

SEQUENCHER®

Tutorial for Windows and Macintosh

Mitochondrial DNA Typing

© 2011 Gene Codes Corporation

Gene Codes Corporation

T C A G E N E
A G T C O D E S

Gene Codes Corporation
775 Technology Drive, Ann Arbor, MI 48108 USA
1.800.497.4939 (USA) +1.734.769.7249 (elsewhere)
+1.734.769.7074 (fax)

www.genecodes.com gcinfo@genecodes.com

Mitochondrial DNA Typing

Open a Project Template	3
Import Sample Data	4
Comparing the Data	4
Validate mtDNA Profiles.....	7
Create an Analyst Report	7
Export mtDNA Profile	8
Evaluate Multiple Samples	9

Mitochondrial DNA Typing Tutorial

There are two versions of Sequencher, the standard version, which has become the favorite for sequence analysis since its release in 1991, and the forensic version. The forensic version of Sequencher has specific functionality to meet the needs of forensic scientists analyzing mitochondrial (mtDNA) sequence. It is the standard for mtDNA analysis in forensic laboratories across the country.

An mtDNA profile is a list of the differences between the mtDNA sequence of a forensic sample and the sequence of a standard reference. The accepted standard sequence for comparison is the revised Cambridge Reference Sequence, rCRS. In most instances the entire mitochondrial genome is not used for comparison, but just the region around the D-loop (16024-576), the control region that does not code for proteins. This region, which crosses the origin of replication, includes 1122 bases. The region around the D-loop has a high degree of variation, particularly in the two hyper variable regions, HV1 (16024–16365) and HV2 (73–340), and so is useful in forensic science.

This tutorial builds on the more generic Reference Sequence Tutorial. Please run through the Reference Sequence Tutorial first if you are not yet familiar with the applications covered there. The Mitochondrial DNA Typing Tutorial begins with forensic samples that have already been assembled and edited. You are now ready for validation and reporting of your results.

This tutorial is structured to give you an overview of the tools available for mtDNA sequence analysis in a forensic environment. It is not a substitute for standard protocols.

The samples in this tutorial consist of four separate Sequencher projects. They are two samples, 90 and 94, prepared by two analysts. In this tutorial we will:

- **Confirm the reported differences in the Variance Table.**
- **Validate the reports from the two analysts.**
- **Create an export of the mtDNA profiles in a format that is compatible with the CODIS data format.**

OPEN A PROJECT TEMPLATE

- **Launch Sequencher.**
- **From the File menu, choose New Project From Template > rCRS.**

Sequencher opens a new project that contains the user preferences, sequences, and contigs contained in the Template Project rCRS. You can see in the Comments section of the Project window that rCRS is the revised Cambridge Reference Sequence. The sequence has already been configured to correspond to the numbering of the mitochondrial genome around the HV1 and HV2 regions.

- Double-click on the rCRS sequence icon to open the sequence.

Residue : 16024	(Sequenced Strand)					Sequence Length : 1122
16024	<u>TTCTTTCATG</u>	<u>GGGAAGCAGA</u>	<u>TTTGGGTACC</u>	<u>ACCCAAGTAT</u>	<u>TGACTCACCC</u>	
16074	<u>ATCAACAACC</u>	<u>GCTATGTATT</u>	<u>TCGTACATTA</u>	<u>CTGCCAGCCA</u>	<u>CCATGAATAT</u>	
16124	<u>TGTACGGTAC</u>	<u>CATAAATACT</u>	<u>TGACCACCTG</u>	<u>TAGTACATAA</u>	<u>AAACCCAATC</u>	
16174	<u>CACATCAAAA</u>	<u>CCCCCTCCCC</u>	<u>ATGCTTACAA</u>	<u>GCAAGTACAG</u>	<u>CAATCAACCC</u>	
16224	<u>TCAACTATCA</u>	<u>CACATCAACT</u>	<u>GCAACTCCAA</u>	<u>AGCCACCCCT</u>	<u>CACCCACTAG</u>	
16274	<u>GATACCAACA</u>	<u>AACCTACCCA</u>	<u>CCCTTAACAG</u>	<u>TACATAGTAC</u>	<u>ATAAAGCCAT</u>	
16324	<u>TTACCGTACA</u>	<u>TAGCACATTA</u>	<u>CAGTCAAATC</u>	<u>CCTTCTCGTC</u>	<u>CCCATGGATG</u>	
16374	<u>ACCCCCCTCA</u>	<u>GATAGGGGTC</u>	<u>CCTTGACCAC</u>	<u>CATCCTCCGT</u>	<u>GAAATCAATA</u>	
16424	<u>TCCCGCACAA</u>	<u>GAGTGCTACT</u>	<u>CTCCTCGTC</u>	<u>CGGGCCCAT</u>	<u>ACACTTGGGG</u>	
16474	<u>GTAGCTAAAG</u>	<u>TGAAGTGAT</u>	<u>CCGACATCTG</u>	<u>GTTCTACTT</u>	<u>CAGGGTCATA</u>	
16524	<u>AAGCCTAAAT</u>	<u>AGCCCACACG</u>	<u>TTCCCCTTAA</u>	<u>ATAAGACATC</u>	<u>ACGATGGATC</u>	
5	<u>ACAGGTCTAT</u>	<u>CACCCATTA</u>	<u>ACCACTCACG</u>	<u>GGAGCTCTCC</u>	<u>ATGCATTTGG</u>	
55	<u>TATTTTCGTC</u>	<u>TGGGGGGTAT</u>	<u>GCACGCGATA</u>	<u>GCATTGCGAG</u>	<u>ACGCTGGAGC</u>	
105	<u>CGGAGCACCC</u>	<u>TATGTCGCAG</u>	<u>TATCTGCTT</u>	<u>TGATTCCTGC</u>	<u>CTCATCCTAT</u>	
155	<u>TATTTATCGC</u>	<u>ACCTACGTT</u>	<u>AATATTACAG</u>	<u>GCGAACATAC</u>	<u>TTACTAAAGT</u>	
205	<u>GTGTTAATTA</u>	<u>ATTAATGCTT</u>	<u>GTAGGACATA</u>	<u>ATAATAACAA</u>	<u>TTGAATGTCT</u>	
255	<u>GCACAGCCAC</u>	<u>TTTCCACACA</u>	<u>GACATCATAA</u>	<u>CAAAAAATTT</u>	<u>CCACCAAACC</u>	
305	<u>CCCCCTCCCC</u>	<u>CGCTTCTGGC</u>	<u>CACAGCACTT</u>	<u>AAACACATCT</u>	<u>CTGCCAAACC</u>	
355	<u>CCAAAAACAA</u>	<u>AGAACCCTAA</u>	<u>CACCAGCCTA</u>	<u>ACCAGATTT</u>	<u>AAATTTTATC</u>	
405	<u>TTTTGGCGGT</u>	<u>ATGCACTTTT</u>	<u>AACAGTCACC</u>	<u>CCCCAACTAA</u>	<u>CACATTTATTT</u>	
455	<u>TCCCTCCCA</u>	<u>CTCCCACT</u>	<u>ACTAATCTCA</u>	<u>TCAATACAAC</u>	<u>CCCCGCCAT</u>	
505	<u>CCTACCCAGC</u>	<u>ACACACACAC</u>	<u>CGTGCTAAC</u>	<u>CCCATACCC</u>	<u>GAACCAACCA</u>	
555	<u>AACCCCAAG</u>	<u>ACACCCCA</u>	<u>CA</u>			

Note that after base position 16569, the sequence numbering starts again at base position 1.

- Close the rCRS sequence window to return to the Project window.

IMPORT SAMPLE DATA

Now that you have your reference sequence you are ready to import your sample data.

- From the **File** menu, choose **Import > Sequencer Project...**
- Navigate to the Sample Data folder inside the Sequencer application folder.
- mtDNA Profile projects are located in the Mitochondrial Sequences/mtDNA Profiles folder. Choose the project 90.SPF and click **Open**.
- Repeat to import 082790.SPF and click **Open**.


The two imported contigs contain data from the same sample, but it was sequenced in two different labs. The process of editing and assembling sample data into a contig is described in detail in the Reference Sequence Tutorial.

COMPARING THE DATA

Because both contigs are assembled to the same Reference Sequence, the rCRS, it is possible to compare equivalent base positions.

- Select both contigs and then select **Contig > Compare Consensus to Reference**.
- In the bottom left corner of the Variance Table, click on the resize  icon button to optimize the column widths for the display of the full names of the contigs.

Description: 2 consensus sequences compared to Reference rCRS.
 Comparison Range: Unfiltered
 Base Positions: 16024..16569 1..576
 Display Options: Large gap insertions (10 or more bases) included.

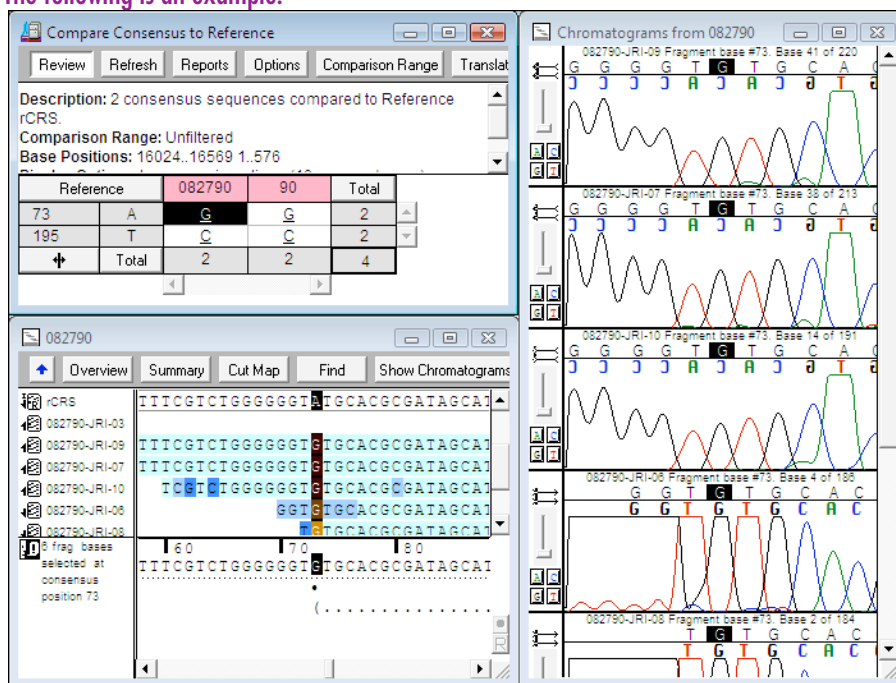
Reference	082790	90	Total	
73	A	<u>G</u>	<u>G</u>	2
195	T	<u>C</u>	<u>C</u>	2
	Total	2	2	4

Each row in the Variance Table represents a position in the Reference Sequence that has a difference in one of the sample sequences. Each column represents a sample sequence. The sample sequences differ from the Reference Sequence at positions 73 and 195. Every difference that is reported in one is confirmed in the other. However, the fact that each of the cells that contain the name of a contig is shaded in pink indicates that the data in the contigs does not cover the full length of the rCRS.

- Double-click on the **G** in column 082790.

This opens the Variance Table in Review mode so that you can now see all of the data that supports each difference reported in the table. You can also customize the layout of the windows, so that Sequencher will remember how you like to work.

- Rearrange the windows so that you optimize the view of the Chromatograms, the Contig Editor, and the Variance Table. The following is an example:



The screenshot displays three windows from the Sequencher software:

- Compare Consensus to Reference (Review mode):** Shows the Variance Table with columns for Reference, 082790, 90, and Total. The 'G' at position 73 in the 082790 column is highlighted.
- 082790 (Contig Editor):** Shows the reference sequence (rCRS) and the consensus sequence for contig 082790. The difference at position 73 (G vs A) is highlighted.
- Chromatograms from 082790:** Shows multiple chromatograms for different fragments of contig 082790, with the peak for 'G' at position 73 highlighted in each.


- From the **Window** menu, select **Remember Window Layout > Single Variance Table Review**.


Now, whenever you execute the Variance Table Review mode, Sequencher will display the windows in the configuration that works optimally for your projects.

When you select the cell to the right in the Variance Table, you display the equivalent base position in the next contig. When you move your cursor down in the Variance Table, you update the display to the next difference in the same contig.

- Move your cursor within the four cells in the Variance Table to navigate to all of the differences in the contig.
- Close the Variance Table and the Chromatogram and Contig Editor windows to return to the Project window.

Because this Variance Table, in contrast to the more simple Variance Tables created from the Sequence menu Compare To commands, is based on a comparison to a consensus sequence, the results of the table will vary depending on the method for calculating the consensus.

- Change the method of calculating the consensus from Plurality by selecting **Contig > Consensus Inclusively**.
- With the two contigs selected, again execute the **Compare Consensus to Reference** command from the **Contig** menu.
- Expand the name columns by clicking on the resize  icon button.

Reference		082790	90	Total
16,163	A	<u>N</u>		1
73	A	<u>G</u>	<u>G</u>	2
183	A	<u>N</u>		1
195	T	<u>C</u>	<u>C</u>	2
	Total	4	2	6

When the consensus is computed inclusively, the rare ambiguous base also makes a contribution to the consensus. Thus 082790 now has two ambiguities, one at position 16,163, and the other at 183.

- Double-click on the "N" cell at reference position 16,163.

Sequencher opens the Contig and Chromatogram Editor windows in the configuration saved from your earlier work. The chromatogram data shows some "G" noise, but there is "A" support in each of the five chromatograms.

- In the Variance Table, type "A" over the "N" at position 16,163.

Sequencher will make the appropriate edits in the Contig Editor, the Chromatogram Editor, and the Variance Table; however, although there are no longer any differences to the Reference Sequence at base position 16,163, this row will persist until you click on the Refresh button.

- Click on the "N" in the Variance Table cell at position 183.
- With your cursor still in the Variance Table, type an "A" to support the chromatogram data.

Now all of the data between the two analysts support each other, so you are now ready to create a report.

- Close the Variance Table and the Chromatogram and Contig Editor windows to return to the Project Window.

VALIDATE MTDNA PROFILES

In order to submit data to some databases, such as the CODIS database, you must validate your results. Sequencher has several reports and views to facilitate the validation of the mtDNA profiles.

- Select the two contigs, 90 and 082790, and then select **Contig > Validate mtDNA Profiles**.

mtDNA Profiles:	Agree			
HV1 and HV2 Ranges:	Agree			
Reference Sequence:	rCRS			
		Reference	Profile 1	Profile 2
		73	A	G
		195	T	C

The top of the Validate mtDNA Profiles view shows you that you have complete agreement between the work done by the two analysts.

The rest of the document gives you additional information concerning the two reports, including the range of bases for which you have data on both the HV1 and HV2 side of the origin, and alerts (in red text) that let you know if your data fails to meet the requirements for HV1 and HV2 coverage.

Profile 1	Incomplete Coverage for HV1,HV2
Name:	082790
HV1 Range:	16024 to 16235
HV2 Range:	33 to 255
Profile 2	Incomplete Coverage for HV1,HV2
Name:	90
HV1 Range:	16024 to 16235
HV2 Range:	33 to 255

Because the experimental data for both HV1, 16024 to 16235, and HV2, 33 to 255, do not cover the full range of these regions (between 16024 – 16365 or between 73 – 340), you are given the “**Incomplete Coverage for HV1, HV2**” warning.

Once you are satisfied with the validation of your data, you may create a report and export the profile data in CMF format. While you cannot generate reports, save, or export in the demo version of Sequencher, you may review an example of the report below.

CREATE AN ANALYST REPORT

A text version of the mtDNA profile validation may be saved or printed for your records.

- Click on the **Create Report** button at the top of the window.

Sequencher presents you with the following report to save or print:

```

mtDNA Profile Validation Report

mtDNA Profiles: Agree
HV1 and HV2 Ranges: Agree
Report Created: 9/10/2007 3:38 PM
Reference Sequence: rCRS

Profile 1: Incomplete Coverage for HV1,HV2
Name: 082790
HV1 Side Range: 16024 to 16235
HV2 Side Range: 33 to 255

Profile 2: Incomplete Coverage for HV1,HV2
Name: 90
HV1 Side Range: 16024 to 16235
HV2 Side Range: 33 to 255

Conflicts      Position  Ref    Profile 1  Profile 2
              73       A      G          G
              195      T      C          C

Analyst #1 Signature: _____
Analyst #2 Signature: _____
Reviewer Signature:  _____

```

- Close the **Analyst Report** window.

EXPORT MTDNA PROFILE

Once you have validated your results, you are ready to export the profile in the CMF format that is compatible with the CODIS database.

- Click on the **Export CMF** button at the top of the **Validate mtDNA Profiles** window.

You are presented with a dialog that allows you to enter the additional data required for the CMF export. The dialog is divided into two portions. The top half contains required information, the bottom half optional. If you float your cursor over any of the fields for entry, Sequencher will provide hints for entering the data to specification. Below is the tip that opens when you hover your cursor over the Specimen ID.

An identifier for the specimen within the import file. Specimen identifiers must be unique within each CMF import file based on the combination of CODISmp Lab ORI and CODISmp Specimen Identifier. Maximum 24 characters.

- Complete the data for each of the required fields.

- Click on the **Export** button.

Note: The Exporting functions are not accessible when you are running a copy of Sequencher in demo mode or using the special demonstration version of Sequencher.

- Select a location to Save to or Cancel. (Note that the file created on export is in xml format.)
- Close the **Validate mtDNA Profiles** window.

EVALUATE MULTIPLE SAMPLES

It is also possible to evaluate multiple samples in the same project.

- From the **File** menu, choose **Import > Sequencher Project...**
- Navigate to the Sample Data folder inside the Sequencher application folder.
- mtDNA Profile projects are located in the Mitochondrial Sequence/mtDNA Profiles folder. Choose the project 94.SPF and click **Open**.
- Repeat to import 120294.SPF and click **Open**.
- Navigate to **Select > Select All**.
- Select **Contig > Consensus By Plurality**.
- Select **Contig > Compare Consensus to Reference**.

Reference		120294	082790	90	94	Total
73	A	X	<u>G</u>	<u>G</u>	X	2
195	T		<u>C</u>	<u>C</u>		2
196.1	:	<u>I</u>			<u>I</u>	2
200	A	<u>G</u>			<u>G</u>	2
263	A	<u>G</u>	X	X	<u>G</u>	2
315.1	:	<u>C</u>	X	X	<u>C</u>	2
⬆	Total	4	2	2	4	12

The resulting Variance Table displays the results from both sets of data. The pink X's reflect positions where there is a difference in one sample that is not supported by data in the sample with the pink X. For instance, at position 73 there is an A to G difference in samples 90 and 082790, but we don't know if this difference also occurs in samples 120294 and 94, because there is no data covering this region.

- Click on the **Total** button in the lower left-hand corner of the **Variance Table**.

The samples are now sorted by their total number of differences. This has the effect of grouping the sequences with similar differences.

- Double-click on any of the cells to enter the Review mode.
- Navigate across the Variance Table to view the data at positions 195 and 200.

A Variance Table with more than one sample source gives you an opportunity to compare a variety of chromatogram data that yield different base calls. This can often be valuable information from which you can draw stronger conclusions.

The Variance Table allows for the comparison of hundreds of contigs. In contrast, the Validate mtDNA Profiles window specifically functions to compare two equal mtDNA profiles. If you select two different mtDNA profiles, the Validate mtDNA Profiles window will alert you that the two sets of data do not support each other.

- Close the Variance Table and the Chromatogram and Contig Editor windows to return to the Project window.
- Select contigs 90 and 94.
- From the **Contig** menu, select **Validate mtDNA Profiles**.

You will get the following display:

mtDNA Profiles:	Disagree
HV1 and HV2 Ranges:	Disagree
Reference Sequence:	rCRS

Reference		Profile 1	Profile 2
73	A	G	
195	T	C	
196.1	:		T
200	A		G
263	A		G
315.1	:		C

The yellow background in the reference column tells you that the data in one mtDNA profile is unsupported by the other. The yellow background in the mtDNA profile columns brings your attention to conflicts between the data.

- Close the project using **File > Close Project**.