

SEQUENCHER®

Tour Guide for Windows and Macintosh

©2016 Gene Codes Corporation

Gene Codes Corporation



Gene Codes Corporation
775 Technology Drive, Suite 100A, Ann Arbor, MI 48108 USA
phone 1.800.497.4939 or 1.734.769.7249 (fax) 1.734.769.7074 E-Mail: gcinfo@genecodes.com
www.genecodes.com

Tour Guide

Introduction

Sequencher has been developed to work with a wide range of sequencing applications. For example, **Sequencher** can be used to:

- Create assemblies for shotgun or EST sequencing projects
- Edit contigs while viewing all relevant trace data
- Assemble multiple sequences to a user-defined Reference Sequence
- Detect and annotate polymorphisms
- Align cDNAs to their genomic sequence using the Large Gap algorithm
- Discover heterozygous peaks
- Create difference reports for SNP discovery
- Display restriction maps, ORF maps, protein translations
- Automatically trim poor quality and vector sequences

Macintosh and Windows Support

Sequencher is available for both Macintosh and Windows platforms.

Sample Data

Some sample data have been included to get you started, but once you have tried **Sequencher** with the sample files, please try using your own data so you can see precisely what **Sequencher** can do for you.

What You Will Learn in This Tutorial

The purpose of this tutorial is to guide you through **Sequencher's** core assembly and editing functions. Additional application-specific tutorials are included in PDF format. In this tutorial, you will:

- | | |
|--|---|
| 1. Where is my copy of Sequencher | 8. Find heterozygotes |
| 2. Create a new project | 9. Work with a Reference Sequence |
| 3. Import data | 10. Translate sequences to amino acids |
| 4. Trim sequences | 11. Annotate a sequence |
| 5. Assemble a contig | 12. Create a Variance Table and Report |
| 6. View contig assembly | 13. Create a Translated Variance Table |
| 7. Edit assembled chromatograms | 14. What else can I do with Sequencher ? |

Once you have mastered these techniques, you will be ready to explore **Sequencher's** other powerful features.

Conventions Used in this Guide

Menu items or keys that you are to select are in **bold**. The purple text provides step-by-step instructions for running through the tour guide and the black text provides additional information. Greater than symbols define menu > submenu commands.

Before you start

Check that you have the appropriate hardware and disk space.

MACINTOSH RECOMMENDED

REQUIREMENTS:

10.7.5 and higher, Intel processor only
3 GB RAM (*working with NGS data requires more)
355 MB hard disk space (plus 320 MBs for NGS data)

WINDOWS RECOMMENDED

REQUIREMENTS:

Windows 8 and higher (64-bit only)
3 GB RAM (*working with NGS data requires more)
280 MB hard disk space (plus 580 MBs for NGS data)

*Additional requirements will vary. For BWA, the Cufflinks suite, and Velvet, at least 8 GB of RAM is recommended. For GSNAP, at least 16 GB of RAM is recommended. Very large datasets may require more RAM.

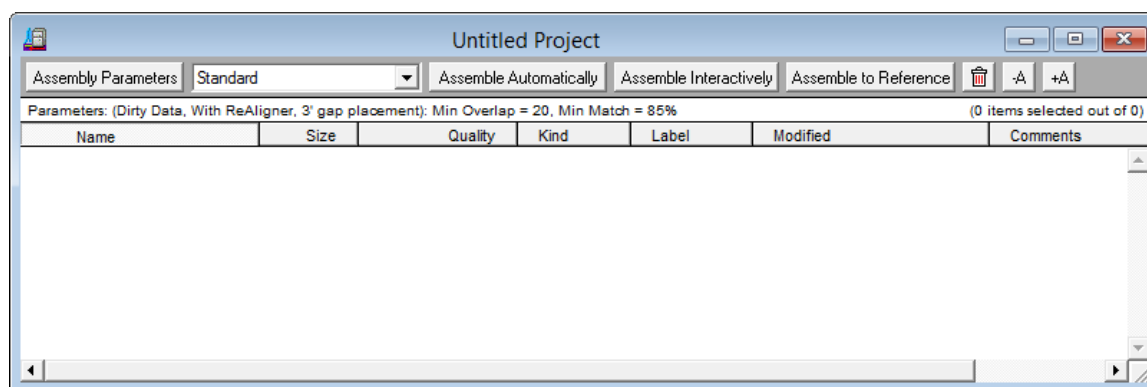
Where is my copy of Sequencer

If you haven't already done so, install **Sequencer** now. Follow the instructions in the Installation Guide. A version-specific **Sequencer** application folder is created in the Applications folder on Mac or in the Program Files > Gene Codes or Program Files (x86) > Gene Codes folder on Windows.

Create a New Project

- Launch **Sequencer** by double-clicking on the **Sequencer** icon. On your Macintosh, this will be in the opened **Sequencer** folder. On a PC, there is a **Sequencer** icon on your desktop.

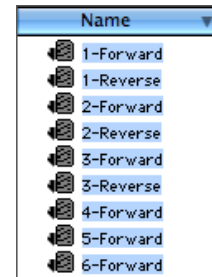
When **Sequencer** is launched, you will be presented with an empty **Project Window** as seen below. If you are using the **Sequencer** Viewer, you'll see ***** Viewer Mode ***** in the title bar text. This is where you import, manipulate, and display sequence fragments and assembled contigs.



New Project Window

Import Data

- To import data, select **File > Import > Folder of Sequences...**
- Browse to **Applications** (Mac) or **Program Files > Gene Codes** (Windows), then to the **Sequencher** application folder, then to the **Sample Data > Demo Sample Data > Demo Sequences** folder, selecting it.
- Click on the **Choose** (Mac) or **OK** (Windows) button.
- When prompted to import the 9 files, select the **Import All Files in Folder** command button. The project now contains the 9 sequences.



Imported Sequences

These imported files, with their associated quality scores, are just one example of the wide variety of file types **Sequencher** accepts for import. Note that the **Quality** column displays the % quality for each of the imported sequences—the percent of bases that are above the low quality threshold as set in the **Confidence User Preference** pane.

Trim Sequences

Sequencher has tools that allow you to trim imported sequences based on several different criteria: ambiguous data, data that have low confidence scores, or data contaminated with vector sequence. The trimmed data are fully recoverable within the **Sequencher** project. To trim the low confidence sequence:

- From the menu bar, choose **Select > Select All** to highlight all sequences if they aren't already highlighted.
- From the menu bar, choose **Sequence > Trim Ends...** **Sequencher** displays the **Ends Trimming** window for the default trimming parameters.

Sequencher recognizes poor quality sequence based on a number of criteria. The confidence score, provided in these samples, is one of the most sensitive. Increasing the stringency of the trim criteria further increases the quality of your data.

- Select the **Change Trim Criteria** button.
- Uncheck all but the three criteria checked below and adjust the values of the two confidence trims to match.

5 prime end

☐ Trim ABI primer blobs, where consecutive bases remain off the scale.

☐ Trimming no more than 25%, trim until the first bases contain less than ambiguities.

☒ Trimming no more than 25%, trim until the first bases contain less than bases with confidences below .

☐ Always trim at least bases from the 5' end.

3 prime end

☐ Trim chromatogram files before the first consecutive peaks below % of the highest peak.

☐ Starting bases after 5' trim, trim the first bases containing more than ambiguities.

☐ Trim from the 3' end until the last bases contain less than ambiguities.

☒ Trim from the 3' end until the last bases contain less than bases with confidences below .

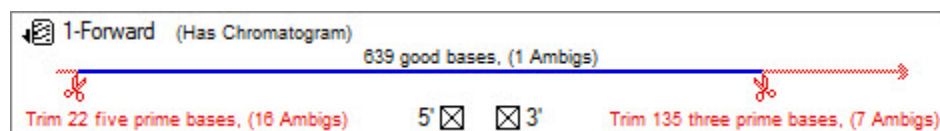
Post fix

☐ Maximum desired length after trimming is bases, trim more from the 3' end if necessary.

☒ Remove leading and trailing ambiguous bases.

- Click **OK** to return to the overview for Ends Trimming.

The **Ends Trimming** window displays how much poor quality data will be trimmed based on the defined criteria. You have the additional option to individually deselect a fragment for trimming on either the 5' or 3' end by removing the "X" from the appropriate box below the trim graphic.



Ends Trimming

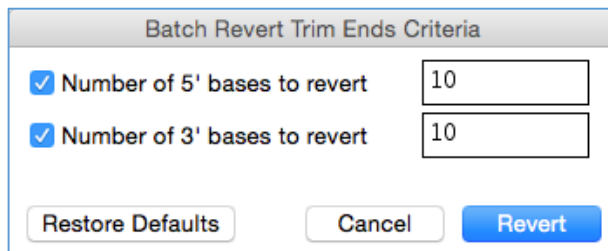
- Click on the **Trim Checked Items** button at the top of the window and then click on the **Trim** button when asked for confirmation.

The data that you have removed are completely recoverable. **Sequencher** always stores two copies of every imported Sequence, the original sequence and the data as you have edited it in **Sequencher**.

- Close the **Ends Trimming** window and return to the **Project Window** by clicking on the close control in the upper corner of the window.

BATCH REVERTING TRIMMED ENDS

After you have made the initial trim of your sequences in the **Project Window**, you may feel that the trim criteria were too stringent. You can revert all of the trim or a portion of the trim for the current sequence selection by going to the **Sequence** menu, selecting the **Batch Revert Trim Ends...** menu item, and entering the number of 5' bases to revert and/or the number of 3' bases to revert.



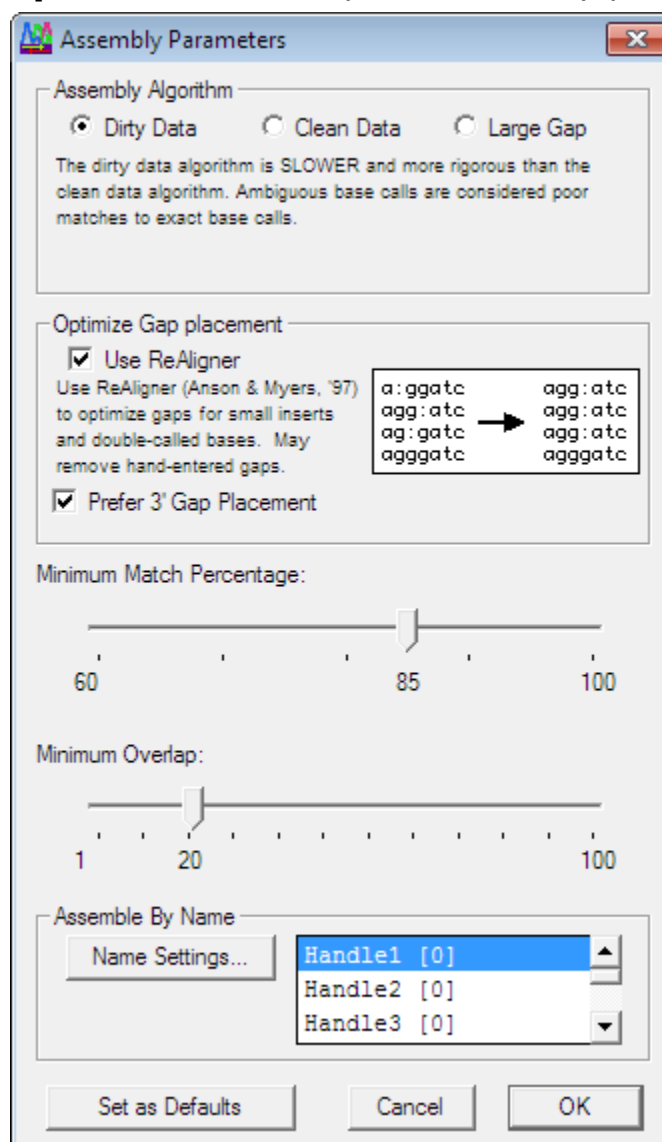
Assemble a Contig

Several alignment algorithms are provided with **Sequencher** to accommodate the wide variety of assembly applications. For this example, you will use the *Dirty Data* algorithm because it is best suited for data that may include the occasional ambiguities or miss-calls generated by automated sequencers. The “Assembly Strategies” and the “Assemble by Name” tutorials in the **Sequencher** Tutorials folder explain the other assembly options.

- Click on the **Assembly Parameters** button at the top of the **Project Window**.
- Accept the defaults for the Assembly Algorithm, Minimum Match Percentage, and Minimum Overlap parameters. They should be Dirty Data, 85% and 20 bases, respectively.
- Optimize gap placement by selecting **Use ReAligner** and **Prefer 3' Gap Placement** if not already selected.
- Click **OK**.

Once you have returned to the **Project Window**, you are ready to begin assembly.

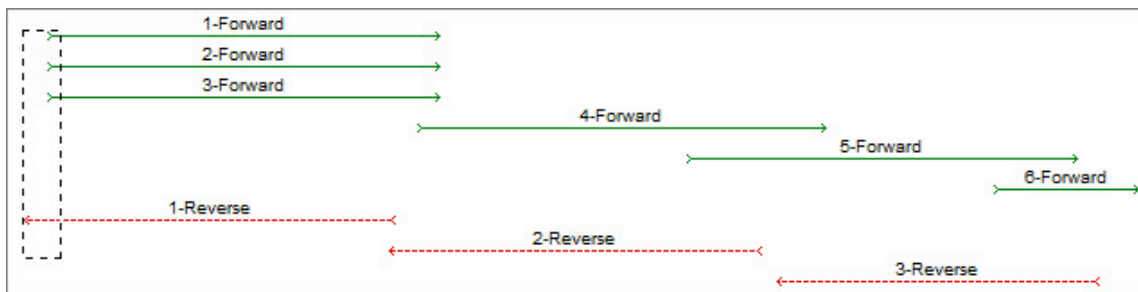
- All of the sequences should be selected. If they are not, select them now using **Select > Select All**.
- Click on the **Assemble Automatically** button at the top of the **Project Window**.
- Click **Close** to dismiss the **Assembly Completed** dialog.



View Contig Assembly

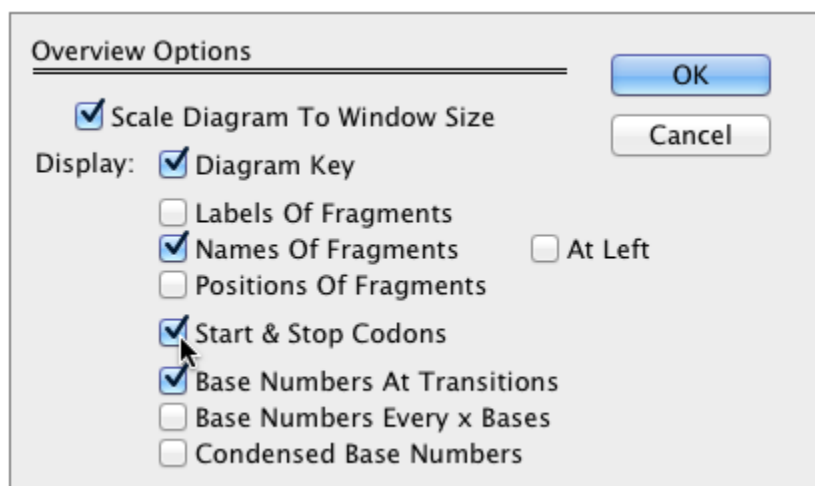
The **Contig User Preference** defines sorting criteria for new assemblies. The default preference sorts the fragments according to position, 5' to 3', within the contig. **Sequencher** provides a number of alternative sorting options.

- Double-click on the contig icon to open the contig **Overview** window.
- Click on the **Sort** button, select the **by Strand** radio button, and click on **OK** to sort the fragments by strand.



Contig Overview

- Click on the **Sort** button, select the **by Position** radio button, and click **OK** to return to the original sorting order.
- Click on the **Options** button to open the **Overview Options** dialog.
- Select **Start & Stop Codons** to turn on the Codon Map display.

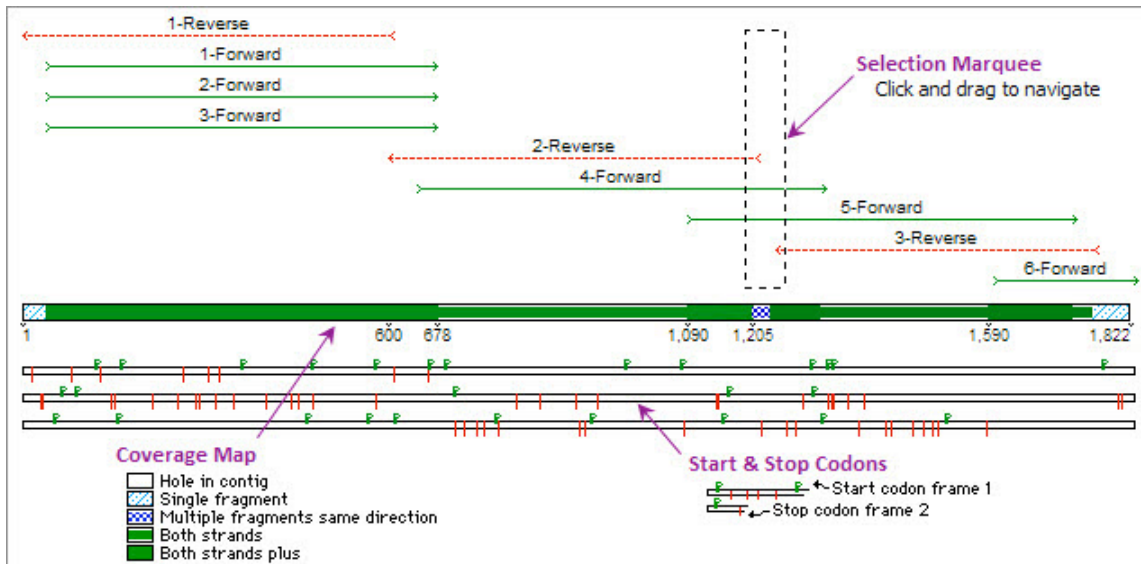


- Click on the **OK** button to dismiss the **Overview Options** dialog.

The **Overview** contains three sections. The top section displays a schematic of how the fragments are assembled in this contig. The arrows indicate the direction of the fragment in relation to the assembly. **Sequencher** provides a Selection Marquee that allows you to navigate from within the **Overview** into the **Bases** window.

The next section provides coverage information. For instance, the consensus called around base 1205 has less coverage than the surrounding consensus bases.

Below the coverage bar is the open reading frame map. Three bars marked with green flags and red lines, representing start and stop codons respectively.

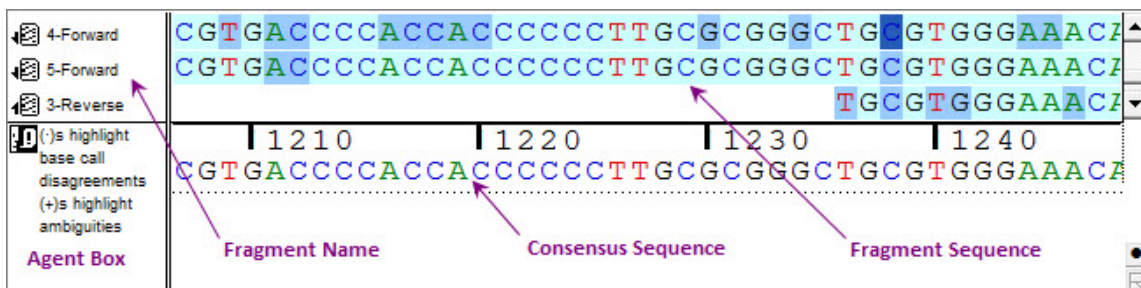


- Click and drag on the **Selection Marquee** in the **Overview** so that it selects the region around base 1,205.
- Select the **Bases** button at the top of the window to open the **Contig Editor** and view the base sequences that assemble at this position.

Edit Assembled Chromatograms

The **Contig Editor** provides the tools for checking and editing sequences. It is divided into four quadrants. You can modify the appearance of the **Contig Editor** from the **View** menu and in your User Preferences.

- Select **View > Display Color Bases**.
- Under **Window > User Preferences**, change **Display > Contig** so the Font is Courier New and the Size is 18.
- Close the **User Preferences** window.



The two upper panels show the individual fragment names to the left with their sequences to the right. The Agent Box contains descriptive information about your sequences and your selection. The lower right panel displays consensus information including ambiguities <+> and disagreements <•> if there are any. The blue shading describes the confidence, low, medium, and high, with the lighter shades defining higher confidence. Note that you may toggle off this shading by selecting **View > Display Base Confidences**.

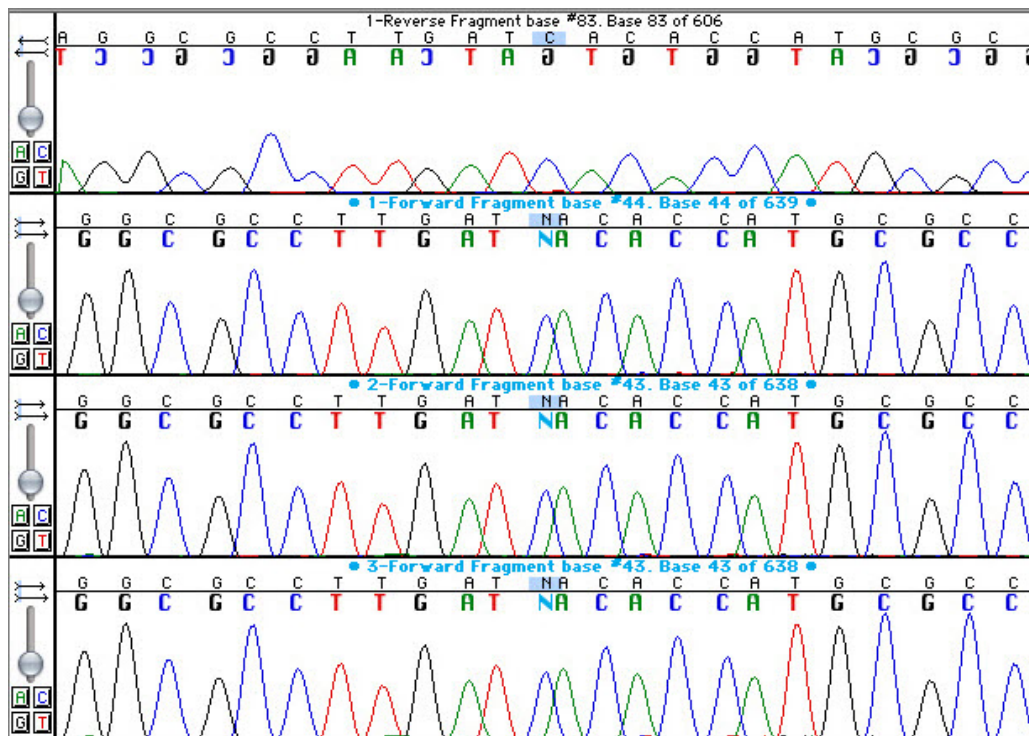
- To begin the editing process, move your selection to base one in the consensus.

The **Select** menu provides several tools to navigate to areas of interest in the contig.

- From the **Select** menu, choose the command **Next Ambiguous Base**. The Agent box reads “4 fragment bases at consensus position 83,” and “Select Next Ambiguous Base = spacebar”.

You have now moved your position in the consensus to the position of the first ambiguous base. An ambiguous position in the consensus is any that includes: 1) A contributing fragment base that is not an A, C, G, or T, 2) Disagreements between other fragment bases, or 3) All fragment bases that contribute to the consensus have low quality scores. Position 83 is flagged, because the base called in three of the fragments is an “N”.

- To view the chromatograms at this position, while your selection is still on the ambiguous consensus position, select the **Show Chromatograms** button at the top of the window.



Assembled Trace Window

Each chromatogram window displays the current version of the base calls in black and above a line that separates them from the original base calls. At the left of the trace window, **Sequencher** provides tools that allow you to

manipulate how the traces are displayed. The volume control bars allow you to adjust the chromatogram peak height. The **A**, **C**, **G**, and **T** buttons allow you to turn off the display of the signal from any or all of the bases. Additional controls to format the traces are available through the **Window** menu under **User Preferences... Display** pane.

- The base call at position 83 is ambiguous in the forward direction, probably because the irregular spacing challenged the original base caller. To change the "N" to "C", with your selection still on the consensus base, type "**C**". Your edit will correct the consensus and every sequence in the contig at that position.

Note that **Sequencer** displays the edited base call in a contrasting color. You are now ready to continue editing your sequence. You can move in your contig from one region of ambiguity to the next using **Sequencer's** navigational tools.

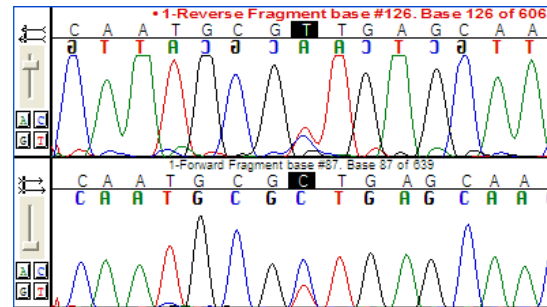
- From the **Select** menu, choose **Next Ambiguous Base** or use the **spacebar** to execute the previous **Select** command.

Sequencer jumps to the next ambiguity in the contig.

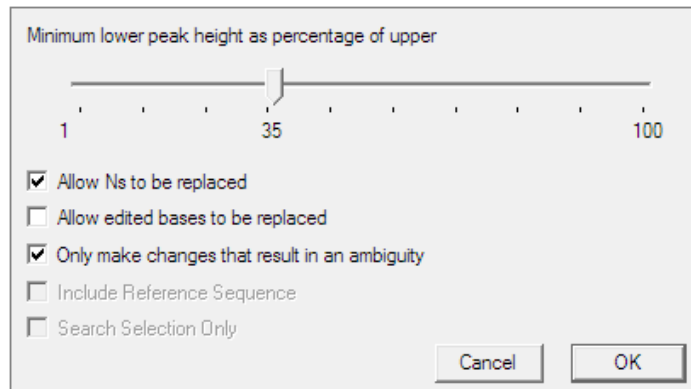
Call Heterozygous bases

The next ambiguous base in the consensus is at position 126. One reverse sequence calls a T and three forward sequences call a C. A quick look at the chromatogram, however, shows that the forward and the reverse sequences have both C and T peaks at position 126. This sequence is from a mixed population.

- Raise the slider for the 1-Reverse sequence to amplify the signal.
- Type "**Y**" in the consensus line at position 126.



In **Sequencer**, you can manually call each heterozygote one at a time, or you can use the **Call Secondary Peaks...** function to automatically find the heterozygous bases.



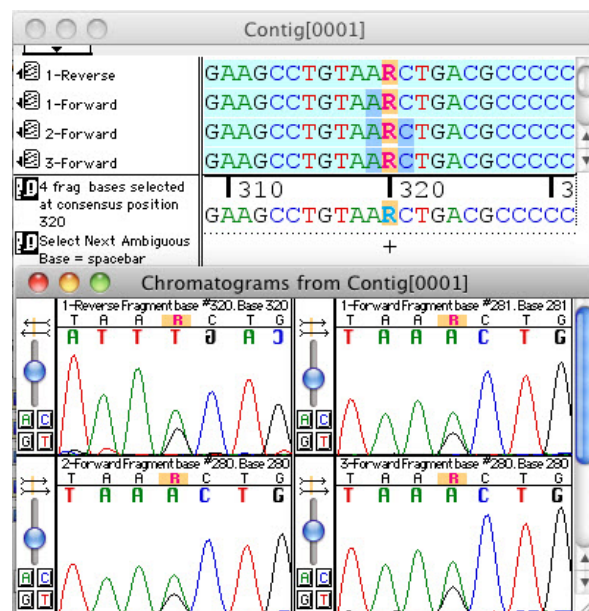
- From the **Sequence** menu, select **Call Secondary Peaks...**
- Change the Minimum lower peak height to **35%**.
- Click on the **Only make changes that result in an ambiguity** checkbox to turn it on.
- Click **OK**.

Sequencher displays a dialog so that you can confirm the changes to 40 base calls.

- Click **Continue**.
- With your cursor still in the consensus, click the **spacebar** to select the next ambiguous base position.

The next ambiguous base position is 320. At this position, you will find that all of the contributing sequences have bold magenta "R"s. The color indicates that the sequences have been edited, in this case automatically by the Call Secondary Peaks... function. The "R" is the IUPAC abbreviation for a mix of puRines, or "A or G". The "+" under the consensus line flags the ambiguous position.

- Continue selecting and editing until all the ambiguities are resolved.



Note that not all of the automatically called bases will be true heterozygotes nor will this function capture all heterozygotes, but it will find all bases that have secondary peaks that are at least 35% of the primary peak.

- Close the **Contig** and **Chromatogram** windows by clicking on the close buttons.

Work with a Reference Sequence

There are numerous potential applications for the Reference Sequence. The Reference Sequence facilitates comparative sequence alignments, defines base numbering, and boosts assembly speed. In this Tour Guide, we will use the Reference Sequence function to compare the assembled trace sequences to a known text sequence.

- From the **Project Window**, select **File > Import > Sequencher Project...** and navigate to the **Demo Sample Data** folder from which you imported the trace files.

- Click on the file **HepC Reference.SPF** and then click on the **Open** button.

The new sequence, AB049090, is now in your project. You can tell that this sequence is already a Reference because it has an "R" in the icon and "Ref:" precedes "DNA Fragment" in the *Kind* column. You can make any sequence a Reference Sequence in **Sequencer** from the **Sequence** menu. You can also right click on the sequence name to invoke the context sensitive menu. The "Reference Sequence" tutorial provides more information on the Reference functions.

You can use the same assembly parameters that we used for the autoseq fragments. Note that the current parameters are listed in the **Project Window** just below the button bar.

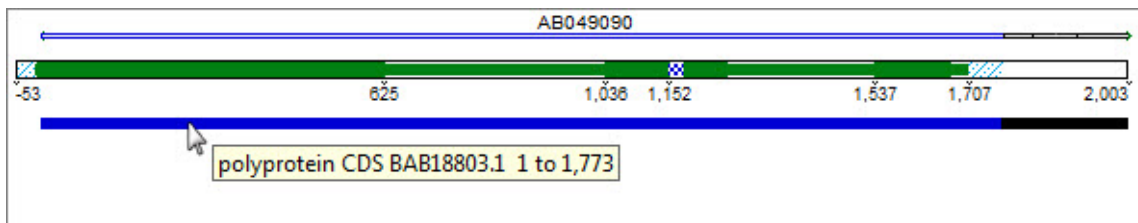
Parameters: (Dirty Data, With ReAligner, 3' gap placement): Min Overlap = 20, Min Match = 85%

- Choose **Select > Select All** and click on the **Assemble to Reference** button.
- Click **Close** to dismiss the **Assembly Completed** dialog.

The Reference Sequence is now incorporated into the new contig.

- Double-click on the contig icon to display the **Contig Overview**.

Immediately you can see that the Reference contributes to the contig in a different manner than the non-Reference sequences. For example, the numbering of the sequence, as it appears on the coverage bar, is negative until the Reference begins to contribute. Note the white space at the 3' end of the coverage bar indicates that the Reference does not contribute coverage at all. The blue bar in the Reference is a graphic representation of the CDS feature. Move the cursor over a feature graphic in the feature map below the coverage bar to see the feature's name and location.



When the data that extends beyond the Reference is not of interest, you can trim the contig sequences to the Reference.

- From the **Contig** menu, select the **Trim to Reference Sequence** command.

The contig now starts at base position 1. It is useful to view a Reference Sequence when editing a sample sequence. The Reference Sequence will guide you so you can critically examine the differences between the sample consensus and the Reference. Yet the Reference will not contribute to the consensus nor will the Reference be affected by edits in the consensus.

Translate Sequences to Amino Acids

Sequencer provides a variety of ways to display the translation of DNA sequence. For instance, you can display one or all three of the reading frames below the sequence while editing.

- Drag the **Selection Marquee** in the **Overview** to the 5' end of the contig.
- Click on the **Bases** button.
- Place your cursor on base position 1 in the consensus line and click on it.
- Click on the [•] button in the lower right-hand corner of the **Contig Editor. Sequencer** displays the translated consensus.

The first click changes the [•] to a [1], displaying the translation in the first reading frame. This button will toggle through each of the reading frames, followed by the display of all three, and then concluding with the display of an [R] which displays both the Reference Sequence translation and the consensus translation in the Reference Frame.

- Click on the upper translation button until it reaches a [1] to leave the translation in the first reading frame.
- To toggle the display of the Reference Sequence translation, click on the lower translation button (the one with the **R** icon). With this button, you may cycle through all 3 reading frames for the Reference Sequence.

The screenshot displays the Sequencer interface. On the left, a menu includes options: 1-Reverse, 1-Forward, 2-Forward, 3-Forward, AB049090, and Select Next Ambiguous Base = spacebar. The main area shows a DNA sequence: ATGTCCTACACATGGACAGGCGCCTTGATCACACCATGCG. Below the sequence, four reading frames are listed: 1-Forward (ATGTCCTACACATGGACAGGCGCCTTGATCACACCATGCG), 2-Forward (ATGTCCTACACATGGACAGGCGCCTTGATCACACCATGCG), 3-Forward (ATGTCCTACACATGGACAGGCGCCTTGATCACACCATGCG), and AB049090 (ATGTCCTACACATGGACAGGCGCCTTGATCACACCATGCG). The protein translation for the first frame is shown as M S Y T W T G A L I T P C. A second row shows the Reference Sequence translation: M S Y T W T G A L I T P C R. A label 'Consensus and Reference Translation Buttons' with an arrow points to the 'R' button in the bottom right corner.

Contig With Protein Translation of Consensus and Reference Sequence

Annotate a Sequence

Sequencer provides for sequence annotation by allowing you to create a feature for a single base or a range of bases.

- If your cursor is not in the consensus, select a base in the consensus.
- From the menu bar, choose **Select > Bases by Number...**
- In the **Select Base** dialog, enter **208** in both input fields.
- Click **OK**.

At this position, the sample sequence has a Valine, GTG, and the Reference codes for a Methionine, ATG.

- Choose **Sequence > Mark Selection As Feature**.
- From the **Feature Key** drop-down menu, choose **variation**.
- In the **Feature Name** input field, give your feature a brief name like "**A-> G, Met-> Val**".

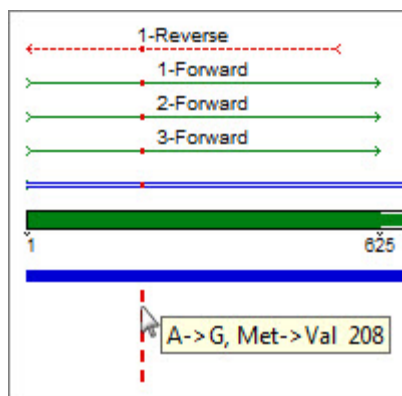
The default Feature Style for a variation is red and underlined, but you may choose any style you wish.

- Click **OK**.
- If the **View > Display Features** menu item is not checked, select it.
- Turn off **Display Color Bases** from the **View** menu.
- From the **View** menu, choose **Colors As Backgrounds**.

The feature will now be displayed in both the **Bases** view and the **Overview**. Note that in the **Overview**, the red single base feature displays over the range of bases covered in the blue CDS feature.

TTAAGGAG	GTGAAGGCG
TTAAGGAG	GTGAAGGCG
TTAAGGAG	GTGAAGGCG
TTAAGGAG	GTGAAGGCG
TCAAGGAG	ATGAAGGCG
200	210
TTAAGGAG	GTGAAGGCG
.	.
..A-G,	Met->Val..
K	E
K	E
M	K
A	A

Feature in the Bases View



Feature in the Overview

Create a Variance Table and Report

After you have edited the sample sequence so that you are confident that the consensus base calls are correct, you can generate a comparison report of how the sample differs from the Reference.

- Close the **Contig Editor** window.
- With the contig selected, choose **Contig > Compare Consensus to Reference**.

Reference		Contig...	Total
18	G	A	1
25	C	T	1
42	T	C	1
51	A	G	1
69	C	T	1
73	T	Y	1
90	C	G	1
93	C	T	1
94	C	A	1
99	C	T	1
153	C	T	1
183	C	T	1
201	C	T	1
208	A	<u>G</u>	1
237	C	T	1
267	G	R	1
⤴	Total	78	78

Sequencher displays a table listing the differences between the consensus sequence that you edited and the Reference Sequence that you imported. Note that base 208 is still red and underlined.

You can also modify the look of this table.

- Click on the symbol that looks like an open elevator button in the bottom left corner to expand the width of the columns.
- Turn off **Colors As Backgrounds** from the **View** menu.

This report lists the differences between the consensus of one contig and the Reference, but it is also possible to create this report for hundreds of contigs, when they share the same Reference. Or choose the **Compare Bases** command from the **Sequence** menu to review differences in sequences within a single contig.

The **Variance Table** also acts as a link to the original data.

- Double-click on the cell at base position 208.

Sequencher rearranges the windows and opens the **Contig** and **Chromatogram Editors** for that base position.

- Use the arrow keys on your keyboard to navigate to bases of interest in the **Variance Table**.
- Create a new variation feature from the **Variance Table** at position 338 by selecting that cell in the **Variance Table** and then executing the **Sequence > Mark Selection As Feature** command.
- Click **OK**.

Sequencher defaults to the variation feature key because this was the last feature key used. The new feature name is "variation" and it is also red and underlined.

- Click on the **Reports** button to open the **Reports** dialog.

You have several reporting options available to you so that you can share the contents of the **Variance Table** outside of **Sequencher**. Reporting is disabled in Viewer Mode but you can see sample reports in **Sequencher Help**. The image below is the result of creating a **Variance Table** from the Row Selection for positions 153 – 338.



Variance Table Report

For research purposes only. Not intended for clinical use.

Date: December 22, 2010
Project Name: donw.SPF
Contig Name: Contig[0001]
Compare Consensus to Reference: AB049090
Comparison Range: Unfiltered
Base Positions: Selected Rows
Options: Large gap insertions (10 or more bases) included.
Matches to ambiguous reference positions excluded.

Reference	Contig[0001]	Total
153	C	T
183	C	T
201	C	T
208	A	G
237	C	T
267	G	R
294	C	T
338	G	A
Set Total	8	8
Total	78	78

Sequence Name	Comparison Range Coverage	Differences
Contig[0001]	Incomplete 1 to 1,769	78

Gene Codes Corporation



Report generated by Sequencher®; The complete software solution for sequencing DNA.

Create a Translated Variance Table

In addition to the **Variance Table** that displays the results of the comparison of DNA sequences, **Sequencher** provides a **Translated Variance Table** that displays the differences in the translation of DNA sequences. The following directs you to create the translated "sister" table to the currently open **Variance Table**, but you can also create a **Translated Variance Table** directly from a contig by selecting sequence names in the contig or by selecting contigs in a **Project Window**.

- While you are still in the **Review** mode of this **Variance Table**, click on the **Translation** button on the button bar of the **Variance Table** window.

Description: 1 consensus sequence compared to Reference AB049090.
Comparison Range: Unfiltered
Base Positions: 1..2003
Amino Acid Positions: 1..667

Reference		Contig...	Total
94 32	CAC H	AAC N	1
208 70	ATG M	GTG V	1
337 113	AGG R	AAG K	1
388 130	CAG Q	GAG E	1
409 137	GTC V	ATC I	1
562 188	GCC A	TCC S	1
631 211	AGA R	ARA ?	1
634 212	TGC C	AAC N	1
+	Total	15	15

Sequencher opens the **Translated Variance Table**.

It is configured in the same way as the **Variance Table**. Note that the positions of the other three windows also adjust to accommodate the new table.

- The row of numbers in the left most reference column refer to both the first base of the codon, above, and amino acid, below, of the Reference Sequence.
- Adjacent to the numbers are the corresponding Reference codon and its translation.
- The sequences for comparison are in the columns to the right of the Reference. In this case, there is only one.
- The pink header is to flag any comparison in which the entire length of the Reference is not covered by the sample sequence.

What else can I do in SEQUENCHER?

You have now tried the basic capabilities of **Sequencher** for assembly and alignment of DNA sequences. Continue to explore the power of **Sequencher** using your own data. We invite you to explore the additional tutorials available as PDFs in the Tutorials folder or from our website <http://genecodes.com/training/tutorials>.

Thousands of laboratories around the world have made **Sequencher** the desktop standard for DNA assembly and alignment. Once you have worked with **Sequencher**, you will understand why.

Contacting Gene Codes

We welcome your questions and input on **Sequencher**. For questions regarding the use of **Sequencher** or to purchase your own copies, please use one of the following methods to contact us:

734.769.7249 (phone) • 734.769.7074 (fax) • 1.800.497.4939 Toll free (USA) • gcinfo@genecodes.com
(Email) www.genecodes.com