SEQUENCHER®

Tutorial for Windows and Macintosh

Assembly Strategies

Gene Codes Corporation

Assembly Strategies

Assembly Strategies Tutorial

Sequencher offers a wide variety of assembly options. The most frequently used options, **Assemble Automatically**, **Assemble Interactively**, and **Assemble to Reference** are available as prominent buttons in the **Project window**. However, there are numerous other options under the **Assemble** menu that may be more suitable for a given project.

In addition to the assembly options described here, investigate the tutorials for Reference Sequence and for **Assemble by Name**. The Reference Sequence has a number of special properties that affect the assembly process. **Assemble by Name** allows you to further automate the assembly process, providing greater quality control while saving you time.
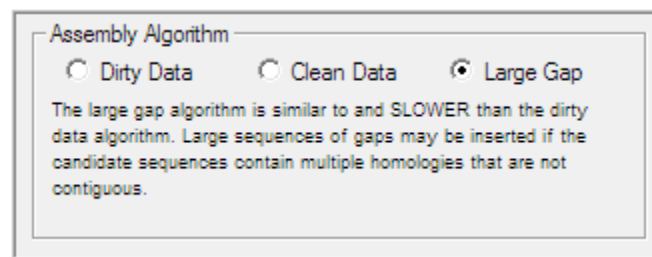
THE LARGE GAP ASSEMBLY ALGORITHM

The **Large Gap** algorithm allows you to assemble sequences that are expected to contain insertions and deletions greater than 9 bases long. Typical examples of assemblies that require the **Large Gap** algorithm include the assembly of a cDNA sequence to a genomic sequence and the assembly of related genes with alternative splicing.

- Launch **Sequencher**.
- Go to the **File** menu, select **Import > Sequencher Project...**, and double-click on **Large Gap.SPF**.

This file contains two sequences, p53 cDNA and p53 genomic sequences.

- Click on the **Assembly Parameters** button.
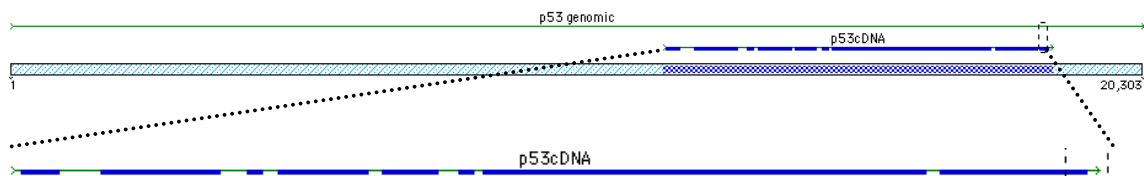- Click on the **Large Gap** radio button, then click **OK**.

- Select both the **cDNA** and **genomic sequences** if they are not already selected.
- Click on the **Assemble Automatically** button.

Sequencher will assemble the cDNA to the genomic sequence in less than one second.
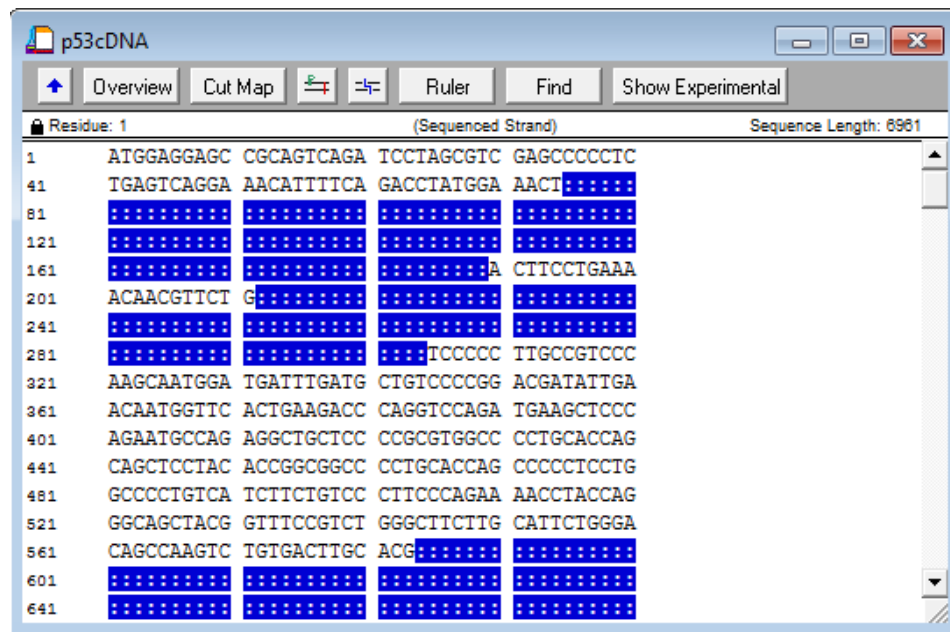
- Close out of the **Assembly Completed** dialog.

- Double-click on the **Contig[0001] icon** to open the **Contig Editor Overview**.
- Make sure that the **Motifs** and **Colors As Backgrounds** menu items on the **View** menu are checked.
- To see a graphic representation of the gap alignment, from the **Window** menu, click on **Motif Definitions…**.
- Type a "**:**" as a Motif and select "**Blue**" from the **Display** drop-down menu.
- Close the **Motifs** dialog.

Now the location of each gap will be highlighted in both the Overview and in the text displays of the sequences.

- Double-click on the name of the **P53cDNA** sequence to open its **Sequence Editor**.
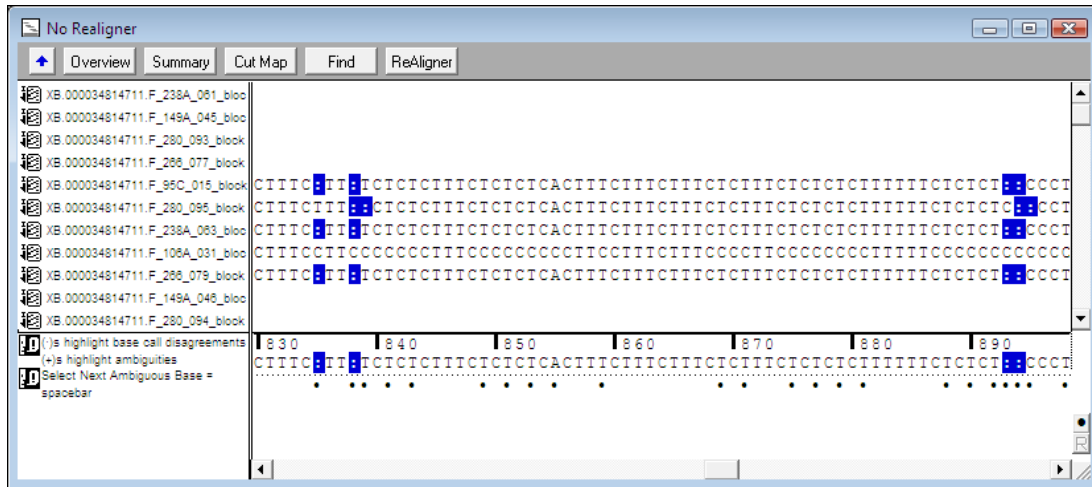

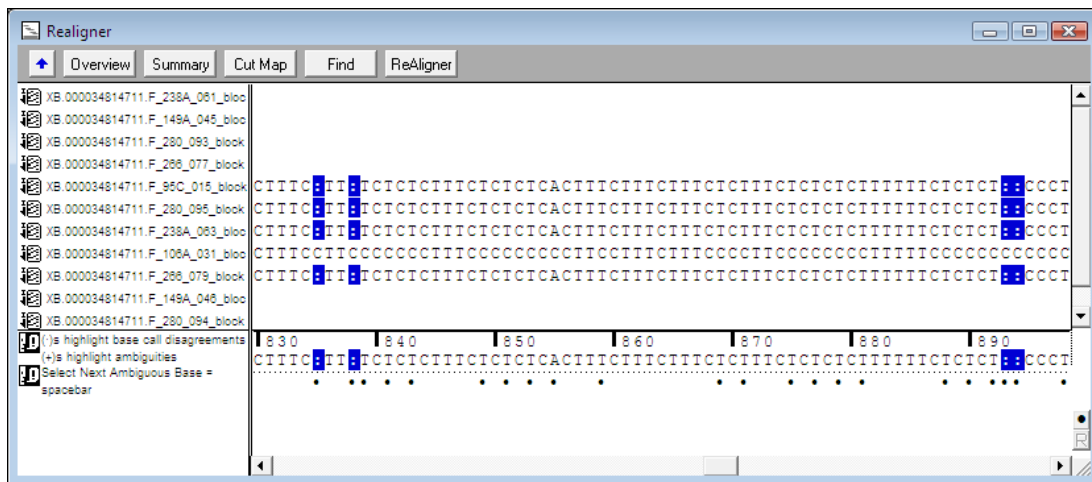Gaps have been placed in the cDNA in order to assemble the sequence to the genomic.

- Close the **Sequence Editor** by clicking on the close box.
- From the **File** menu, select **Close Project** without saving.

ASSEMBLY WITH REALIGNER

Realigner is an optional step that can augment the standard assembly algorithm. The **ReAligner** function evaluates the placement of gaps within a contig and optimizes their placement. **ReAligner** facilitates editing in the consensus sequence and more clearly displays the effect of insertions and deletions.

The sequences in the assemblies above and below are the same. The top contig was assembled without **ReAligner**. The bottom contig was assembled with the **ReAligner** option. Note that the location of the inserted bases is much clearer in the assembly below than it is in the one above.

To test the **ReAligner** option, use the following steps to open the **Sequencher** Realigner project:

- Select **File > New Project**.
- Go to the **File** menu, select **Import > Sequencher Project**..., and then double-click on **Realigner.SPF**.

The project contains the two contigs, Realigner and No Realigner.

- Select the contig **Realigner** and dissolve the assembly by choosing **Dissolve Contig** from the **Contig** menu.
- Dismiss the warning dialog by clicking on the **Dissolve Contig** button.

The contig will revert to its original 11 sequences.

- To reassemble the contig with the **ReAligner** option, click on the **Assembly Parameters** button.
- The checkboxes to **Use ReAligner** and **Prefer 3' Gap Placement** should be on by default. If not, select them.
- Click **OK**.
- Ensure that all the sequences from the dissolved contig are still selected.
- Click on the **Assemble Automatically** button.

You will assemble all of the 11 sequences and run the ReAligner in less than one second.

Note: **Sequencher** assembles and dissolves contigs so quickly that it is easy to optimize assembly parameters and options.

- Select the **Close** button to close the **Assembly Completed** dialog.
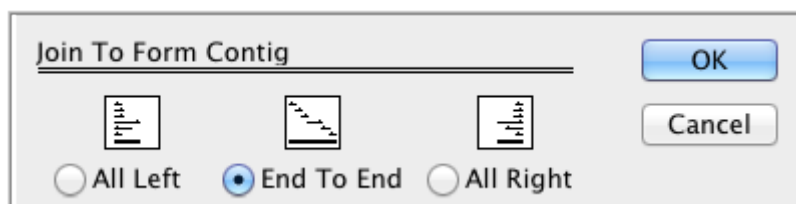- From the **File** menu, select **Close Project** without saving.

MINDLESSLY JOIN

Mindlessly Join is the kind of tool that every assembly program ought to have, but very few do. It is the kind of tool that allows you to really tinker with sequences outside of the conventional sets of algorithms.

- From the **File** menu, select **New Project**.
- Go to the **File** menu, select **Import > Sequencher Project…**, and then double-click on **Mindlessly Join.SPF**.

The **Project window** now contains the sequences for each of the eleven exons of p53. You can use the **Mindlessly Join** command to perform a virtual ligation and create a cDNA sequence from these exons by following the steps below.
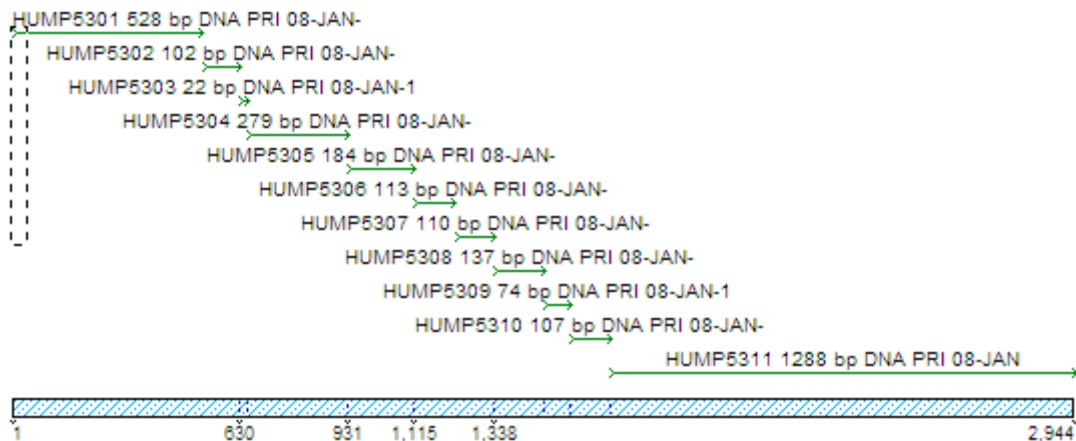
- Use your mouse to drag a selection box around the exons in the order in which you wish them to assemble.
- Select **Assemble > Mindlessly Join…**

- Select **End to End**, then click **OK**.

Sequencher will build the new contig instantly.

- Double click on the new **Contig[0001] icon** to open the **Overvie**w.

Note that there is no overlap between the sequences. The other options, All Left and All Right, are valuable for aligning sequences that do not assemble based on standard assembly criteria.

- From the **File** menu, select **Close Project** and don't save changes when prompted.
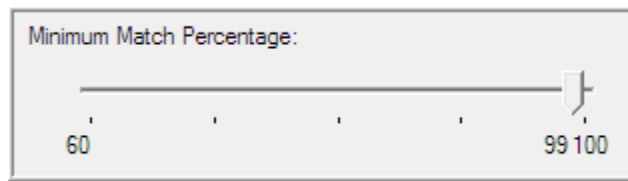
ADD SELECTED ITEMS TO OTHERS

Add Selected Items To Others is the favorite algorithm of those working in large sequencing projects. Assemblies of tens of thousands of sequences are greatly expedited by using this command in lieu of the standard, **Assemble Automatically** command. **Add Selected Items To Others** not only saves you time, but it also gives you more control over the direction of your assemblies. Uses for this function include genomic sequencing and clustering ESTs.

- From the **File** menu, select **New Project**.
- Go to the **File** menu, select **Import > Sequencher Project…**, and select **Add Selected Items.SPF**.
- Click **Open**.

The project contains 83 sequences, all related to each other, and all about 1100 bases in length. Our objective is to identify which, if any, of the sequences in the project are most closely related to the first sequence in the project – HLA:HLA00001.

- To do this, modify the **Assembly Parameters** to require a **Minimum Match Percentage** of 99%. Click on the **Assembly Parameters** button and slide the **Minimum Match Percentage** control to 99%.

Minimum Match Percentage:

60   99 100

- Click **OK** to dismiss the dialog.
- Select only the sequence **HLA:HLA00001 icon** in the **Project window**.
- Under the **Assemble** menu, select **Add Selected Items To Others**.

Sequencher will only assemble the selected sequence with the sequences that match the original selected sequence with a 99% identity.

This assembly option is also useful when you have a growing data set. For instance, if you generate 96 new EST sequences per day, you may wish to assemble the newly generated sequences to the previously assembled sequences. Newly imported data is automatically selected in **Sequencher**. By using the **Add Selected Items to Others** assembly command, you direct **Sequencher** to perform all of the possible comparisons with the new data without repeating the comparisons that you have already performed with the old data.

- **Close** out of the **Assembly Completed** dialog.
- From the **File** menu, select **Close Project** without saving.

ASSEMBLE INTERACTIVELY

The **Assemble Interactively** option can be used whenever you want the ultimate control in the assembly of a batch of sequences. The **Assemble Interactively** window provides you with detailed information regarding overlap, mismatch, and gapping for any given sequence and a set of candidate sequences. The candidate assemblies are driven by your user-defined **Assembly Parameter** settings.

- Go to the File menu and select **New Project**.
- Then, from the **File** menu again, select **Import > Sequencher Project…** and then select **Assemble Interactively.SPF**.
- Click **Open**.

You will open a project that contains the same 83 sequences that we used in the Add Selected Items project.

- Click on the **Assembly Parameters** button.
- Set the following parameters:
  - ➢ Dirty Data
  - ➢ Use ReAligner and Prefer 3'Gap Placement: Off
  - ➢ Minimum Match Percentage: 90%
  - ➢ Minimum Overlap: 20 bases

- Click **OK**.
- Ensure that the **Assembly Mode** setting is set to **Standard**, like in the following picture:

- With all of the sequences in the project still selected, click on the **Assemble Interactively** button.

- Select the first sequence in the **Candidates** list.

The **Matches** window displays a list of all of the selected sequences that would assemble to the candidate sequence given the current assembly parameters. The list also displays the %Match, the length of the overlap with the candidate sequence, the number of mismatches, the number of gaps, and the length of the new contig. The list is sorted by %Match.

Matches	%Match	Overlap	Mismatch	Gaps	New Contig Len
HLA:HLA01496	99	1,098	1	0	1,098
HLA:HLA00004	≠99	≠1,099	≠0	≠1	≠1,099
HLA:HLA00110	≠99	≠1,098	≠4	≠0	≠1,098
HLA:HLA00002	≠99	≠1,098	≠5	≠0	≠1,098
HLA:HLA00043	≠98	≠1,098	≠12	≠0	≠1,098
HLA:HLA01297	≠98	≠1,098	≠13	≠0	≠1,098
HLA:HLA00047	≠98	≠1,098	≠14	≠0	≠1,098
HLA:HLA00046	≠98	≠1,098	≠14	≠0	≠1,098
HLA:HLA00040	≠98	≠1,098	≠18	≠0	≠1,098
HLA:HLA00037	≠98	≠1,098	≠19	≠0	≠1,098

- Select the sequence with only one mismatch, **HLA:HLA01496**, in the **Matches** list.

Sequencher automatically displays the alignment of the selected sequence with the candidate sequence, and provides the following description of the putative alignment.

The view at the bottom of the window displays and scrolls through the new assembly and the new consensus.

- If you choose to create this assembly, click on the **Assemble** button.
- Name the new contig and click **OK**.

The new contig will appear in the **Candidates** window and in your **Project window**.

- Close the **Assemble Interactively** window by clicking on the **Done** button.
- Close the project without saving and exit **Sequencher** if desired.