SEQUENCHER™

Tutorial for Windows and Macintosh

Assemble by Name

Gene Codes Corporation

TCA **G** ENE
AGT **C** ODES

Assemble by Name

The **Assemble by Name** feature allows you to take information that you already have in your sample name and:

> ➢ Sort the sequences in your Project window
> ➢ Assemble sequences that have a portion of their names in common
> ➢ Automatically name contigs

In general, **Assemble by Name** follows in the tradition of other Sequencher functions; it saves you time and reduces errors. **Assemble by Name** is dependent on a new element in the Sequencher project that we call the *Assembly Handle*. The following example demonstrates how to define Handles in your project, and how to use these Handles to control **Assemble by Name**. We based this tutorial on what we anticipate is a typical **Assemble by Name** application, but we anticipate that you will find many more ways to take advantage of this flexible tool.

- Launch Sequencher to open an Untitled Project.
- Go to the **File** menu and select **Import/Sequencher Project…**
- Select the **Assemble by Name** project and select **Open**.

The Assemble by Name project contains 212 sequences. In addition to a Reference Sequence, the sequences include auto-sequencing data from fifteen different samples with nine different primers under a variety of conditions. The name of each sequence is written in the following format: *Source_Sample_Date_Oligo*. A typical example is A_FDR_01301882_Primer1F. Note that each of the four elements of the sequence name is separated from the adjacent elements by an underscore.

This tutorial will walk you through assembling 15 separate contigs, one for each sample, from the 212 sequences. Without **Assemble by Name**, this task requires three steps: selecting, assembling and naming. You would first have to select each sequence from a given sample, click on the **Assemble Automatically** button, and then rename the resulting contig with the sample name. For fifteen samples, you would have to repeat the three steps fifteen times. To create fifteen contigs using **Assemble by Name**, you select all of the sequences in the project and click on the **Assemble by Name** button. Sequencher separately selects and assembles the set of all sequences associated with each sample and then names each contig according to the resolved sample name.
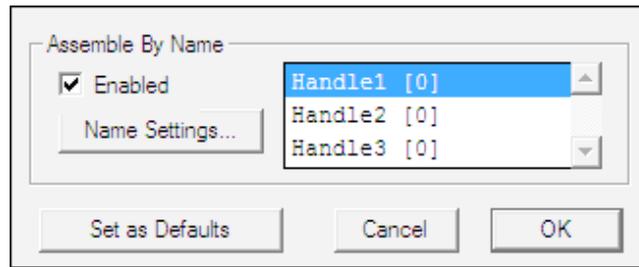
DEFINING THE ASSEMBLY HANDLE

In this example project, each element of the sequence name is separated by an underscore. In your own sequencing reactions you may have a different delimiter, such as a dash, or a period, or no delimiter at all. As long as you use a consistent naming scheme with your sequencing reactions, you should be able to use the **Assemble by Name** function to more quickly and efficiently assemble your fragments. You can choose the most common name delimiters from the **Assembly Parameters** dialog. [1]
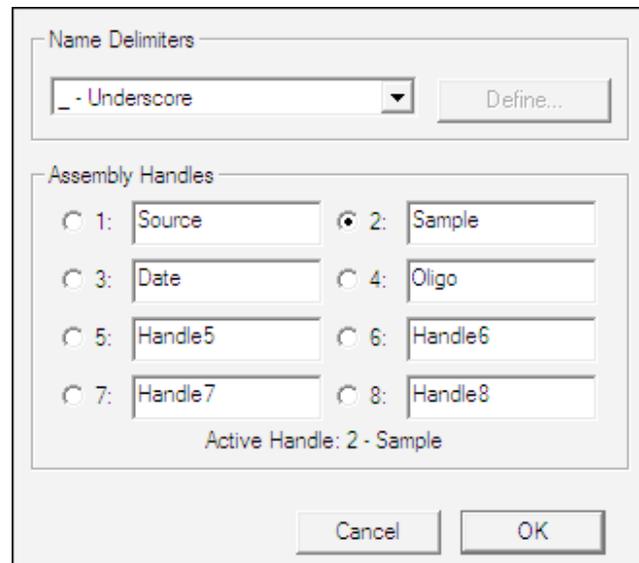
- Click on the **Assembly Parameters** button.

---

[1] In this example, we will focus on simple parsing rules; however, Sequencher also allows you to use regular expressions to define the Assemble by Name Handles. Regular expressions are described in greater detail in the Regular Expression tutorial.

The **Assemble by Name** parameters are at the bottom of this dialog.

- Place a check in the **Assemble by Name Enabled** box.
- Click on the **Name Settings…** button.
- Choose Underscore as the **Name Delimiter** from the drop down menu.
- Name the **Assembly Handles** as detailed below.

- Select the Radio button for **Sample** to define your **Active Handle**.
- Click on the **OK** button to return to the **Assembly Parameters** dialog.
- Close the **Assembly Parameters** dialog to set your new parameters by clicking on the **OK** button.

Sequencher displays the updated **Assembly Parameters** at the top of the Project window.

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

When you view the Project window now, you will notice several changes. First, you created a new column labeled Handle. Because you just selected the second element of the sequence name, Sample, as your Active Handle, the resolved handle name for each of the sequencing fragments is listed in the Handle column.
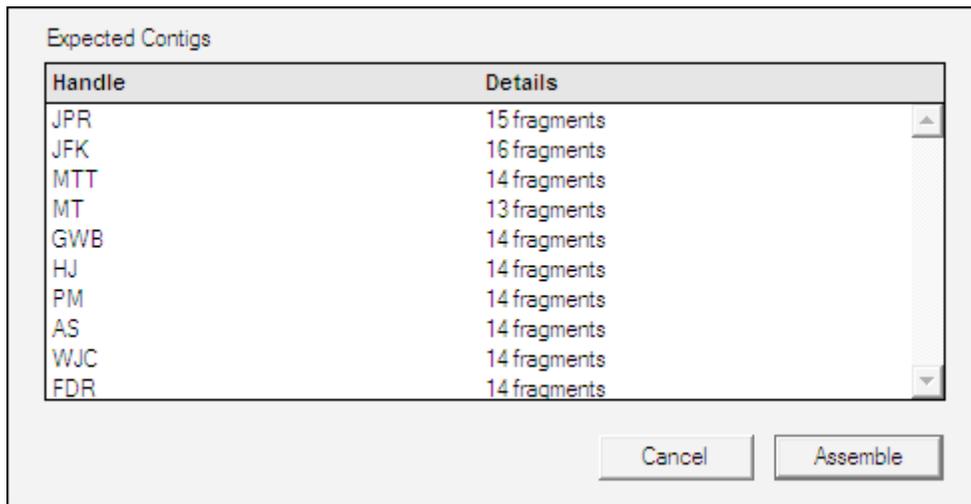
Also note the changes in the button bar. The text on the **Assemble Automatically** and **Assemble to Reference** buttons have changed to reflect the **Assemble by Name** function, and the **AbN** button is now darkened on Macintosh and depressed on Windows to further indicate that **Assemble by Name** is active.
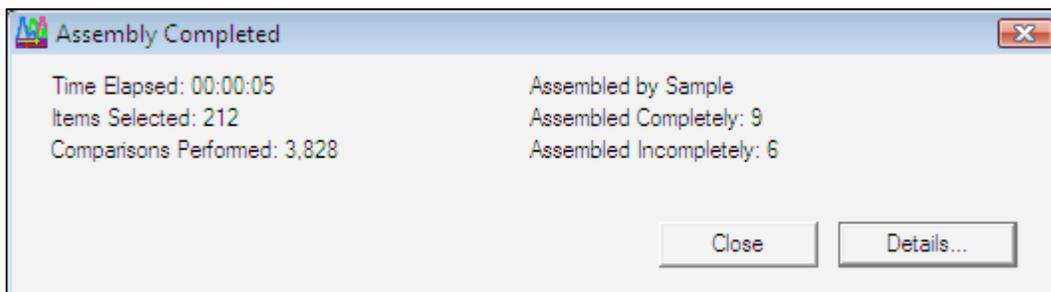
| ↑ | Assembly Parameters | AbN | Auto Assemble by Name | Assemble Interactively | To Reference by Name |

- From the **Select** menu, select **Select All** if all items in the Project window are not already selected.
- Click on the **Auto Assemble by Name** button.

Prior to performing Assembly by Name, Sequencher will provide you with an **Assembly Preview** dialog. The dialog lists each unique resolved handle name and the number of associated fragments and contigs.

Expected Contigs

Handle	Details
JPR	15 fragments
JFK	16 fragments
MTT	14 fragments
MT	13 fragments
GWB	14 fragments
HJ	14 fragments
PM	14 fragments
AS	14 fragments
WJC	14 fragments
FDR	14 fragments

Cancel   Assemble

- Click on the **Assemble** button.

Following assembly, Sequencher displays a summary of the results. Note that of the 15 potential contigs, 9 assembled completely into one contig while 6 did not.

Assembly Completed

Time Elapsed: 00:00:05   Assembled by Sample
Items Selected: 212   Assembled Completely: 9
Comparisons Performed: 3,828   Assembled Incompletely: 6

Close   Details...

A more detailed report is available for viewing and for export using the **Details...** command button.

- Click on the **Details...** command button.

Assembly Detail

Time Elapsed: 00:00:05
Items Selected: 212
Comparisons Performed: 3,828

Assembled by Sample
Assembled Completely: 9
Assembled Incompletely: 6

Completed Assemblies

Handle	Details
JFK	16 fragments
MTT	14 fragments
GWB	14 fragments
AS	14 fragments
WJC	14 fragments
FDR	14 fragments
PET	14 fragments
ABV	14 fragments
JK	12 fragments

Partial Assemblies

Handle	Items	Details
JPR	JPR[0008]	2 fragments
	JPR[0009]	8 fragments
	JPR[0010]	5 fragments
MT	MT[0031]	2 fragments
	MT[0032]	11 fragments
HJ	HJ[0046]	4 fragments
	HJ[0047]	10 fragments
PM	PM[0055]	4 fragments
	PM[0056]	10 fragments
TB	TB[0083]	2 fragments
	TB[0084]	2 fragments
	TB[0085]	4 fragments

Close   Save As...

The contigs created have names based on the resolved assembly handles and the contigs will include every selected sequence with that Assembly Handle. Note the names of the contigs in the upper portion of this detail window are the same as their Assembly Handles. Sequencer adds a numeric suffix to the names of Contigs that assemble incompletely. These are listed in the bottom portion of this window along with unassembled fragments.

If you click on the **Save As...** command button, you will generate a text document which also lists the details of the assembly. It may be useful in determining which samples require additional attention.

- **Close** the Assembly Detail dialog.

As with previously available assembly strategies, if you are not able to fully assemble a contig with the default assembly parameters, you may decrease the stringency and retry the assembly.

- Click the **Assembly Parameters** button on the Project window.
- Decrease the minimum overlap from 20 bases to 10 bases.
- Click **OK** to return to the Project window.
- Click on the **Auto Assemble by Name** button.
- Proceed through the Assembly Preview and the Assembly Complete dialogs.

Sequencher now assembles all of the sequences derived from a given sample into one contig, resulting in the 15 expected contigs.

Because it is so easy and fast to create large numbers of contigs with Assemble by Name, this feature allows you to easily dissolve your contigs and reassemble based on different handles.

- From the **Select** menu, select **Select All** if all items in the Project window are not already selected.
- From the **Contig** menu, select **Dissolve Contig**.
- From the next Sequencher dialog, select the **Dissolve Contigs** command button.
- Click the **Assembly Parameters** button on the Project window.

The Assembly is currently set to **Assemble By: Sample [15]**, as you chose in the Assemble by Name Settings
User Preferences. However, you can choose to assemble by any of the other available handles.
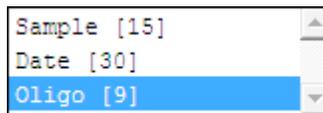
The numbers in square brackets count the items in your selection that share a common handle name. Therefore, you should anticipate 9 contigs if you assemble by Oligo.

- Change the **Assemble By Name** parameter to **Oligo**.
- Click **OK** to return to the Project window.
- Click the **Auto Assemble by Name** button.
- Proceed through the **Assembly Preview** and the **Assembly Complete** dialogs.

You'll notice that all 9 of the contigs assembled to completion.

ASSEMBLING TO REFERENCE[2] BY NAME

The **Assemble by Name** function saves you time and reduces the probability of your introducing error in the naming of a contig or in the selection of its fragments. The Reference Sequence facilitates sequence comparisons and instills consistency between assemblies. The combined function, Assemble to Reference by Name, provides you with the simultaneous advantages of both of these functions.

Assemble to Reference by Name duplicates the behavior of the standard Assemble by Name function, but with two modifications; a Reference Sequence is used as the template to direct the assembly, and a copy of the Reference is included in each contig. Including a Reference in your assembly also increases the likelihood that all of the related fragments will assemble into one contig.

- From the **Select** menu, select **Select All** if all items in the Project window are not already selected.
- From the **Contig** menu, select **Dissolve Contig**.
- Select **Dissolve Contigs** on the next dialog.
- Click the **Assembly Parameters** button on the Project window.
- Increase the minimum overlap from 10 bases back to 20 bases.
- Reset the **Assemble Handle** parameter to **Sample**.
- Click **OK** to return to the Project window.

In the first example, we used these settings for Assemble Automatically and only 9 of the potential 15 contigs assembled to completion. The others only assembled after we modified the assembly parameters. However, when assembling to Reference, you are not dependent on the fragments' overlap to build your contig. The Reference Sequence acts as a scaffold on which you can build your assemblies.

- Click the **To Reference by Name** button.
- Proceed through the **Assembly Preview** and the **Assembly Completed** dialogs.

You now have 15 contigs assembled, named, and ready for editing and comparison.

- Close the project without saving and exit Sequencer if desired.

---

[2] If you are not yet familiar with the advantages of working with a Reference Sequence, see the Reference Sequence Tutorial.