited and all maternally-related individuals are expected to have the same mtDNA profile. Also, unrelated individuals may have the same mtDNA profile within the sequenced range.

The following interpretations are possible for sequence comparisons:

1. CANNOT EXCLUDE:
   If the samples have the same sequence, or are concordant, they cannot be excluded as coming from the same source. Sequence concordance is defined as having a common DNA base at each position at which sequence data were obtained in the sample.

2. INCONCLUSIVE:
   If samples differ at only a single nucleotide position, no conclusion can be reached as to whether they originate from the same source.

3. EXCLUSION:
   If samples differ at two or more nucleotide positions, they are excluded as coming from the same source."

- When an EMPOP search is conducted, the final sentence of the first paragraph may be modified, as appropriate:

  "Matching profiles may be searched against the CODIS mtDNA population database as well as the EMPOP mtDNA population database to provide an upper bound frequency estimate."

5.6.2.2 If No Comparison is Conducted

"DNA is extracted from each sample and portions of the control region of the mitochondrial DNA are amplified using the polymerase chain reaction (PCR). The amplified regions are sequenced using fluorescent dye-labeled chemistry. The sequences obtained are aligned and compared to the rCRS. Differences between the sample sequence and the rCRS are noted by nucleotide position and DNA base.

Mitochondrial DNA cannot be used to conclusively identify an individual because mtDNA is maternally inherited and all maternally-related individuals are expected to

have the same mtDNA profile. Also, unrelated individuals may have the same mtDNA profile within the sequenced range."

5.6.3 *Remarks Section*

As appropriate, the Remarks section of the report may include one or more statements similar to those below:

5.6.3.1 *Standard Remarks Section*

"The work described in this report was conducted at the Quantico Laboratory, and the results will be maintained by the FBI Laboratory for possible future comparisons. This report contains the opinions and interpretations of the issuing examiner and is supported by records retained in the FBI Laboratory file. This report conforms to the Department of Justice Uniform Language for Testimony and Reports for Forensic Mitochondrial DNA Examinations. For questions about the content of this report, please contact Forensic Examiner Jane Doe at (insert phone number) or (insert email address)."

"The submitted item will be returned to you under separate cover. In addition to the evidence in the case, secondary evidence was generated that will also be returned to you. The secondary evidence can be found in a package marked DNA Secondary Evidence."

5.6.3.2 *Consumption of Evidence*

"Item 1 was consumed during the DNA Casework Unit examinations."

5.6.3.3 *Missing Person Cases*

"Mitochondrial DNA profiles from unidentified human remains that are uploaded to the National DNA Index System (NDIS) require a complete mtDNA profile (at a minimum, nucleotide positions 16024-16365 and 73-340). If uploading to NDIS is required, additional skeletal material is needed. Please submit more skeletal material such as a femur, tibia, or other whole bone sample if available."

"A portion of item 3 was retained in the National Missing Person DNA Database Repository for possible future testing. Archive samples will be retained by the FBI Laboratory until an association is made. Archive samples will be returned to the contributor if requested."

5.6.3.4 *Return of Archival Sample*

"On February 28, 2011 Dr. Ben Jones notified Examiner John Doe by telephone and facsimile that the UNKNOWN FEMALE DECEASED has been identified by dental examination. The return of all skeletal material was requested and the DNA profiles

that were placed into the Combined DNA Index System (CODIS) will be removed. The portion of item 1 that was retained for archive will be returned under separate cover."

5.6.3.5   TEDAC Cases

"The supporting records for the opinions and interpretations expressed in this report are retained in the FBI files. For questions about the content of this report, please contact mtDNA Examiner Jane Doe at (insert phone number)."

5.6.3.6   No mtDNA Examination - Remains Non-Human Origin

"The item 1 bone was examined by Dr. Jane Smith, Curator, Department of Anthropology, Smithsonian Institution, Washington, DC. Dr. Smith's report is enclosed herewith. Because the bone is not of human origin, no mitochondrial DNA examinations were conducted."

5.6.3.7   Requests Not Conducted

"The DNA typing results obtained from the tested items are not eligible for entry into the Combined DNA Index System (CODIS)."

5.6.3.8   CODIS Association Reports

"The information provided by the University of North Texas Center for Human Identification for the missing person is:

Missing Person Name:	JOHN DOE
Sex:	Male
DOB:	6/10/1950
Race:	Caucasian
Date of Last Contact:	3/21/1980
Location of Last Contact:	Charlottesville, Virginia
Agency Submitting Sample:	University of Virginia Police Department
	Lt. Jane Smith
	2304 Icy Road
	Charlottesville, VA 21901
Agency Case Number:	198111718
NAMUS MP#	9766
NCIC #	M196777667

Please contact Lt. Jane Smith of the University of Virginia Police Department at (insert phone number) or jsmith@virginia.edu for potential investigative information.

A portion of Item 1 was retained for archival purposes for possible future testing. Archive samples will be retained by the FBI Laboratory until an association is made. Archive samples will be returned to the contributor if requested.

It is noted that the results of the nuclear DNA CODIS association will be the subject of a separate report. Please contact the DNA Casework Unit at (703) 632-8446 for further information regarding the nuclear DNA association.

The supporting records for the opinions and interpretations expressed in this report are retained in the FBI files. For questions or to advise the FBI Laboratory of any identifications made based on the content of this report, please contact mtDNA Examiner Jane Doe at (insert phone number).

This information is provided only as an investigative lead."

6 CALCULATIONS

A 95% confidence interval is calculated by the CODIS Popstats software in order to determine an upper bound frequency estimate of individuals within each population who are estimated to be included as potential contributors of a particular profile, as described in Section 5.5.3.

7 LIMITATIONS

- Mitochondrial DNA analysis is subject to limitations defined in the extensive validation studies conducted by FBI scientists prior to the implementation of the procedure.
- Because of the sensitivity of mtDNA analysis, contamination is a concern. To this end, measures are taken to minimize and monitor any DNA foreign to the sample. To remove contaminating DNA present on the sample prior to extraction, samples are cleaned with detergents, bleach, and/or other chemicals whenever possible. In some instances, this cleaning may not remove all of the contaminant and a mixed mtDNA profile may result. Extensive quality control procedures (e.g., cleaning of work surfaces and equipment with bleach and/or ultraviolet light prior to use, assessing all critical reagents prior to use on casework samples, physical separation of pre-amplification and post-amplification work areas) are employed to minimize the introduction of foreign DNA into the sample. In addition, RBs and NCs monitor for contamination introduced during the analytical processes. Although the optimal outcome is for no mtDNA to be detected in these controls, controls with sequence the same as that of the associated sample can yield partial or no interpretable results for that sample.
- Although mtDNA analysis is a sensitive technique, it is possible that the quantity of DNA present in some samples is too low to yield results with the current mtDNA methodologies.
- Mitochondrial DNA is maternally inherited and is not unique to an individual. All maternal relatives are expected to have the same mtDNA sequence, absent mutation. In addition, unrelated individuals may share the same mtDNA sequence.

- Mixtures of mtDNA from more than one individual are not interpretable and are reported as such.

8 REFERENCES

SWGDAM, Interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories, most recent version, www.swgdam.org.

SWGDAM, Mitochondrial DNA nomenclature examples document. Additional guidance for the application of the SWGDAM interpretation guidelines for mitochondrial DNA analysis for forensic DNA testing laboratories, most recent version, www.swgdam.org.

9 REVISION HISTORY

Revision	Issued	Changes
00	02/04/2022	Reformatted DNA 410-3 into new template and assigned new Doc ID. Updates to RB and NC guidance to mirror testing SOP and updates to report wording.

10 APPENDIX A: SEQUENCHER SETTINGS
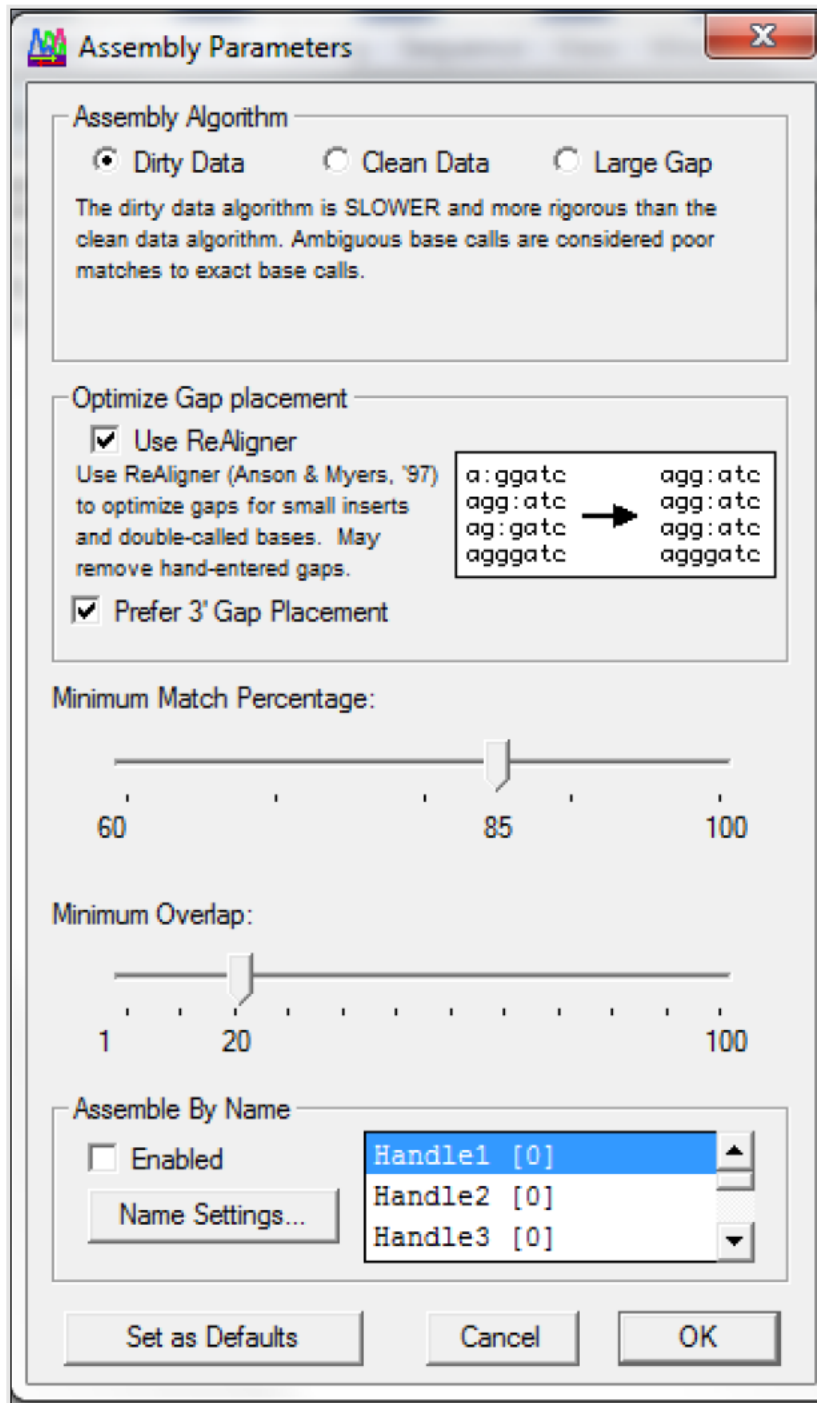
Figure 1 – Assembly Parameters

Appendix A: *Sequencher Settings (cont.)*
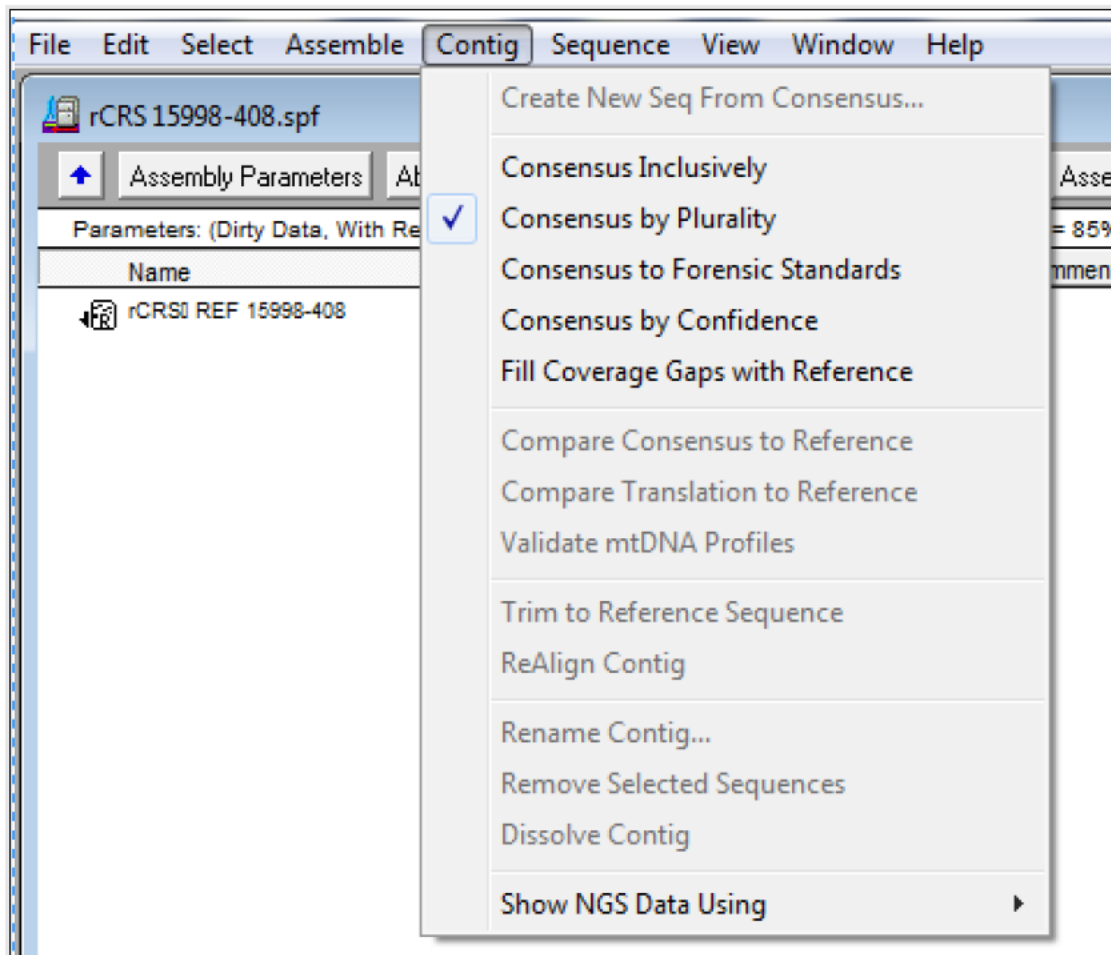
Figure 2 – Contig Consensus

Appendix A: *Sequencher Settings (cont.)*
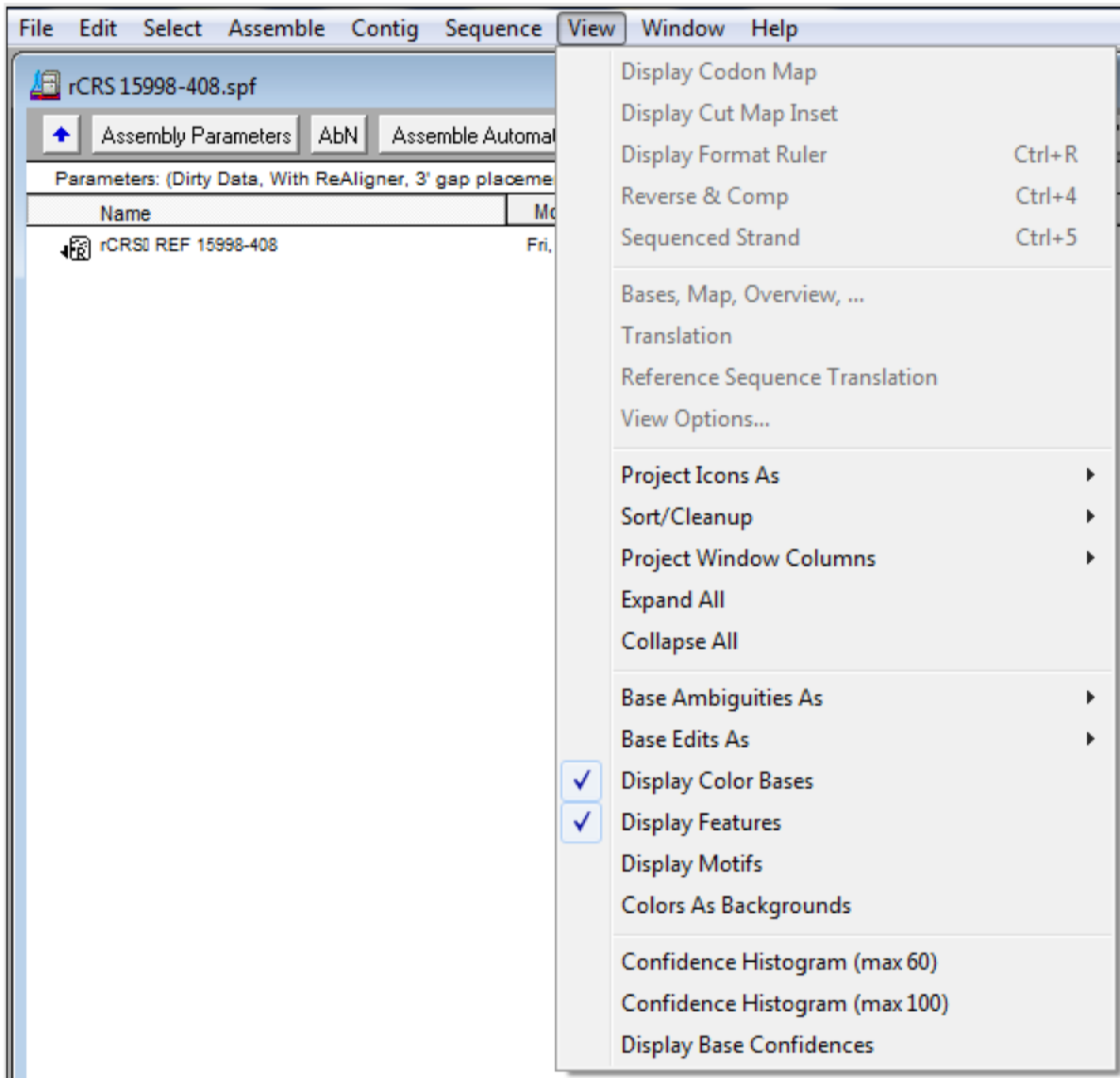
Figure 3 – View settings

Appendix A: *Sequencher Settings (cont.)*

Figure 4 – User Preferences

Default settings for CODIS mtDNA population search

12 APPENDIX C: EMPOP

Figure 1 - Settings for EMPOP database search

Appendix C: *EMPOP (con't)*

Figure 2 - EMPOP Search Results Screen

13 APPENDIX D: SEQUENCES OF AMPLIFICATION AND SEQUENCING PRIMERS

Primers used only for sequencing are marked with an asterisk (*).

Standard Primers

A1   (L 15997) 5'-CAC CAT TAG CAC CCA AAG CT-3'

B2   (H 16237) 5'-GGC TTT GGA GTT GCA GTT GAT-3'

A2   (L 16159) 5'-TAC TTG ACC ACC TGT AGT AC-3'

B1   (H 16391) 5'-GAG GAT GGT GGT CAA GGG AC-3'

A4*   (L 16209) 5'-CCC CAT GCT TAC AAG CAA GT-3'

B4*   (H 16164) 5'-TTT GAT GTG GAT TGG GTT T-3'

C1   (L 048) 5'-CTC ACG GGA GCT CTC CAT GC-3'

D2   (H 285)5'-GGG GTT TGG TGG AAA TTT TTT G-3'

C2   (L 177) 5'-TTA TTT ATC GCA CCT ACG TTC AAT-3'

D1   (H 409)5'-CTG TTA AAA GTG CAT ACC GCC-3'

317*   (L 317) 5'-CCC CCC CTC CCC CCG C-3'

Whole Control Region Primers

A1   (L 15997) 5'-CAC CAT TAG CAC CCA AAG CT-3'

B4*   (H 16164) 5'-TTT GAT GTG GAT TGG GTT T-3'

A4*   (L 16209) 5'-CCC CAT GCT TAC AAG CAA GT-3'

16511*   (L 16511) 5'-CCG ACA TCT GGT TCC TAC-3'

19*   (H 19)   5'-TCC CGT GAG TGG TTA ATA G-3'

296*   (H 296)5'-RRR GGG GGG GTT TGG TG-3'

317*   (L317)   5'-CCC CCC CTC CCC CCG C-3'

617   (H 617)5'-TGA TGT GAG CCC GTC TAA AC-3'

557*   (H 557)5'–GGG GGG TGT CTT TGG GG–3'

Appendix D: Sequences of Amplification and Sequencing Primers

Mini-Primers

1A	(L 16055)	5'-GAA GCA GAT TTG GGT ACC AC-3'
1B	(H 16142)	5'-ATG TAC TAC AGG TGG TCA AG-3'
2A	(L 16131)	5'-CAC CAT GAA TAT TGT ACG GT-3'
2B	(H 16218)	5'-TGT GTG ATA GTT GAG GGT TG-3'
3A	(L 16209)	5'-CCC CAT GCT TAC AAG CAA GT-3'
3B	(H 16303)	5'-TGG CTT TAT GTA CTA TGT AC-3'
4A	(L 16287)	5'-CAC TAG GAT ACC AAC AAA CC-3'
4B	(H 16356)	5'-GTC ATC CAT GGG GAC GAG AA-3'
5A	(L 16347)	5'-CGT ACA TAG CAC ATT ACA GT-3'
5B	(H 16410)	5'-GCG GGA TAT TGA TTT CAC GG-3'