SEQUENCHER®

Tutorial for Windows and Macintosh

De Novo Sequence Assembly with Velvet

Gene Codes Corporation
525 Avis 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

De Novo Sequence Assembly with Velvet

About File Formats 3

Getting Started 3

Aligning Your Data with Velvet 5

Aligning Your Data with a Different Hash Length 6

Using the Advanced (Edit) Options 7

Using the Advanced (Edit) Options To Capture Unaligned Reads 8

Conclusion 10

De Novo Sequence Assembly with Velvet

Velvet is a de novo assembler, like **Sequencher's** default algorithm "Assemble Automatically". If you do not have a reference sequence or the reference sequence you do have is too evolutionarily distant to be useful, then you need to use de novo assembly. You can use this type of assembly whether you are working with a new organism, genome, chromosome, or region of DNA. Velvet itself consists of two programs, velveth, which analyzes the reads, and velvetg, which uses that analysis to perform the assembly.

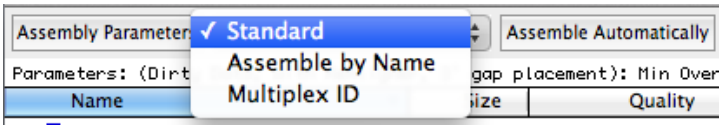
About File Formats

As with GSNAP reference-guided alignment, you can use reads in **FastA** or **FastQ** format. In single-end de novo assembly, all reads in a single file are compared to each other and contigs are built. In paired-end de novo assembly, reads from two files are used along with the paired information (distance between pairs) to construct longer scaffolds. As this is an NGS assembly, you don't need any reads to be imported into **Sequencher**; instead you tell **Sequencher** where to find the files it needs.

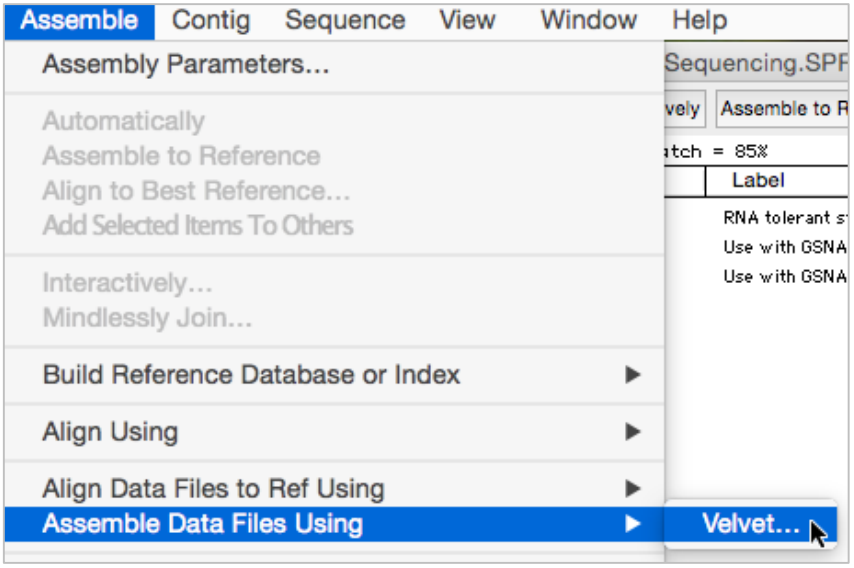
Getting Started

In this tutorial, you will use one of the Next-Gen algorithms in **Sequencher** to align your reads. We provide a sample set of data for you to use with this tutorial.

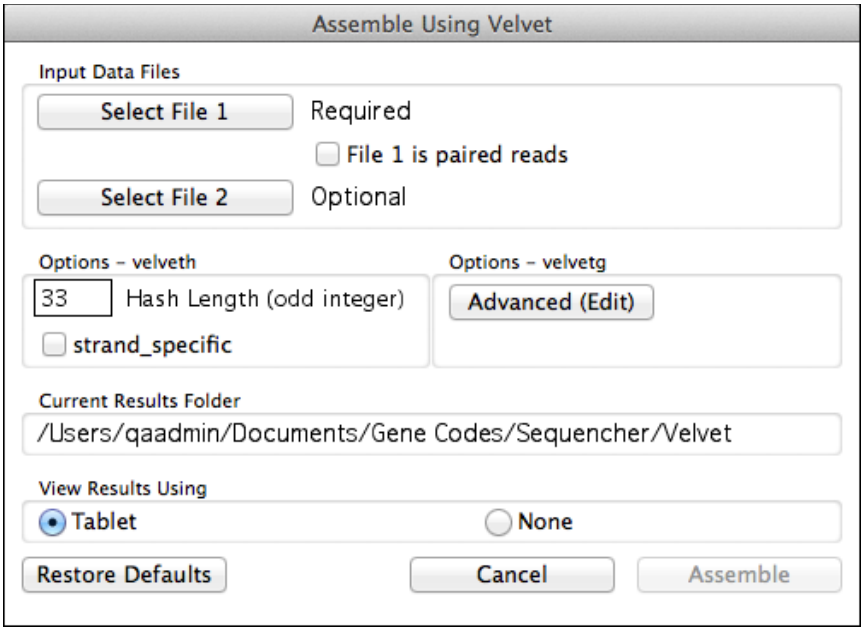
- Launch **Sequencher**.
- Set the **Assembly Mode** to **Standard**.



- Go to the **Assemble** menu and select **Assemble Data Files Using>Velvet...**



The **Assemble Using Velvet** and **External Data Browser** dialogs appear. The **External Data Browser** dialog is where you will go to monitor the progress/status of the assembly. The **Assemble Using Velvet** dialog is the launch point for your de novo assembly. You use this dialog to choose the data files that you are going to work with. You will be working with paired-end data so you will need to select two files. You may also want to decide whether you will view your results immediately or not.



- Click on the **Select File 1** button.
- Navigate to the **Sample Data** folder inside the **Sequencher** application folder, then to the **NGS Data** folder.
- Choose **read1.fq** and click on the **Open** button.

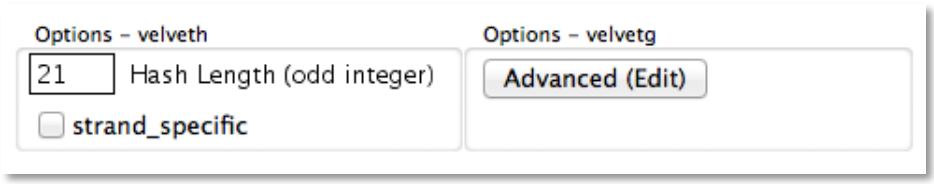
- Click on the **Select File 2** button.
- The file picker will automatically take you to the last location you chose which, conveniently, is where the second file you need is located.
- Choose **read2.fq** and click on the **Open** button.

Aligning Your Data with Velvet

The **Hash Length** is probably the single most important parameter when working with Velvet. It sets the minimum read length for consideration in an assembly. Setting the value higher will mean that fewer reads are used by Velvet and the assembly will occur more quickly, but the downside is that coverage may be reduced and may affect the overall accuracy.

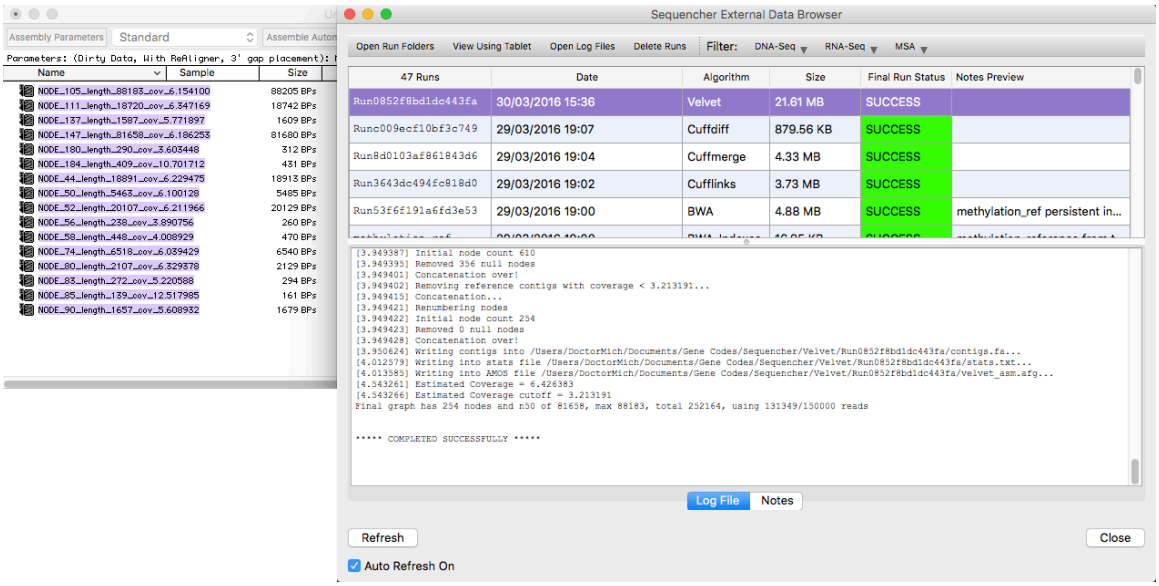
In this tutorial, you will be looking at the effect of changing the **Hash Length** on the same data set.

- Change the value in the **Hash Length** input field to **21**.



- Click on the **None** radio button in the **View Results Using** groupbox.
- Click on the **Assemble** button.

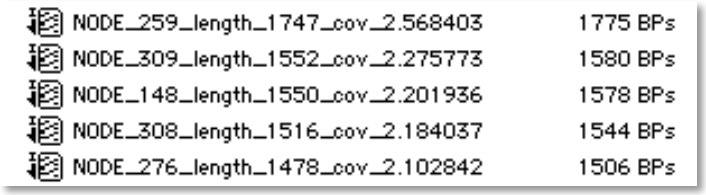
To monitor the progress of your assembly, look at the **External Data Browser** and watch as the Log File pane updates. Notice that, once your assembly is complete, the status field updates to SUCCESS. Note how many results are returned to **Sequencher's Project Window**. Each result is the consensus of a contig created by Velvet (which you can view using the Tablet genome viewer). Your run may not return exactly the same results as seen in the image below.



- Click on any sequence.
- Go to the **Contig** menu and select **Show NGS Data Using>Tablet**.
- When you have finished browsing with Tablet, quit the program.
- Before moving to the next section, ensure that no data is selected in the **Project Window** by going to the **Select** menu and choosing **Select None**.

Aligning Your Data with a Different Hash Length

- Go to the **Assemble** menu and select **Assemble Data Files Using>Velvet...**
- The **External Data Browser** and the **Assemble Using Velvet** dialogs will appear.
- Click on the **Select File 1** button.
- Navigate to the **Sample Data** folder inside the **Sequencher** application folder, then to the **NGS Data** folder.
- Choose **read1.fq** and click on the **Open** button.
- Click on the **Select File 2** button.
- The file picker will automatically take you to the last location you chose which conveniently is where the second file you need is located.
- Choose **read2.fq** and click on the **Open** button.
- Change the value in the **Hash Length** input field to **29**.
- Click on the **None** radio button in the **View Results Using** groupbox.
- Click on the **Assemble** button.



There are many consensus sequences in the **Project Window**. In the above image, the longest sequence's overall length is 1,775 bases. (Your results may vary from this.) Another thing to note is that, included in the name of each consensus, is the coverage of its original coverage, which has decreased. So in this case, increasing the **Hash Length** has not improved the result.

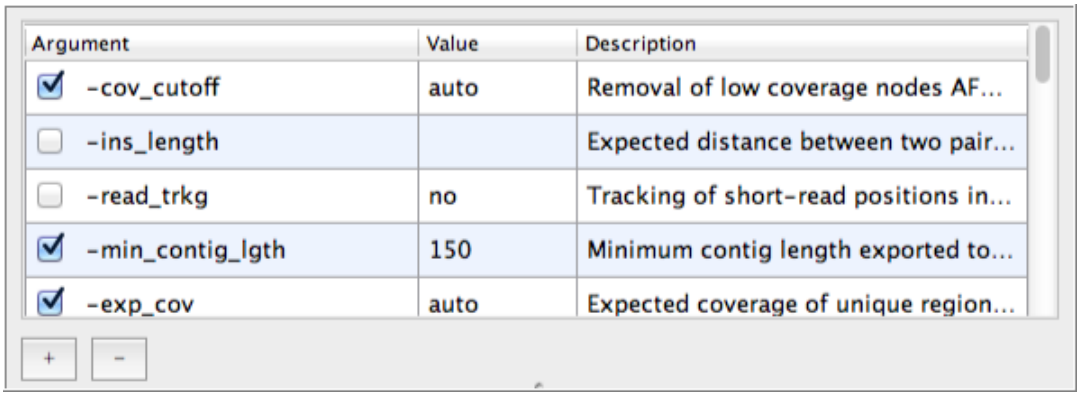
Using the Advanced (Edit) Options

Another way you can control the outcome of the assembly is to use the Advanced Parameters.

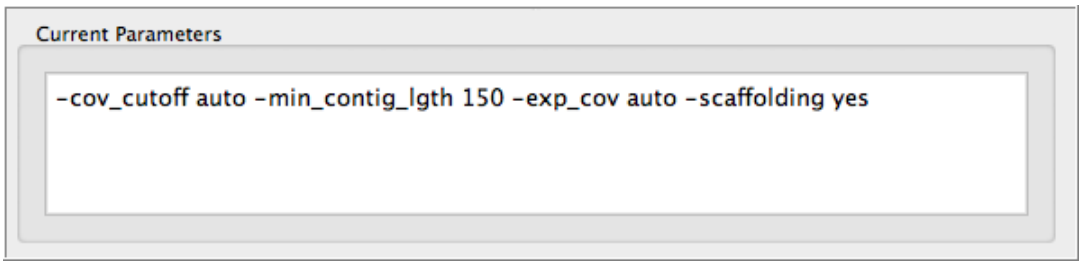
- Go to the **Assemble** menu and select **Assemble Data Files Using>Velvet**...
- Click on the **Advanced (Edit)** button.

The **Velvet Advanced Options** dialog opens. The dialog consists of two parts, a table with 3 columns and a **Current Parameters** preview window. You can read more about these parameters in the Velvet manual at this link. http://www.ebi.ac.uk/~zerbino/velvet/Manual.pdf.

You click in a checkbox (in the first column) to enable or disable a parameter. If you are used to using Velvet from the command line, this is the equivalent of choosing whether or not to include this argument on the command line. You click in the Value field (in the second column) to change the parameter's value. If you are uncertain about a specific parameter, then read its description (in the third column) first for more information.



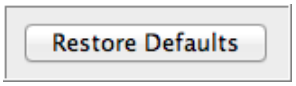
If there is a set of parameters you are not interested in, select the parameter you wish to remove and click on the – (dash) button. If there is a parameter you wish to add, click on the + (plus) button and enter the relevant information.



- Change the value for the parameter called **min_contig_lgth** to **100**.
- Scroll down the table to the parameter called **min_pair_count** and enable it by clicking on its checkbox.
- Click on the **OK** button.
- Click on the **Select File 1** button.
- Navigate to the **Sample Data** folder inside the **Sequencher** application folder, then to the **NGS Data** folder.
- Choose **read1.fq** and click on the **Open** button.
- Click on the **Select File 2** button.
- The file picker will automatically take you to the last location you chose which conveniently is where the second file you need is located.
- Choose **read2.fq** and click on the **Open** button.
- Click on the **Assemble** button.
- The consensus sequences appear in the **Project Window**. Select one of the newly created consensus sequences so it is highlighted.
- Go to the **Contig** menu and select **Show NGS Data Using>Tablet**.

If you find that your results are not getting better, then you may need to revert the parameters to their original values.

- Click on the **Restore Defaults** button on the **Velvet Advanced Options** dialog.



Note how the parameters in the **Current Parameters** preview window change back to the initial values you saw at the beginning of this tutorial.

Using the Advanced (Edit) Options To Capture Unaligned Reads

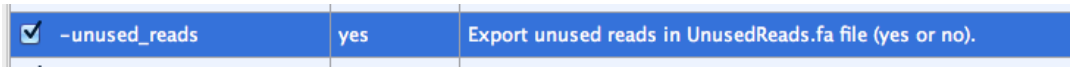
If you wish to capture any unaligned reads you will need to use an Advanced setting.

- Go to the **Assemble** menu and select **Assemble Data Files Using>Velvet**...
- If the **External Data Browser** dialog is not already opened, it will be.

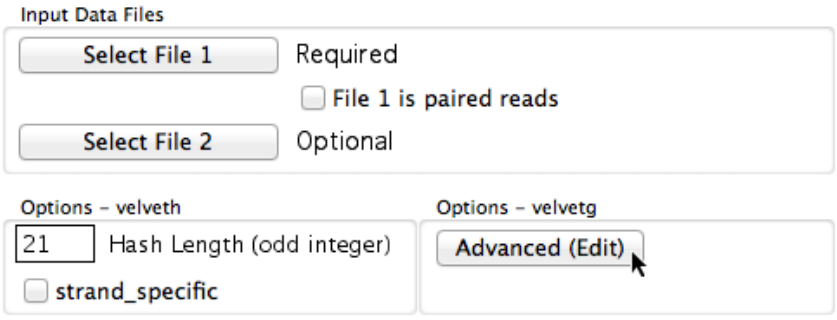
- Click on the **Advanced (Edit)** button.

The **Velvet Advanced Options** dialog opens. The dialog consists of two parts, a table with 3 columns and a **Current Parameters** preview window.

- Ensure the value for the parameter called **unused_reads** is set to **yes**.



- Set any other arguments you wish to use or change.
- Click on the **OK** button.



- Click on the **Select File 1** button.
- Navigate to the **Sample Data** folder inside the **Sequencher** application folder, then to the **NGS Data** folder.
- Choose **read1.fq** and click on the **Open** button.
- Click on the **Select File 2** button.
- The file picker will automatically take you to the last location you chose which conveniently is where the second file you need is located.
- Choose **read2.fq** and click on the **Open** button.
- Click on the **Assemble** button.
- The consensus sequences appear in the **Project Window**. Click anywhere in the **Project Window** to deselect all items.
- Select one of the newly created consensus sequences so it is highlighted. Right-click and select the **Open External Data Folder** menu item.

You will see a file called UnusedReads.fa that contains the unaligned reads in the appropriate Run folder.

Conclusion

In this tutorial, you have learned how to use the Velvet de novo assembler. You have learned how to use the advanced settings to refine its parameters in order to improve your results. You have also seen how using **Sequencher's** internal assembly algorithms can refine your results even further.

For more information on using **Sequencher**, this tutorial and others are a good place to start. You can also read the manual or consult our website by visiting http://www.genecodes.com.