Interpretation of Legacy Mitochondrial DNA Data

Table of Contents

Interpretation of Legacy Mitochondrial DNA Data

1 INTRODUCTION

These procedures describe criteria for the interpretation or reinterpretation of mitochondrial DNA (mtDNA) data from legacy typing procedures (i.e., Sanger Sequencing).

2 SCOPE

These procedures apply to DNA personnel in the DNA Casework Unit (DCU) who are authorized to interpret and reinterpret Sanger sequencing data and/or report comparisons to previously reported mtDNA typing results generated using legacy mtDNA procedures that are no longer in use in the DNA Casework Unit.

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 the current interpretation procedures (i.e., BIO-571) for generalized guidance on reagent blanks and negative controls.

4.1 Positive Control

A. The HL60 cell line DNA is used as the positive control for legacy mtDNA analysis.

Region	HL60 (Whole Control Region)
HV1	16069 T 16193 T 16278 T 16362 C
HV2	73 G 150 T 152 C 263 G 295 T 315.1 C 489 C (outside of HV2)

Table 1 - HL60 variants from rCRS (np 15998-616)

4.2 Assessment of samples and associated controls

A. A sample re-amplified with the same primers, sample volume, and amplification parameters (cycle #) does not require a re-amplified associated RB, provided that an HL60 and NC were included with the re-amplification.

B. Samples reprocessed at cycle sequencing must include the associated HL60 and NC.

C. Samples reprocessed at injection must include HL60 reinjects.
    1. HL60 reinjects do not require reinjection of associated samples.

D. 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.

E. The sample sequence for each individual amplicon can be used under the following conditions (Note: Each amplicon is assessed for contamination individually, and reporting criteria is not dependent on the results of other amplicons):
    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
        i. Any controls with analyzable sequence must have at least one difference from the associated sample, with no underlying peaks at the position of difference that are concordant between the two sequences.
        ii. A difference in length variants is not considered as a base difference when comparing RB and/or NC sequences to the sample sequence.

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 sequence data is collected.

A. 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.
    1. 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.

5.2 Assembly of Sequences Using Sequencher™

A. Trimmed sequence data is assembled and edited using the Sequencher™ software. (See Appendix A for Sequencher™ settings.)
    1. Primer sequences are manually trimmed from all forward and reverse primer amplicons within Sequencher. (See Appendix B for sequencing primers and their positions.)
    2. Sequence with dye blobs, spacing issues and/or noisy baseline encountered at the beginning of a sequence read may be trimmed.
    3. Peaks with low resolution at the end of a sequence read may be trimmed.
    4. Two strands (one from the forward primer and one from the reverse primer) must be used to edit sequence whenever possible.
        i. 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.
        ii. Artificial insertions, deletions, or ambiguities created by the Sequencing Analysis software may be edited in single-stranded sequence if supported by the surrounding context.
    5. Manual verification of the base calls is required at each base position.

5.3 Sequence Nomenclature

For general information regarding sequence nomenclature, see the appropriate mtDNA interpretation procedure (i.e., BIO-571).

A. Differences from the rCRS 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). (Refer to the current interpretation protocol [i.e., BIO-571] for base designations.)

B. 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.

C. 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 14^{th} 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.

D. 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.
    1. The sequence range is defined as the shortest length of sequence shared in common with the positive control and associated sample.

5.4 Data Interpretation

For information regarding data used in sequence comparisons, interpretation of sequence comparisons, weight assessment and report wording, see the appropriate DNA interpretation procedure (i.e., BIO-571).

5.4.1 *Evaluation of Control Sequences*

- The positive control must type correctly (see Table 1) and be of requisite quality in order to report the sequence for an associated sample.

5.4.2 *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 "Inclusion/Cannot Exclude" has been made, refer to the current version of the procedures for mtDNA analysis, interpretation, and reporting (i.e., BIO-571) for information on procedures for weight assessment and statistical calculations.

5.5 Report Writing

Refer to the current version of the procedure for mtDNA Analysis, Interpretation, and Reporting (i.e., BIO-571) for suggested reporting language. Reporting language specific to legacy mtDNA data is included below.

5.5.1 *Methodology and Interpretation Section*

5.5.1.1 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. INCLUSION:
   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."

5.5.1.2 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."

6 CALCULATIONS

For information regarding calculations, see the current mtDNA interpretation procedure (i.e., BIO-571).

7 LIMITATIONS

- These procedures describe the parameters that should be used in conjunction with the current interpretation and reporting procedures (i.e., BIO-571). Refer to the current interpretation and reporting procedures for additional information regarding the limitations of interpretation.

- It is not possible to anticipate the nature of all potential DNA typing results or the nature of the evidentiary specimens from which they may be obtained. These procedures do not exhaust the possible list of the results that may be encountered by the Examiner, nor the conclusions that an Examiner may render, based on his/her interpretation of those results. For those results not specifically described, conclusions should be drawn using the procedures given for the results above that are similar in concept and/or origin.

8 REVISION HISTORY

Revision	Issued	Changes
00	12/16/2024	Moved Sanger Sequencing interpretation information from BIO-571 to create new Legacy Interpretation Procedure for mtDNA data.

9 APPENDIX A: SEQUENCHER SETTINGS

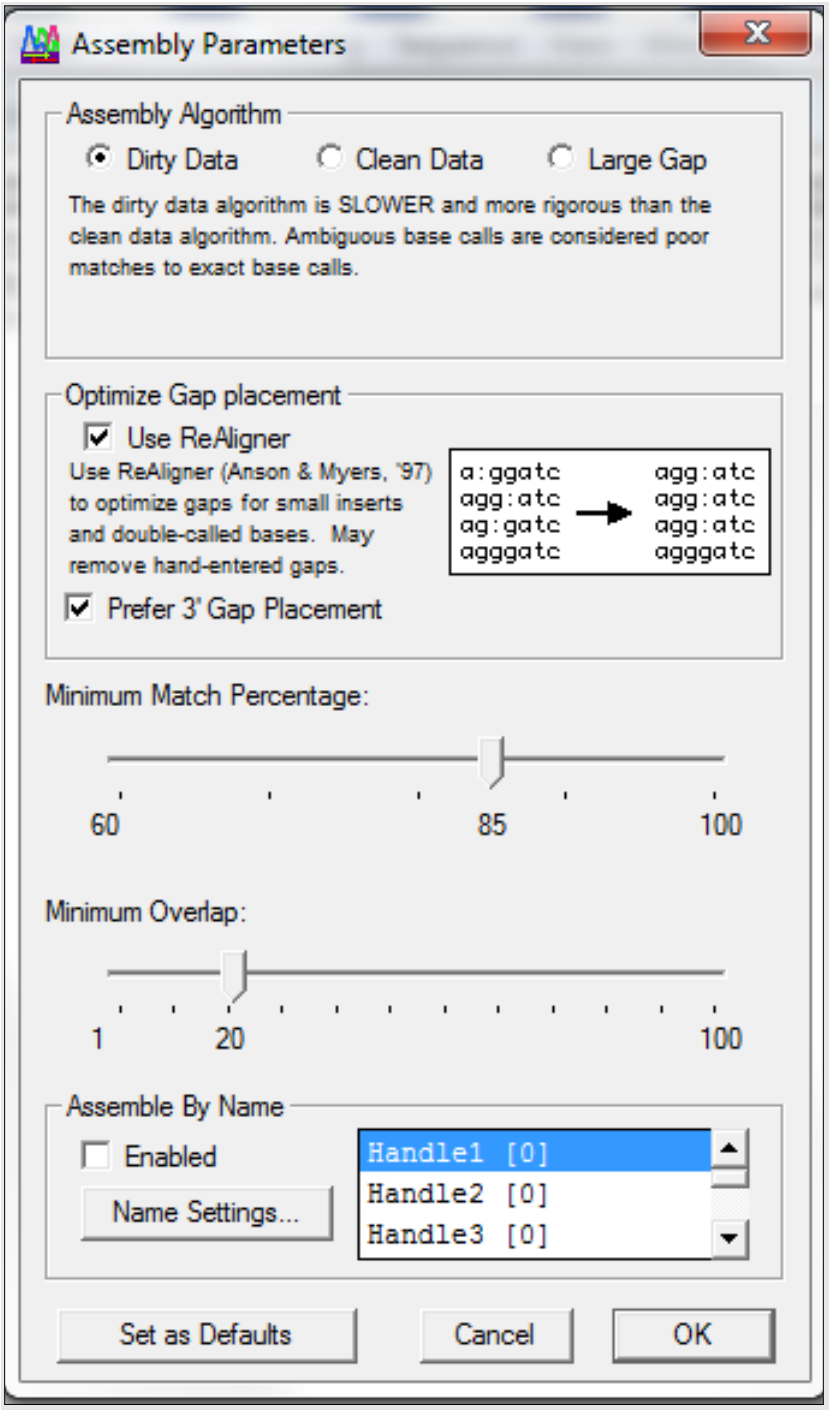

Figure 1 – Assembly Parameters

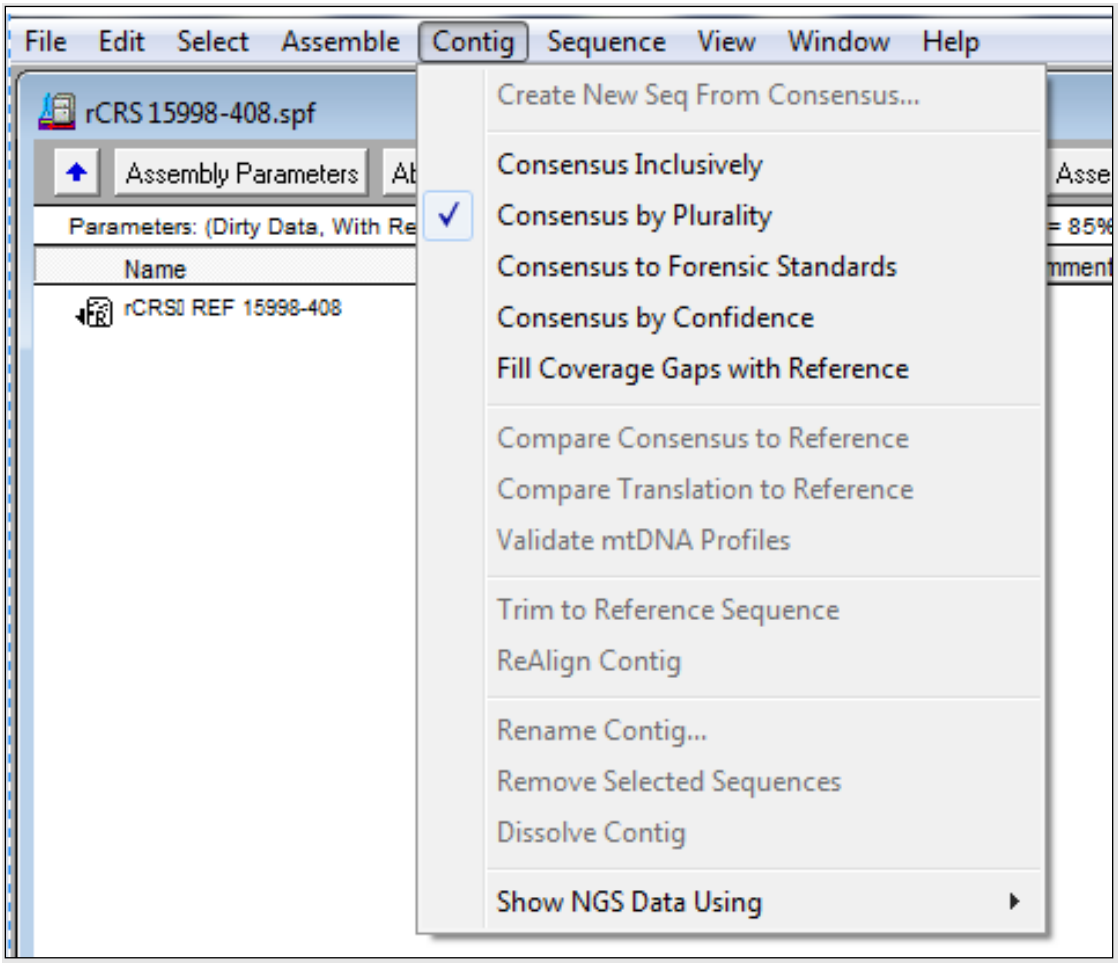

Figure 2 – Contig Consensus

APPENDIX A: SEQUENCHER SETTINGS (CONT.)

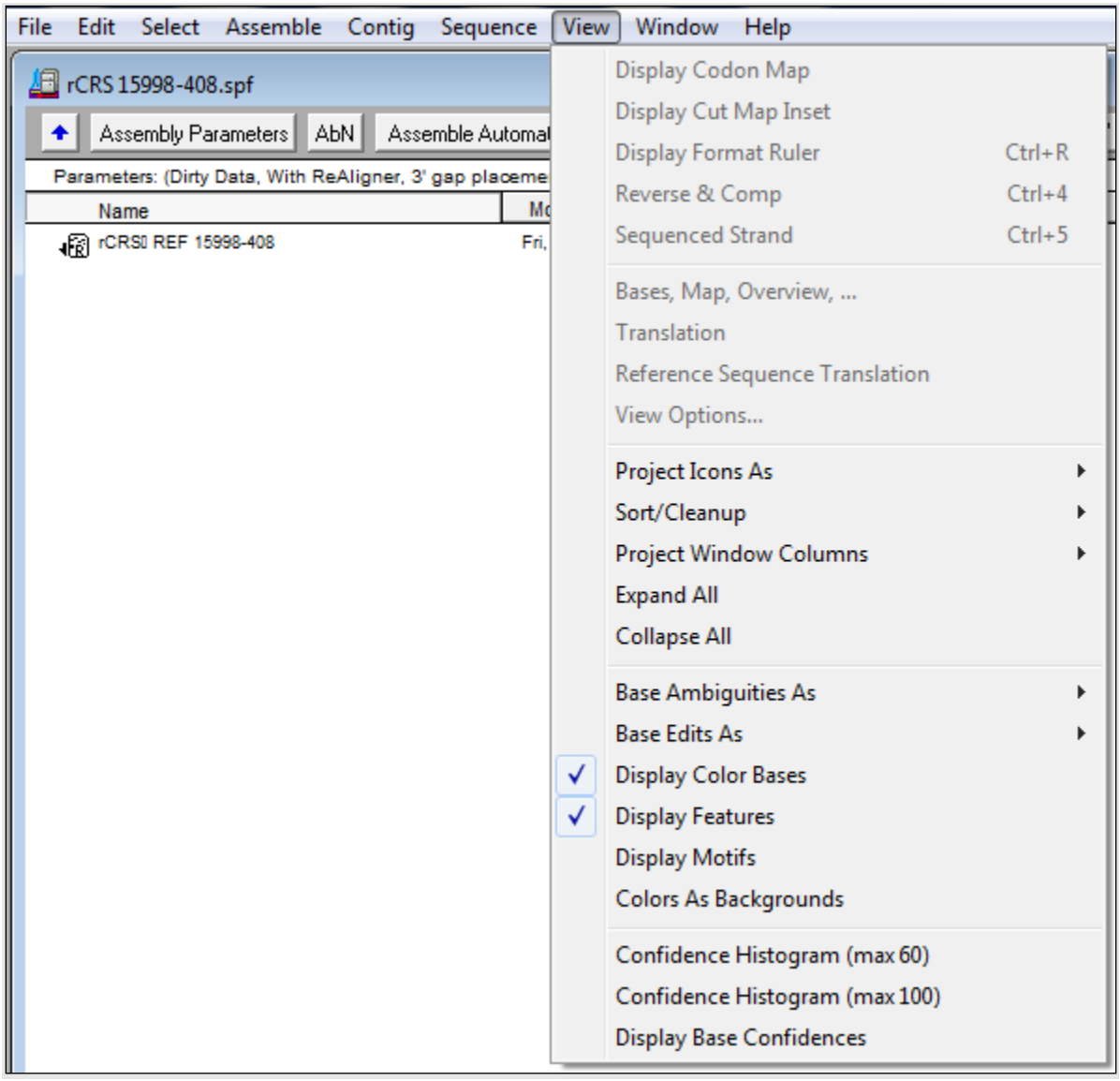

Figure 3 – View settings

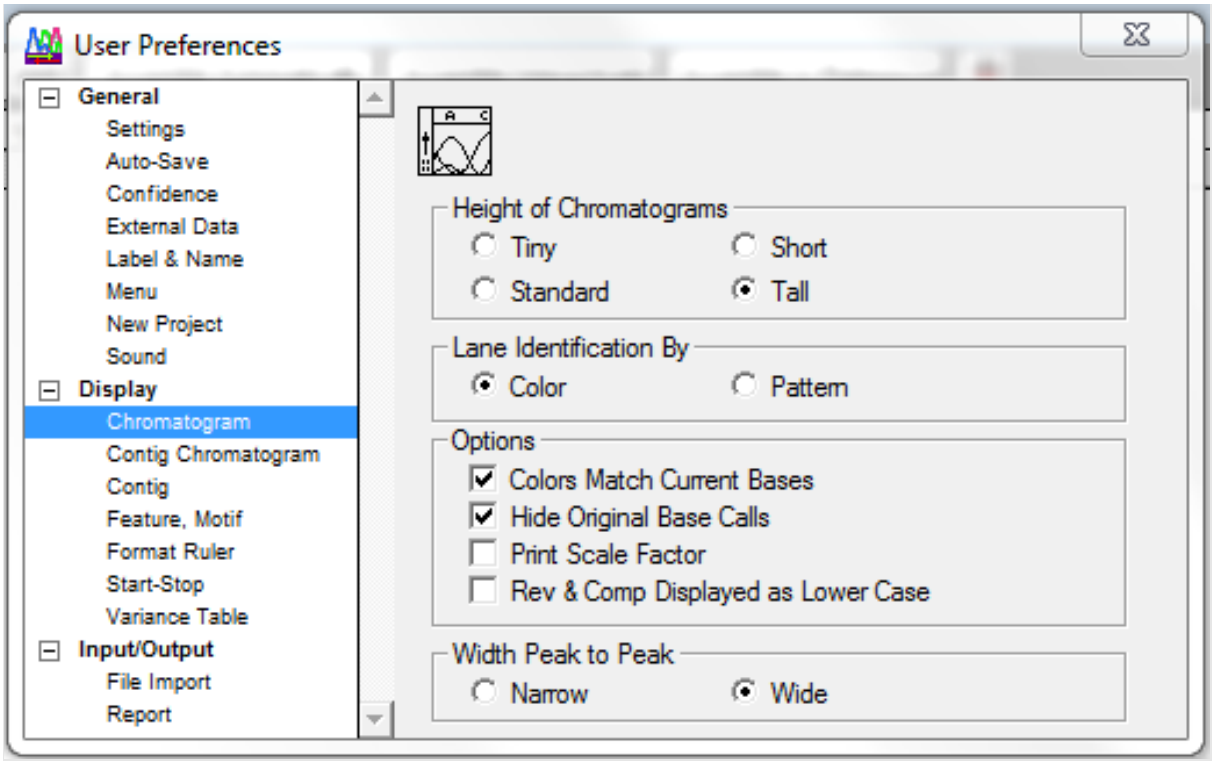

Figure 4 – User Preferences

10 APPENDIX B: 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'

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'