mtDNA Sequence Analysis, Interpretation, and Report Writing

Table of Contents

mtDNA Sequence Analysis, Interpretation, and Report Writing

1 INTRODUCTION

Upon completion of the technical aspects of mitochondrial DNA analysis, the results must be verified and interpreted by an Examiner. The verification of the accuracy of the results involves a review of information generated by the appropriate DNA typing software, as well as an evaluation of quality controls. Following this assessment, the Examiner makes comparisons among samples as applicable and draws conclusions that are captured for documentation and communication purposes within an FBI *Laboratory Report* (7-1 or 7-1 LIMS).

The results are derived through application of the appropriate software during and after sequence detection of amplified DNA that is generated for each specimen using the PowerSeq™ CRM Nested System. Based on a visual comparison of the final DNA typing results, the Examiner may conclude that a person of interest (POI) is excluded as a possible contributor, may be included as a possible contributor, or the results of the comparison are inconclusive. Profiles yielding inclusionary results are searched in a population database to generate an estimate of statistical rarity in major US populations and to generate a statistic that describes the magnitude of the support for the conclusion.

These procedures describe the use of the software packages for mitochondrial DNA (mtDNA) sequence analysis, 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 typing results obtained from the PowerSeq CRM Nested System using the CLC Genomics Workbench and GeneMarker HTS DNA typing software for forensic analysis, perform interpretation, generate statistical analyses, and write reports.

3 EQUIPMENT

- GeneMarkerHTS (Softgenetics®, version 1.2.2 or higher)
- CLC Genomics Workbench (Qiagen, version 12.0.3 or higher)
- PopStats (CODIS, version 11.0 or higher)
- EMPOP MtDNA Database (Current version available at http://empop.online/)

4 STANDARDS AND CONTROLS

The controls used throughout mtDNA analysis are the reagent blank (RB), negative control (AMPBLANK or NC), and positive control (2800M). Refer to Section 5.3 for the interpretation of these controls.

4.1 Extraction Control (i.e., Reagent Blank (RB))

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.

B. At least one RB must be processed in parallel with each batch of samples subjected to a specific extraction process.
C. Amplification or post-amplification reworks must contain RB controls if the reworks are being processed with an increased level of sensitivity (e.g. higher pool concentration or lower number of total pool samples.)

4.2 Negative Amplification Control (i.e., AMPBLANK)

A. The AMPBLANK monitors for the presence of exogenous DNA from the amplification procedure through the sequencing procedure.
B. An AMPBLANK must be processed in parallel with each amplification set or batch of samples.
C. Amplification or post-amplification reworks must contain all associated AMPBLANK controls.

4.3 Positive Amplification Control (i.e., 2800M)

A. The use of a positive control monitors the success of the amplification through sequencing processes. The 2800M cell line DNA is used as the positive control for mtDNA analysis.
B. One positive control must be processed in parallel with each amplification reaction or batch of samples.
C. Amplification or post-amplification reworks must contain a positive amplification control.
D. If the positive control does not exhibit the expected results (see table 1), the positive control may be re-amplified, or re-sequenced, together with all samples that were typed in parallel.
    1. If the re-processed positive control displays the expected mtDNA type, all associated samples that were re-processed in parallel may be interpreted.
    2. If the positive control displays partial results, incorrect results or no results, any samples amplified in parallel with this control may not be interpreted.
E. See Table 1 for the expected positive control typing results obtained using the PowerSeq™ Nested Kit.

Range (np)	2800M
16013-592	152 C 263 G 315.1C 477 C 16519C

Table 1 - Expected DNA Typing Results of the Positive Control 2800M using the PowerSeq™ Nested Kit, expressed as differences from the rCRS

5 PROCEDURE

5.1 Sequencing Analysis

5.1.1 Read filtering using CLC Genomics Workbench

A. The CLC Genomics Workbench software trims raw FASTQ files obtained from the MiseqFGx to remove reads <90 bp that therefore do not span targeted amplicon regions.
B. The CLC Genomics Workbench settings are represented in Appendix A: Settings for CLC Genomics Workbench.

5.1.2 Assembly of Reads Using GeneMarker™ HTS

A. Trimmed sequence reads are assembled using GeneMarker™ HTS.
B. The GeneMarker HTS settings are represented in Appendix B: Analysis Parameters for GeneMarker HTS 1.2.2 and above.
    1. Read depth (i.e., read coverage or read count) is the absolute number of reads of a given base or position
    2. The read coverage threshold/analytical threshold (AT) is 200 reads. Any region, position or variant under the AT is below the level of detection and therefore not considered in profile determination.
    3. The variant percentage threshold is set to 10%. Any variants below 10% of the reads at a position are below the threshold, and are therefore not considered in profile determination.
    4. The stochastic threshold (ST) is the minimum total amplification input where dropout is not expected to occur. This is defined as either:
        i. ≥ 300 total copies with a degradation index ≤ 1 or
        ii. ≥ 1250 total copies with a degradation index > 1

5.2 Interpretation of DNA Typing Results

A. Typically, the profile range of a sample or control is defined as the span of sequence obtained above 200 reads and is denoted by the 5’ and 3’ ends of contiguous amplicon(s).
B. If the profile range of a sample or control only includes data from the amplicon overlap regions, the sequence will not be used.
C. Differences from the rCRS are noted within the GeneMarker HTS™ files within a variant table by noting the nucleotide position followed by the International Union of Pure and Applied Chemistry (IUPAC) code for the polymorphic base (e.g., 263 G) as listed in Table 2.

IUPAC Code	DESIGNATIONS
A	**Adenine**
B	C, G, or T (Not A)
C	**Cytosine**
D	A, G, or T (Not C)
G	**Guanine**
H	A, C, or T (Not G)
K	G or T (**K**eto)
M	A or C (a**M**ino)
N	a**N**y base or combination of bases
R	A or G (pu**R**ine)
S	G or C
T	**Thymine**
V	A, C, or G (Not T)
W	A or T
Y	C or T (p**Y**rimidine)
OTHER DESIGNATIONS	
del	Deletion
X.1(.2,...)	Insertion after position X
lowercase base code	Minor variant deletion

Table 2- Nucleic acid designations

D. 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* (current version):

- **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 as "del."
    - All sequences containing insertions or deletions outside the expected HV1, HV2, or HV3 C-stretch regions, as well as nucleotide positions (np) 249, 290, 291, and between 514 - 525, will be compared to known phylogenetic alignments.
    - For assistance with identifying known patterns of polymorphisms, the sequence should be 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 should be interpreted 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.

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

- **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.

5.2.1 Point or sequence heteroplasmy

Point or sequence heteroplasmy is defined as more than one base at a given position (independent of the insertion/deletion of bases created by length variants) present within a sample.

A. Typically, point heteroplasmy is detected at only one or two positions within the mtDNA control region for a given sample.
B. A point heteroplasmy is designated as more than one base called above the 10% variant threshold at one specific position, excluding insertions or deletions.
C. Positions consistent with point heteroplasmy must be verified within GeneMarker HTS to rule out untrimmed primer artifacts.
    1. If an untrimmed primer creates a false variant, the primer must be trimmed within the GeneMarker HTS project file. This may require changing the amplicon settings.
    2. The amplicon regions within GeneMarker HTS amplicon settings include the start and end (5' and 3' end) positions after primer trimming and before quality trimming within GeneMarker HTS. The corresponding amplicon report lists ranges that are 1 bp smaller on the 5' and 3' ends than the actual

amplicon settings (e.g. amplicon 1 analyzed data, prior to quality trimming, includes 16013-16126, while the corresponding amplicon report range lists 16012-16125).

5.2.2 Mixture Determination

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.

5.2.3 Length heteroplasmy

Length heteroplasmy is defined as more than one length variant present within a sample in a given region.

A. Length variants that differ from rCRS and are present over the 10% variant threshold are displayed within casefile profile documentation.
    1. Because length variants matching rCRS and variants below the variant threshold are not listed within the GeneMarker HTS software, major and minor length variants are not determined.
B. Length variants in HV1 (i.e., 16193), HV2 (i.e., 309) and HV3 (i.e., 573) resulting from homopolymeric C stretch regions, and in HV3 (i.e., AC repeat region) resulting from a dinucleotide repeat region, are commonly observed.
    1. Length heteroplasmy between HV1 positions 16183-16194 most commonly arises when there is a substitution of a C for a T at position 16189. The reference type in HV1 is C5TC4.
    2. Length heteroplasmy in HV2 is observed in two formats:
        i. Length heteroplasmy in HV2 is most often observed between positions 302-310 in the number of C residues preceding a T residue at position 310. The reference type at these locations is C_7TC_5.
        ii. Length variants can also be observed when a T to C transition occurs at 310.
    3. Length heteroplasmy in HV3 is commonly observed in two formats:
        i. HV3 length variants are commonly observed in the number of C residues between positions 568-574.
        ii. Length variants may also be observed in the dinucleotide repeat region between positions 514-525.
            a. The AC repeat is treated as a single unit, with indels grouped immediately 5’ to np 525.
            b. A single repeat insertion is noted as 524.1 A, 524.2 C
            c. A single repeat deletion is noted as 523 del, 524 del.

5.3 Evalution of control sequences

5.3.1 Positive Controls

A. The positive control must type correctly (see Section 4.3) and include np 16024-576, at a minimum, in order to report the sequence for an associated sample.
B. The sequence range for the associated sample(s) may not extend beyond the sequence range for the positive control.

5.3.2 Contamination Assessment

Although appropriate quality assurance practices are stringently applied and enforced, it is not unexpected that lowlevels of adventitious DNA may be detected due to the highly sensitive nature of the amplification and DNA typing process. Because mtDNA is not unique to an individual, a contaminant may not be attributable to a specific source (e.g., laboratory personnel), but a reasonable assumption of logical sources of contamination is attempted. Instances of contamination will be evaluated on an individual basis according to the following guidelines.

A. The typing results of samples associated with an AMPBLANK and/or RB may be used under any of the following conditions:
    1. The RB and/or AMPBLANK has < 200 reads at all positions
    2. The RB and/or AMPBLANK has ≥ 200 reads but < 10% of the read depth for each corresponding position of the associated sample(s).
    3. The RB and/or AMPBLANK has ≥ 200 reads and ≥ 10% of the read depth for a corresponding position of the associated sample(s), but all of the following conditions are met:
        i. The RB and/or AMPBLANK read depth is less than the associated sample(s) read depth at all positions ≥ 200 reads
        ii. The sample amplification reaction included:
            1. ≥ 300 total copies with a degradation index ≤ 1 or
            2. ≥ 1250 total copies with a degradation index > 1
        iii. The RB and/or AMPBLANK typing results are not concordant with the sample(s) in one or more amplicons
B. If the above criteria to allow for the use of the sample with a possible RB or AMPBLANK contaminant are not met, the associated sample(s) may be reprocessed as appropriate at the Examiner's discretion. If reprocessing fails to resolve the issue, the associated sample(s) may not be used.

5.4 Procedure for Sequence Comparisons

5.4.1 Data Used for Comparisons

A. When comparing sequences, only the regions with a common range will be evaluated (e.g., If a partial profile with np 16013-16224 is compared to a profile with np 16013-592, only the portion of the profile range shared in common [i.e., np 16013-16224] will be compared.)

B. Cytosine insertions and/or length variants in homopolymeric regions (e.g., at nucleotide positions 16193, 309, and 573) will not be used for comparison purposes.
C. Cytosine insertions and deletions between np 302-316 in the presence of np 310C are not used for comparison purposes.
D. Insertions and deletions in the AC repeat region between np 514-525 are considered as follows:
    1. In the absence of length variants, insertions and deletions in the AC repeat region are considered for sequence comparisons.
        i. An insertion or deletion is counted as one difference per AC repeat unit. For example, when compared to rCRS:
            a. 524.1A, 524.2 C is counted as one difference
            b. 524.1A, 524.2C, 524.3A, 524.4C is counted as two differences
            c. 523 del, 524 del is counted as one difference
    2. When length variants are present, insertions and deletions in the AC repeat region are not considered for sequence comparisons.
E. DNA mixtures are not interpreted, and therefore are not included in profile comparisons.

5.4.2 Interpretation of Sequence Comparisons

The following conclusions may be determined for sequence comparisons:

5.4.2.1 Exclusion

If the samples differ at two or more positions, they are excluded as coming from the same source or same maternal lineage.

5.4.2.2 Inconclusive

If the samples differ at a single position only, the comparison is inconclusive.

- To account for 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. Samples of known origin may include blood, buccal swabs, and hair. Multiple hair fragments from a hair standard of known origin may be combined and processed as a single known sample.

5.4.2.3 Inclusion/Cannot Exclude

If the samples have the same sequence, or demonstrate sequence concordance, they are included as potentially coming from the same source or same maternal lineage. Sequence concordance is defined as having a common base at each position at which sequence data were obtained in the sample.

- The terms include and cannot exclude are synonymous and can be used interchangeably. Each inclusion interpretation must be clearly and properly qualified with a statistic. For information on reporting statistics, proceed to Section 5.5 (Procedure for Weight Assessment). For other cases, proceed to Section 5.7 (Suggested Reporting Language).

5.5 Procedure for Weight Assessment

5.5.1 Database Searches for weight assessment

A. Database searches are performed when a sample(s) of unknown origin and a sample(s) of known origin are concordant.
B. Database searches may be performed using the EDNAP (European DNA Profiling Group) Mitochondrial DNA POPulation Database (EMPOP, v 4/R13 or higher) or the CODIS Popstats Database (v 11.0 or higher).
    1. Database search rarity estimates are calculated for four general United States population groups (African American, US Caucasian, US Hispanic, and US Asian).
    2. The US Native American population database group is used for specimens that potentially originate from these populations.
    3. Additional population database groups may be searched in a database but not reported.
    4. Additional relevant populations may be included and/or reported on a case-specific basis.
    5. The profile search range will reflect the control region range (16024-576) for profiles that span this range, or greater. For partial profiles, the profile search range will not include positions outside of the partial profile range.
    6. Ambiguous base positions are searched using the appropriate IUPAC symbol (see Table 2), with the following exceptions:
        i. Deletions are entered as "-" following the base position in PopStats (e.g. "247-").
        ii. Lowercase base designations may be entered into EMPOP to include minor variant deletions.

5.5.2 Special Considerations

A. Sequence polymorphisms are included in database searches; however, all length variants at nucleotide positions 16193, 309, and 573 are ignored. Hence, length variability in these regions will not add any additional rarity to a database profile search.
B. Cystosine insertions and/or deletions between np 302-316 in the presence of np 310C are ignored in database searches.
C. When length variants are present in the AC repeat region between np 514-525, they will be ignored in database searches.
    1. E.g., if length variants 524.1A 524.2C and 524.1A 524.2C 524.3A 524.4C are present in a sample, length variants will be excluded by omitting them from the searched range (e.g., begin – 524; 525-end).
    2. Alternate database search methods that ignore insertions at 524 may also be used.

5.5.3 *Searching Profiles in the European MtDNA POPulation Database (EMPOP)*

A. The primary database used for estimating the haplotype frequency of a single mtDNA profile is the EMPOP database (http://empop.online), which is used to provide likelihood ratios from frequency estimates for African American, Caucasian, Hispanic, and other populations, when applicable.
B. EMPOP may be used to estimate profile rarity for relevant population groups using the settings represented in Appendix D: EMPOP.
C. Relevant metapopulations are selected, if necessary, in the "Find metapopulation..." box.
D. Lowercase nucleotide designations can be entered into EMPOP to designate minor deletion variants. These entries require the "Use extended IUPAC code" selection in the Query tab.
E. If a one-tailed Clopper and Pearson Upper Bound Frequency Estimate (UBFE) is not provided within the EMPOP search, it will be calculated using the formula in Section 5.6, and the value entered into the report table.

5.5.4 *Searching Profiles in the CODIS Popstats Database*

CODIS 11.0 or higher is used to search casework profiles in the database using the settings represented in Appendix C: Analyst Workbench – CODIS Popstats.

A. Deletions are searched using a "-".
B. Lowercase designations are not entered into the CODIS Popstats database.

The CODIS Popstats database search result provides the number of database profiles that match the casework profile and the one-tailed Clopper and Pearson 95% Upper Bound Frequency Estimate. 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.

5.6 Statistical Calculations

5.6.1 *Upper Bound Frequency Estimate (UBFE)*

A. For instances in which the haplotype has been observed in the database (x > 0), the one-sided 95% confidence interval is calculated using the Clopper and Pearson formula.

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

Where k = 0,1,2,3, ... x observations; n = database size, and x = number of observations of the haplotype in the database. This 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 and n is the number of profiles in that population. This equation will require a computer to solve.

B. For instances where the profile has not been observed in the database (x=0), the following formula is used to estimate the upper bound of the 95% confidence limit, p_0:

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

Where α = 0.05 or 5% and n = database size. For example, if a particular haplotype is not observed in a database of 2000 haplotypes, then the 95% upper confidence limit is estimated to be $1-(0.05)^{1/2000} = 0.0014977$ or 1 in every 667 haplotypes. Note that when α = 0.05, p_0 is very close to 3/n (e.g., 3/2000 = 0.0015).

C. The highest UBFE from the relevant US population groups searched will be chosen for reporting purposes, and converted into a percentage (UBFE*100).
   1. Generally for EMPOP searches, the relevant populations are the African American, US Caucasian, and US Hispanic populations.
   2. Generally for CODIS searches, the relevant populations are the African American, US Caucasian, US Hispanic, and (when applicable) Native American populations.

5.6.2 Likelihood Ratio

A. The likelihood ratio (LR) of the upper bound frequency estimate describes how much more likely the results are if the matching profiles are from the same source or the same maternal lineage rather than if the profile is from an unknown, unrelated individual or lineage. The LR is calculated as the reciprocal of the UBFE:

$$\text{LR} = \frac{1}{95\%\ \text{UBFE}}$$

B. The LR calculated from the highest UBFE from the relevant US population groups searched will be chosen for reporting purposes.

5.7 Suggested Reporting Language

The results and/or conclusions for specimens subjected to DNA analysis will generally be reported in narrative form. The formatting and administrative information required in a *Laboratory Report* are described in the appropriate FBI Laboratory level 1 documents and applicable DNA procedure (i.e., BIO-500). Additional administrative statement examples (e.g., comparisons to previously reported results) can be found in the STR Interpretation procedure (i.e., BIO-570).

5.7.1 Introductory Statement

The report must indicate (a) that the items were subjected to DNA typing, (b) that mtDNA technology was used in DNA analysis, and (c) specify which Amplification Kit was used. The report should contain the item listing followed by the general introductory statement with the appropriate associated endnote.

The items listed above were subjected to mitochondrial deoxyribonucleic acid (mtDNA) typing.[A or B]

5.7.2 Results of Examinations Section

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. See Section 5.7.3 for endnote language, denoted as A-G in these examples.

Results and conclusions are generally preceded by a description of the item tested and/or, where appropriate, the sampling tested from an item.

5.7.2.1 Typing Results Statement

An initial statement will be reported for each item when mtDNA results are obtained. Alternative statements for other types of results (e.g., partial, no results, mixture, missing persons) are listed in the sections below.

Mitochondrial DNA typing results were obtained from item 1.

5.7.2.2 Inclusion (Full Sequence)

For concordant samples of unknown origin and known origin, the Report of Examination will include a table (see below) containing the results of the applicable database search and likelihood ratio (LR) calculation. When database searches have been conducted for multiple items in a case and have the same LR results, a single table may be incorporated into the report.

5.7.2.2.1 Database search results, US populations

Mitochondrial DNA typing results were obtained from item 1. The mtDNA typing results from item 1 are 900 times more likely if DOE is the contributor than if an unknown, unrelated person is the contributor.[C]

Person of Interest (POI)	Upper Bound Frequency Estimate (UBFE)[D]	Likelihood Ratio (1/UBFE)[E]	Level of Support[G]
DOE	0.11%	900	Moderate Support for Inclusion

5.7.2.2.2 EMPOP search results, OCONUS populations

Typically, a database search of US populations will be conducted along with an EMPOP search for other relevant populations. Therefore, the wording from section 5.7.2.2 would be used in addition to the following (Note: The subpopulation groups are case dependent):

Mitochondrial DNA typing results were obtained from item 1. A search of the European DNA Profiling Group Mitochondrial DNA Population Database (EMPOP; currently available online at http://empop.online) was performed to provide additional information regarding population groups relevant to this case.

The mtDNA typing results from item 1 are 900 times more likely if DOE is the contributor than if an unknown, unrelated person is the contributor.[C]

Geographic Affiliation	Upper Bound Frequency Estimate (UBFE)[F]	Likelihood Ratio (1/UBFE)[E]	Level of Support[G]
ASIA	0.10%	900	Moderate Support for Inclusion
South East Asia			
Philippines	0.10%	900	-
Indonesia	0.05%	400	-
Malaysia	0	123	-
Thailand	0	190	-
Vietnam	0	187	-
East Asia			
China	0	544	-
Japan	0	402	-
Korea	0	692	-

5.7.2.3 Inclusion (Partial Sequence)

Partial mtDNA typing results were obtained from item 1 (nucleotide positions 16013-16126 and 136-257 only). The mtDNA typing results obtained from item 1 and DOE are the same across the sequence range obtained in common to both samples.

- Include the LR statement and database search results table as in section 5.7.2.2.

Note: A limited sequence range will only be indicated in the report if it is less than the minimum range of the whole control region (i.e., less than 16024-576).

5.7.2.4 Concordance (Sequence Heteroplasmy)

Mitochondrial DNA typing results were obtained from item 1. 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. The sequences obtained from item 1 and DOE are concordant.

- Include the database search results table and explanatory wording as in section 5.7.2.2.

5.7.2.5 Exclusion

Mitochondrial DNA typing results were obtained from item 1. DOE is excluded as a potential contributor to the mtDNA typing results obtained from item 1.[H]

5.7.2.6 Inconclusive (One Base Difference)

Mitochondrial DNA typing results were obtained from item 1. 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 included or excluded as the contributor of item 1.[I]

5.7.2.7 Evidentiary Samples Concordant

Mitochondrial DNA typing results were obtained from items 1 and 2. The mtDNA results from items 1 and 2 are x times more likely if items 1 and 2 are from the same contributor, or maternally related contributors, than if items 1 and 2 are from unrelated contributors.

- A database search will be conducted in such cases.

5.7.2.8 Mixture

The mitochondrial DNA typing results 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.

5.7.2.9 No Sequence Obtained

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

5.7.2.10 Not of Requisite Quality

The mitochondrial DNA typing results obtained from item 1 were not of the requisite quality for comparisons or interpretations to be conducted.

5.7.2.11 CODIS Statements

A statement must be included in the report of DNA examinations that indicates when an item's typing results will be initially entered into the CODIS or other appropriate databases. Alternatively, a statement is typically added when no results are entered into CODIS. CODIS information should generally be reported using the following statements:

5.7.2.11.1 Missing Person Reference Sample

Mitochondrial DNA typing results were 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.

5.7.2.11.2 Missing Person Unknown Origin Sample

Mitochondrial DNA typing results were 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.

5.7.2.11.3 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.

5.7.2.11.3.1 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].

- A database search will be conducted in such cases.

5.7.2.12 Final Sentence

As appropriate, the final sentence in the Results of Examinations section will be:

No other mtDNA examinations were conducted.

5.7.3 Methods/Limitations Section and Associated Endnotes

5.7.3.1 If no Comparison is Conducted

[A] *Mitochondrial DNA is analyzed using the Promega PowerSeq™ CRM Nested System for Next Generation Sequencing (NGS). DNA is extracted from each sample and portions of the control region 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.*

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.

5.7.3.2 If a Comparison is Conducted

[B] *Mitochondrial DNA is analyzed using the Promega PowerSeq™ CRM Nested System for Next Generation Sequencing (NGS). DNA is extracted from each sample and portions of the control region 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. The annotated*

profiles for all of the samples are then compared. Matching profiles may be searched against an mtDNA population database to provide an upper bound frequency estimate (UBFE) and likelihood ratio (LR).

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.

[C] *If the samples have the same sequence, or demonstrate sequence concordance, they are included as potentially coming from the same source or same maternal lineage. Sequence concordance is defined as having a common base at each position at which sequence data were obtained in the sample. Barring mutation, any relative within the same maternal lineage has the same mtDNA profile and would also be expected to be included as a possible contributor.*

[D] *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 (UBFE). Calculations were performed using the African American, Caucasian, and Hispanic populations in the EMPOP/CODIS population database (EMPOP Rxx/CODIS v.XX). The highest calculated population upper bound frequency estimate is reported. Mitochondrial DNA profiles were placed into the US 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.*

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

[E] *The likelihood ratio is a statistical approach that compares the probabilities of observing the DNA results under two alternative propositions and is calculated by taking the inverse of the highest untruncated upper bound frequency estimate.*

[F] *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 (UBFE). Mitochondrial DNA profiles were placed into the US population groups based on self-identification by the donor (first table). Other population groups within the EMPOP database are separated based on geographic affliliation (second table). 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.

[G] *These likelihood ratio (LR) ranges provide the following support for mtDNA conclusions:*

Likelihood Ratios:	Qualitative Equivalent:
≥ 1,000,000	Very Strong Support for Inclusion
10,000 to < 1,000,000	Strong Support for Inclusion
100 to < 10,000	Moderate Support for Inclusion
2 to < 100	Limited Support for Inclusion
1	Uninformative
> 1/100 to ½	Limited Support for Exclusion
≤ 1/100	Exclusion

[H] *If samples differ at two or more nucleotide positions, they are excluded as coming from the same source. Barring mutation, any relative within the same maternal lineage has the same mtDNA profile and would also be expected to be excluded as a possible contributor.*

[I] *If samples differ at only a single nucleotide position, no conclusion can be reached as to whether they originate from the same source.*

5.7.4 Remarks Section

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

5.7.4.1 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. Additionally, this report conforms to the FBI Approved Standards for Scientific Testimony and Reporting for Haplotype DNA Testing. 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.

5.7.4.2 Consumption of Evidence

Item 1 was consumed during the DNA Casework Unit examinations.

5.7.4.3 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 and returned to the contributor if requested.

5.7.4.4 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.

5.7.4.5 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.

5.7.4.6 Requests Not Conducted

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

5.7.4.7 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 and 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 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 5.6.1.

A likelihood ratio is calculated to express the likelihood of observing the evidence profile under alternative propositions, as described in Section 5.6.2.

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.
- 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 result that may be encountered by the Examiner, nor the conclusions that an Examiner may render based on their 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.
- 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

Parson and Dür (2007) EMPOP – A forensic mtDNA database. For Sci Int 1(2): 88-92.

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

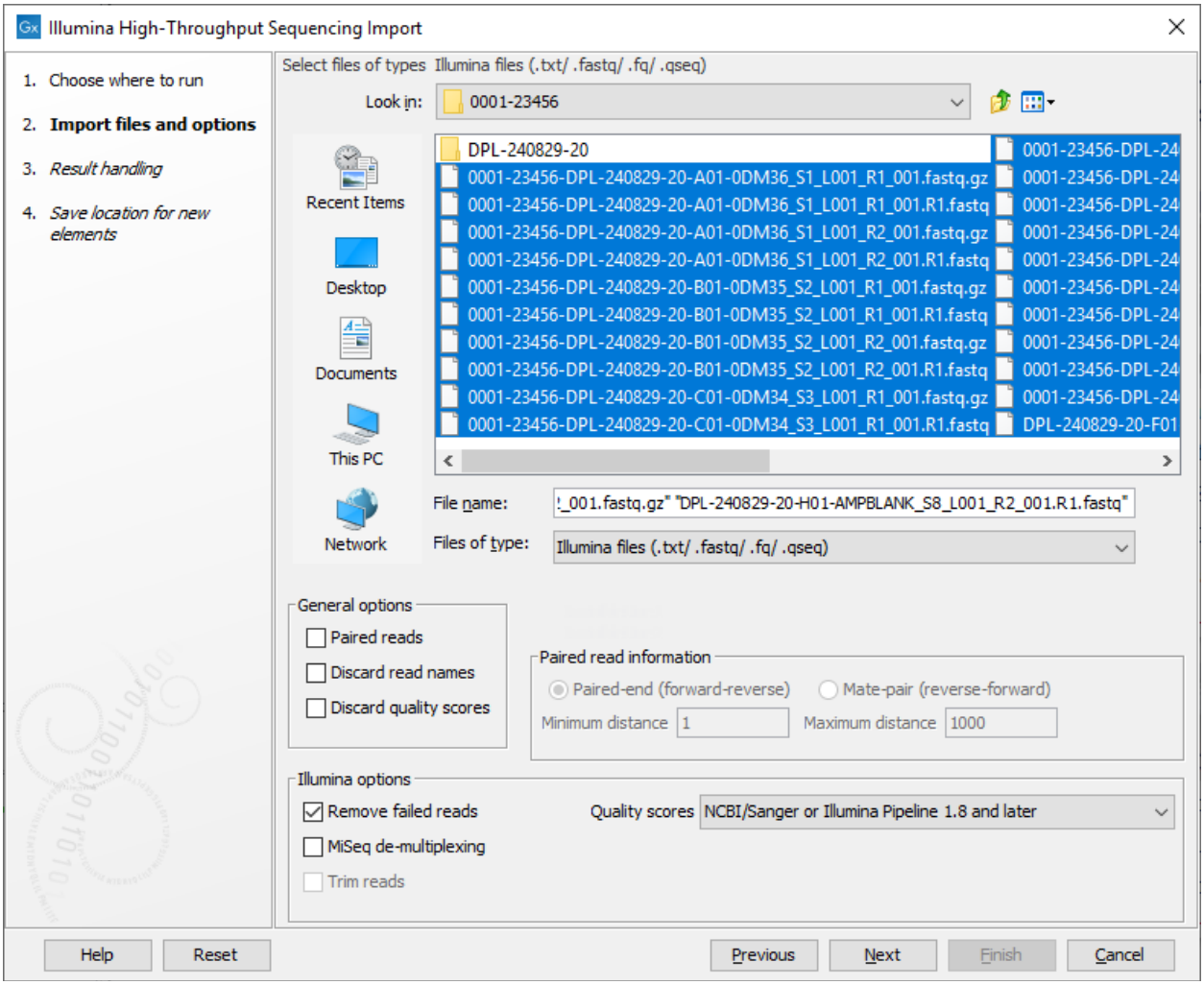
SWGDAM, Supplemental Information for the SWGDAM Interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories, most recent version, www.swgdam.org.

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.
01	12/16/2024	Updates for NGS including thresholds, interpretation software instructions, interpretation for length variants. Updated interpretation parameters for positive and negative controls. Added likelihood ratio calculations and revised report wording. Changed upper bound frequency estimate calculation table to report highest UBFE (lowest LR). Changed primary database for profile weight assessment to EMPOP. Updated Appendix for NGS computer program settings. Sanger Sequencing interpretation info moved to new mito Legacy Interpretation SOP BIO-575.

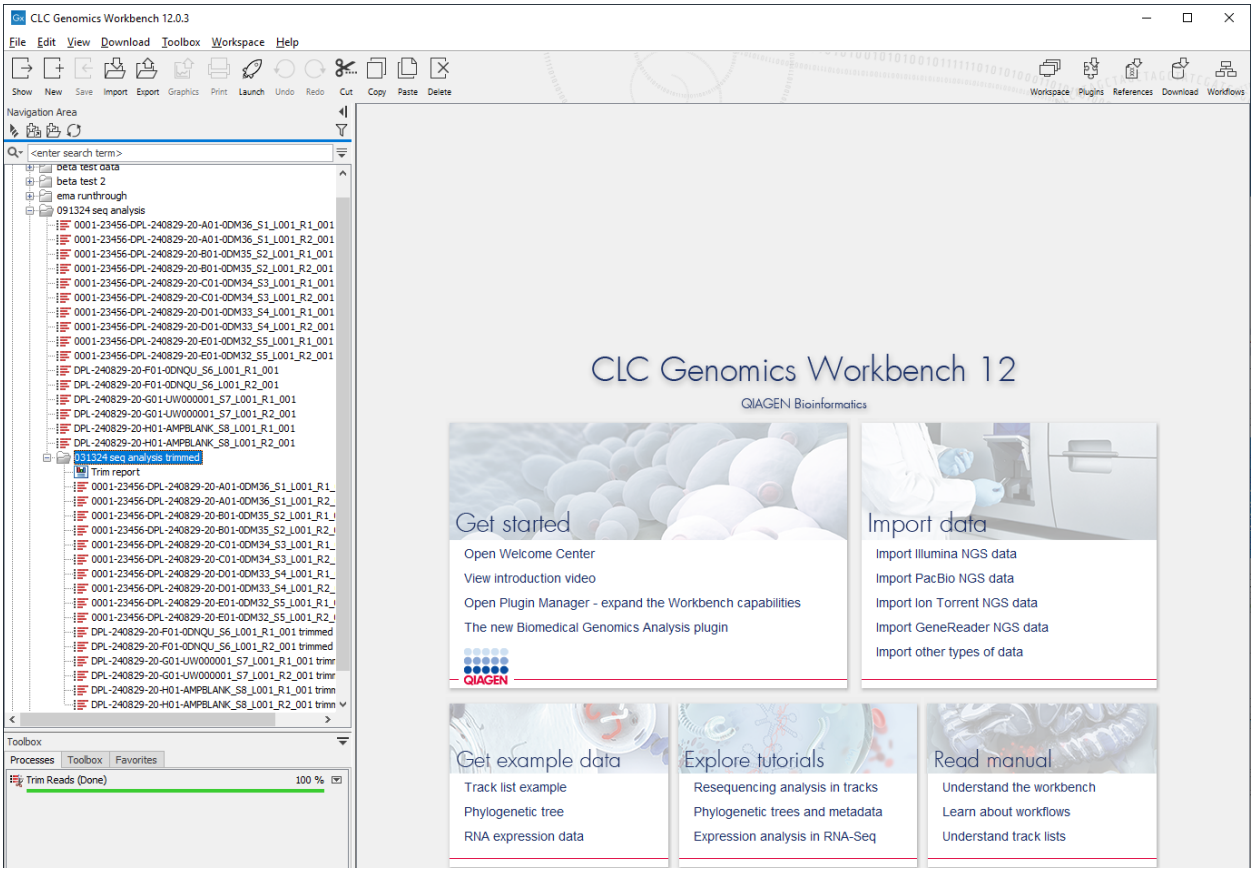
10 APPENDIX A: SETTINGS FOR CLC GENOMICS WORKBENCH

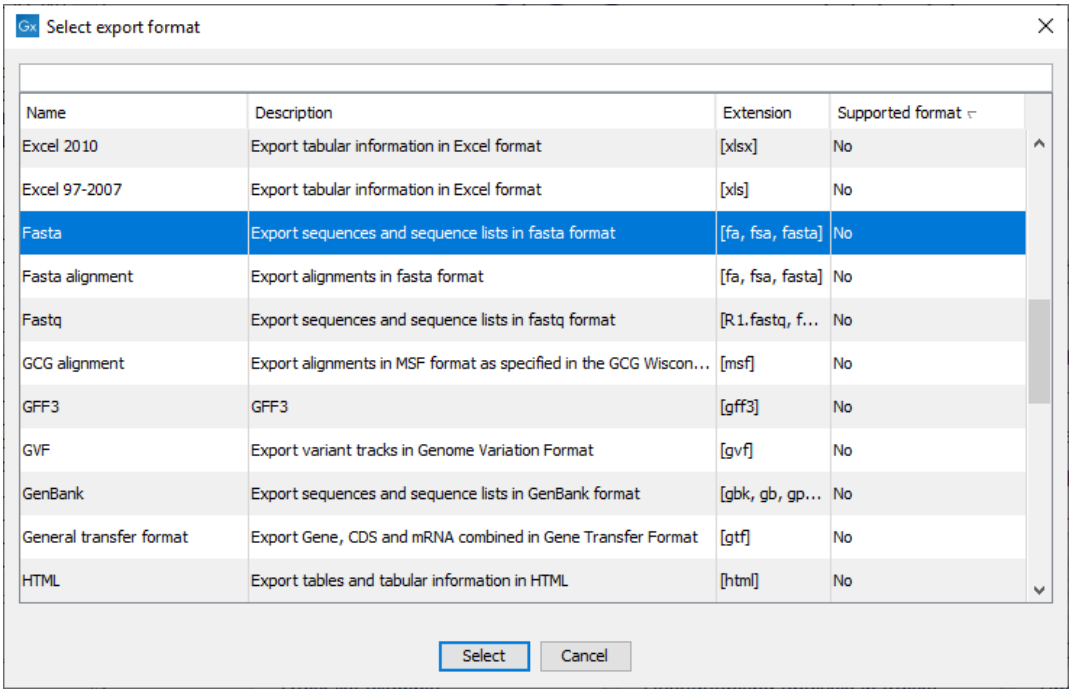
Sequence file import settings



APPENDIX A: SETTINGS FOR CLC GENOMICS WORKBENCH

Export trimmed file settings





11 APPENDIX B: ANALYSIS PARAMETERS FOR GENEMARKER HTS 1.2.2 AND ABOVE

New Project dialog box settings	
Option	
Use Default Reference	Selected - This uses the revised Cambridge reference sequence (rCRS) that is supplied with GeneMarker HTS
Use Default Motif	Not Selected - A path is specified to a motif file with current phylogenetic alignment rules

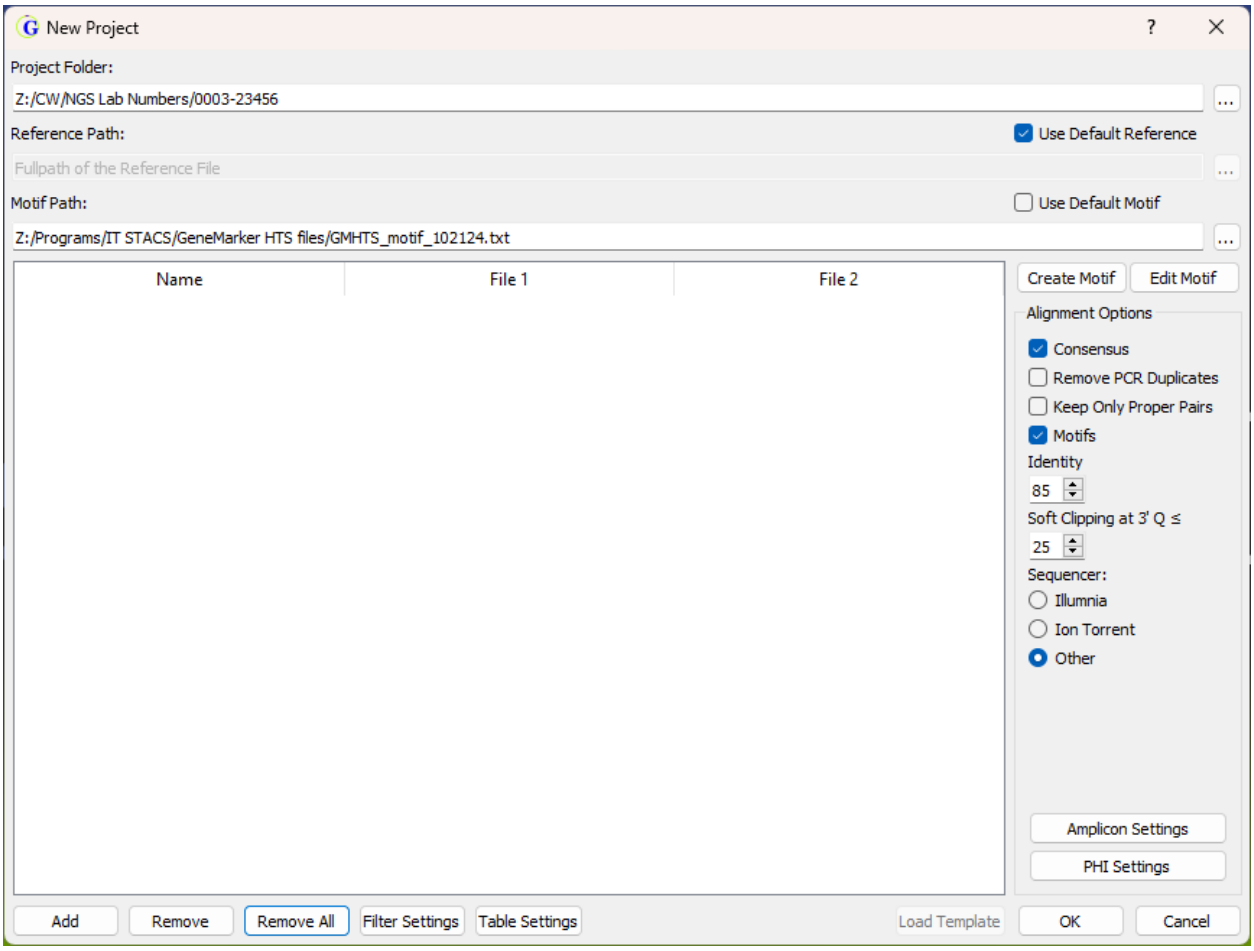
Alignment Options	
Option	
Consensus	Selected
Remove PCR Duplicates	Not Selected
Keep Only Proper Pairs	Not Selected
Motifs	Selected
Identify	85
Soft Clipping at 3' Q ≤	25
Illumina	Not Selected
Ion Torrent	Not Selected
Other	Selected

Filter Settings dialog box, Variant Filters tab	
Option	**Updated Setting**
Variant Percentage	10.00%
Variant Allele Coverage	***200***
Total Coverage	***200***
Allele Score Difference	Selected; Set to 10
Allele Balance Ratio SNP Indel	Selected Set to 2.5 Set to 5.0

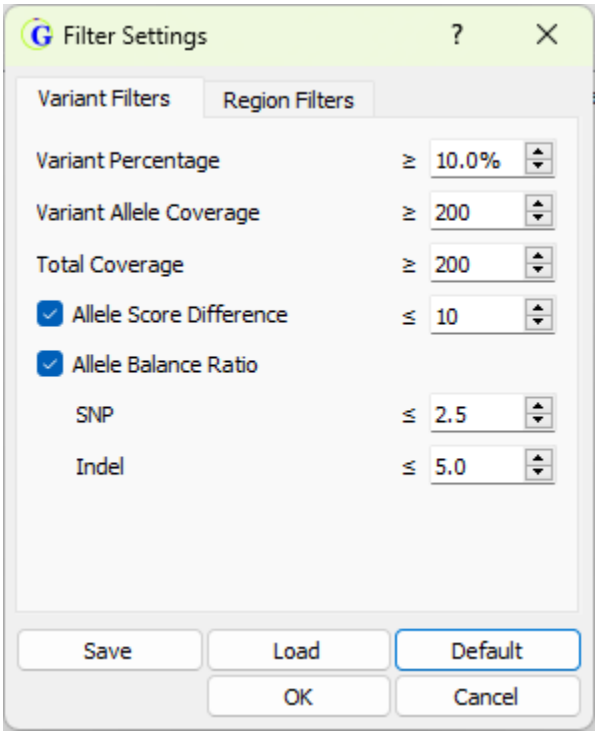
Filter Settings dialog box, Region Filters tab	
Option	
Entire Reference or Input Region	Input Region: Added two regions 1st region: Start=16013 End=16569 2nd region: Start=1 End=592

Amplicon Settings dialog box	
Option	**Updated Setting**
Select Amplicon Settings Button	Open by Selecting and added 10 amplicon regions: 16013-16126, 16116-16225, 16223-16408, 16387-16486, 16474-30, 16555-152, 136-257, 246-364, 342-436, and 429-592

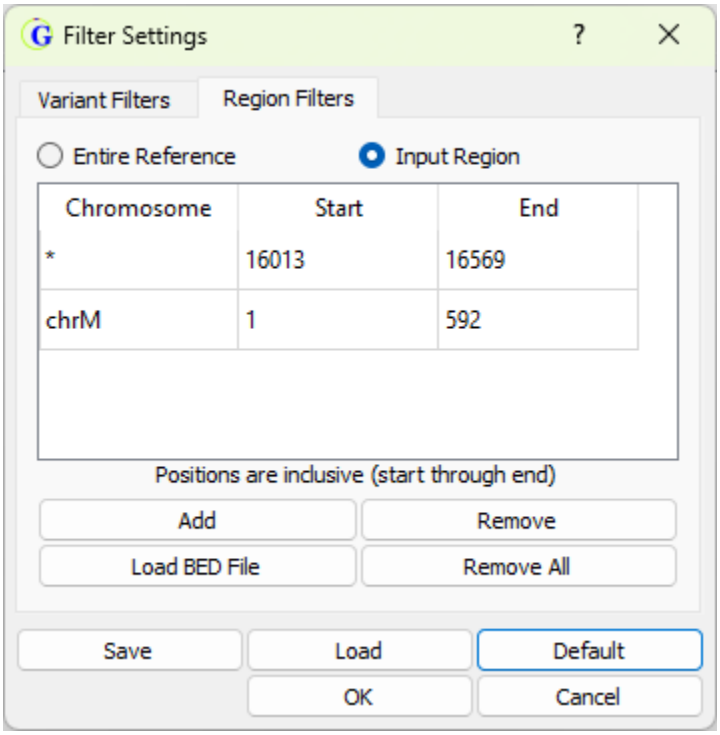
Project dialog box settings:



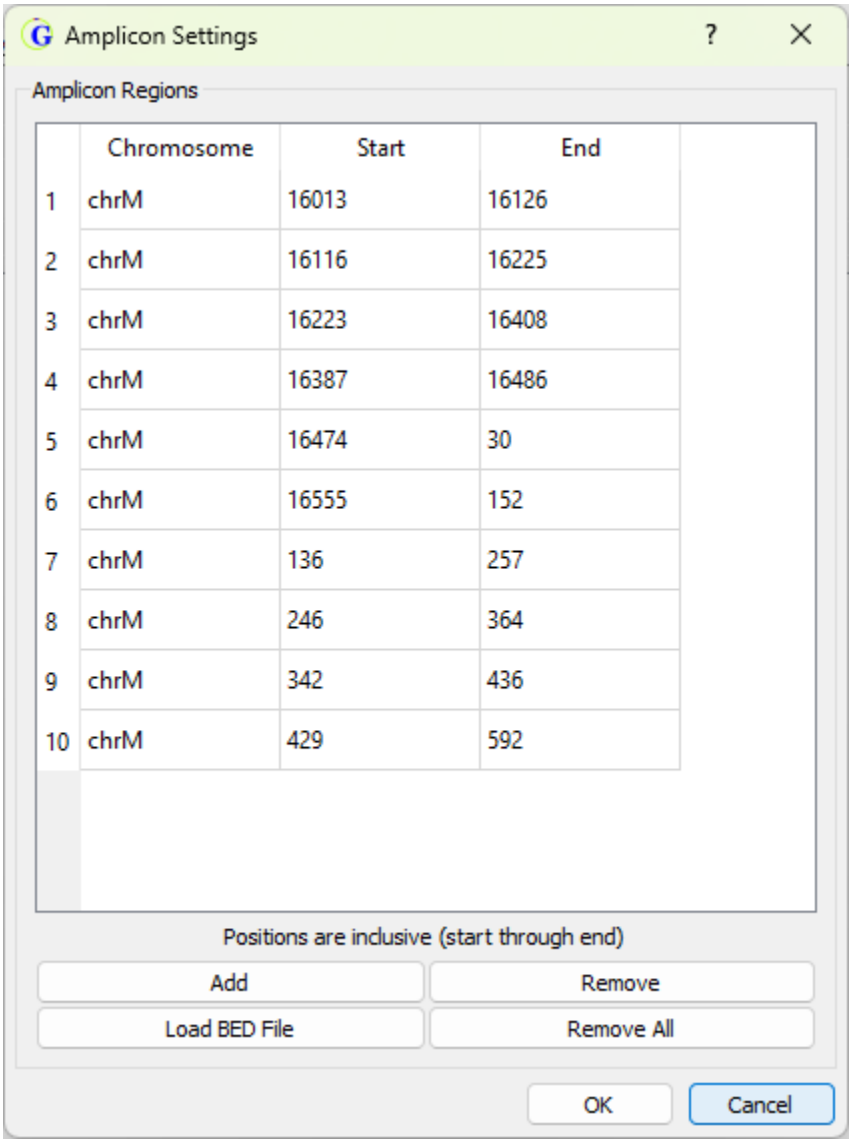
Filter Settings dialog box, Variant Filters tab



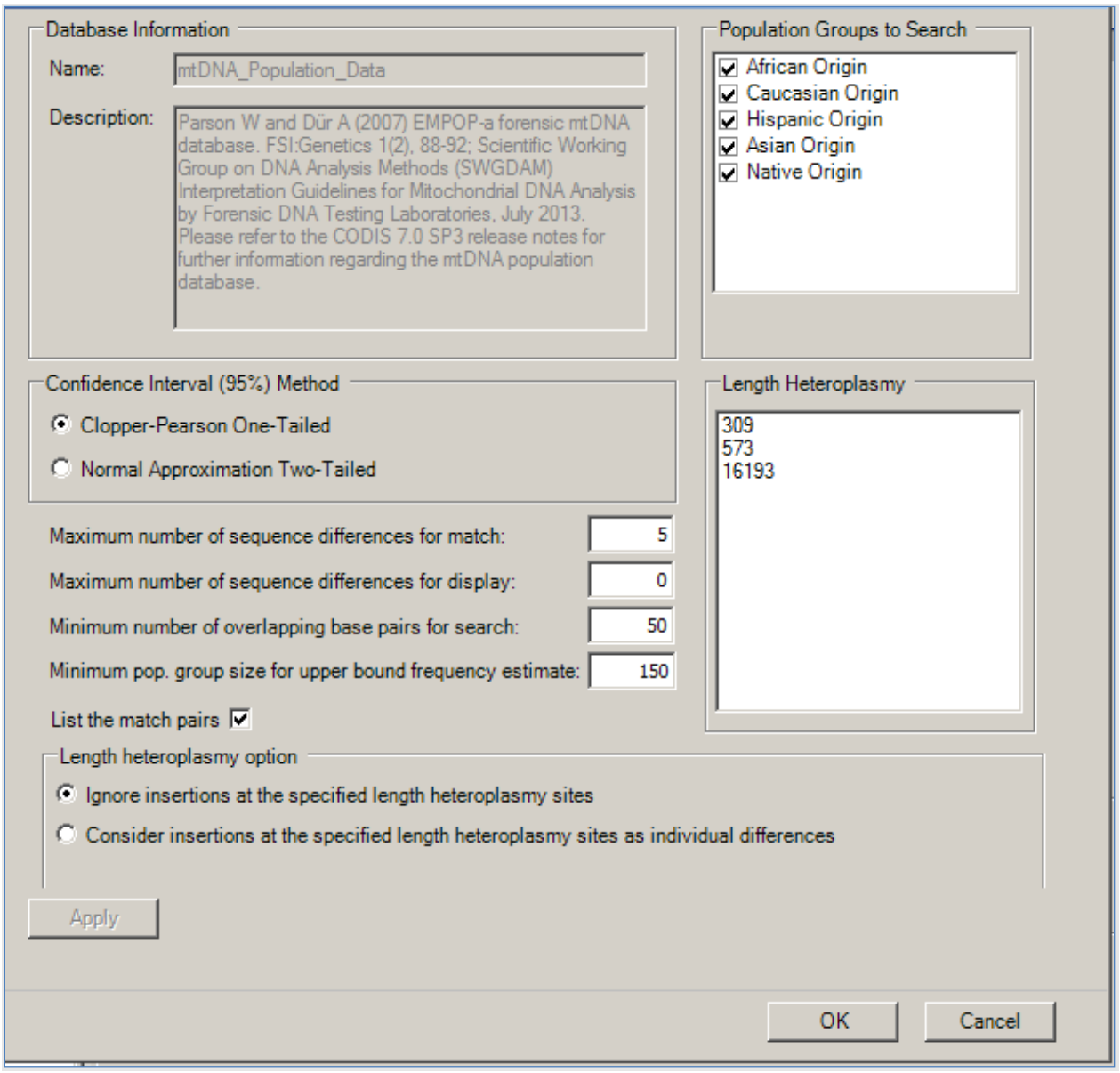
Filter Settings dialog box, Region filters tab



Amplicon Settings dialog box:



12 APPENDIX C: ANALYST WORKBENCH – CODIS POPSTATS



Default settings for CODIS mtDNA population search

13 APPENDIX D: EMPOP

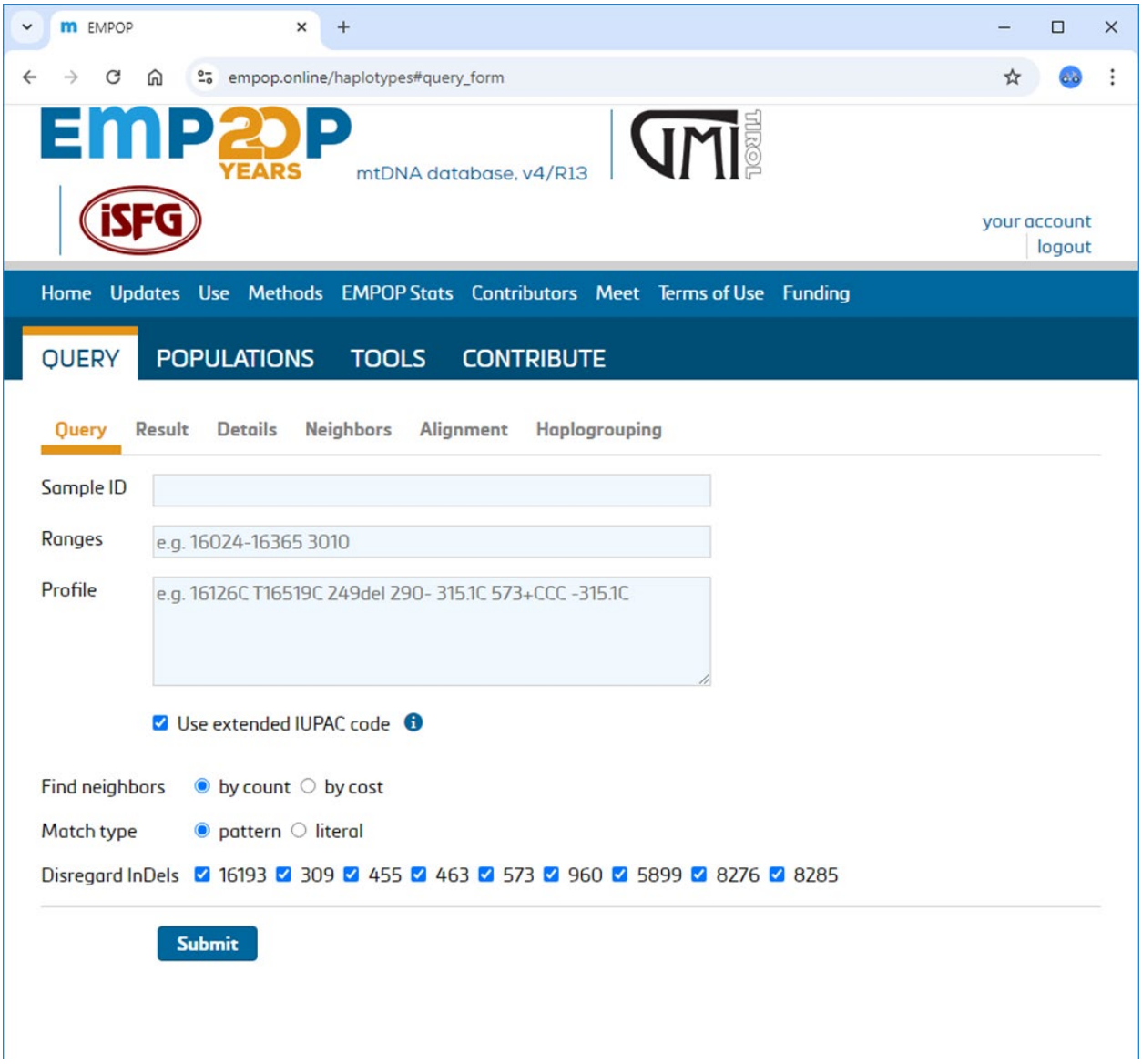


Figure 1 - Settings for EMPOP database search

APPENDIX C: EMPOP (CON'T)

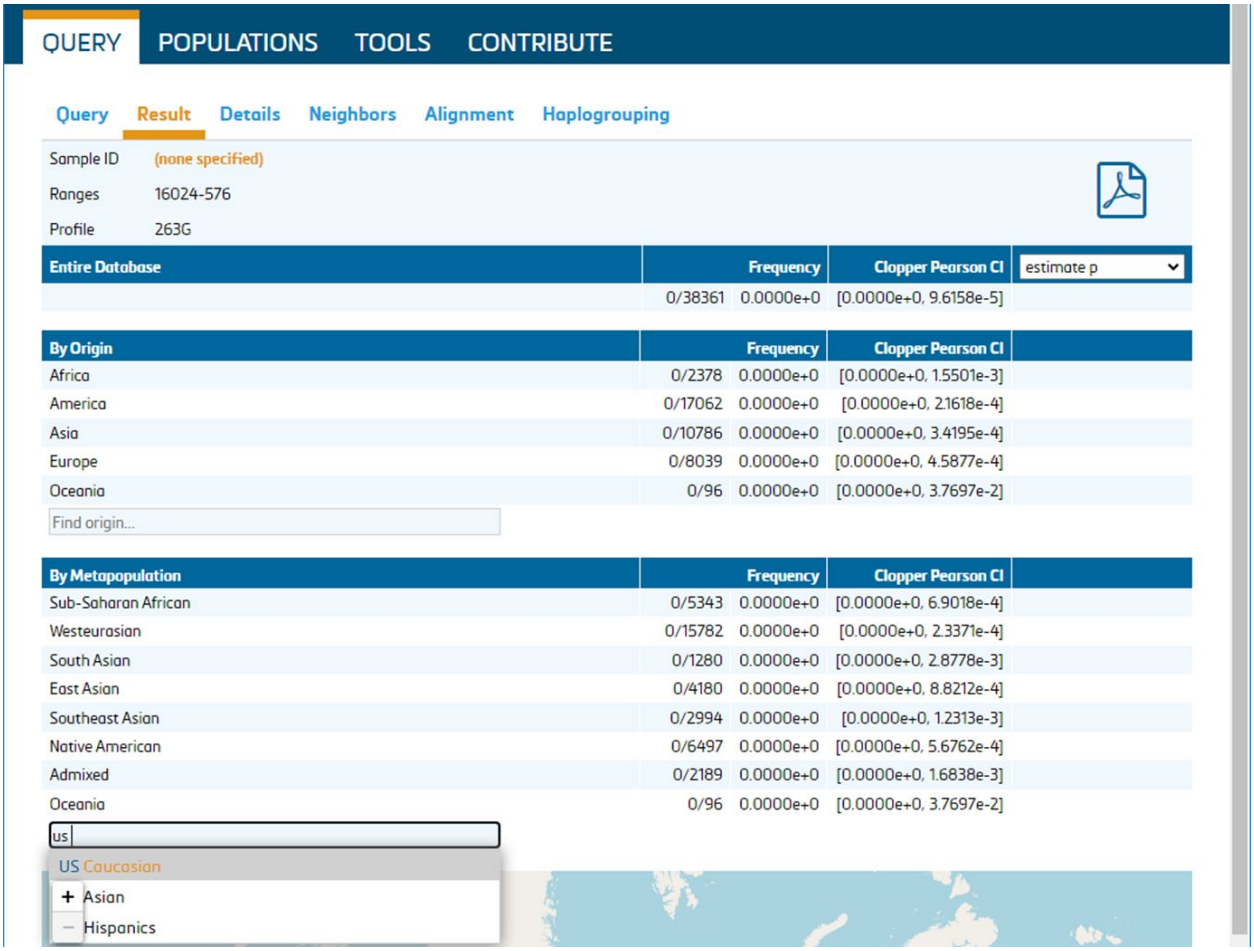


Figure 2 - EMPOP Search Results Screen