

# User Guide to



# SeqMan Ultra<sup>TM</sup>

**DNASTAR, Inc. 2021**

# Table of Contents

<b>SeqMan Ultra Overview .....</b>	<b>3</b>
<b>SeqMan Ultra Tutorials.....</b>	<b>4</b>
Whole genome reference-guided workflow .....	5
Whole genome de novo workflow with mate pair data .....	15
Analysis of a whole genome de novo assembly.....	21
De novo assembly using Sanger data .....	25
RNA-Seq de novo transcriptome workflow .....	30
Gap closure workflow.....	37
Reference-guided alignment of Sanger/ABI data .....	40
<b>Assembly Projects .....</b>	<b>41</b>
Create a new assembly.....	42
Open an existing assembly .....	44
Getting Started .....	46
Navigate between open assemblies .....	47
View information about the currently open assembly .....	48
Project Overview .....	49
Project Details .....	51
Trimming .....	53
<b>Contigs.....</b>	<b>55</b>
View and select contigs in the Explorer panel .....	56
Organize contigs and scaffolds .....	58
Close a gap between contigs .....	62
Display contig information in a graphical or tabular view .....	65
Alignment View .....	67
Alignment view toolbar .....	69
View data tracks in the Alignment view .....	71
View multi-sample assemblies in the Alignment view .....	73
Edit sequences in the Alignment view.....	75
Strategy View .....	76
View data tracks in the Strategy view .....	77
Contig Report View .....	80
Contig Information .....	81
Contig Coverage .....	82
Contig Structural Variation.....	83
Search within a contig .....	85
<b>Data Tracks.....</b>	<b>87</b>
Edit track options .....	91
Change track height or spacing between tracks .....	92

<b>Features .....</b>	<b>93</b>
<b>Variants .....</b>	<b>96</b>
Filtering in the Variants view .....	99
Exporting variant data .....	102
<b>Make a Selection .....</b>	<b>103</b>
<b>View Information About a Selection .....</b>	<b>104</b>
View information about a selected track .....	106
View information about multiple selected tracks .....	107
<b>Search for Sequences Online .....</b>	<b>108</b>
Search for sequence matches (BLAST).....	109
Search wizard: Query .....	111
Search wizard: Search .....	114
Search wizard: Options.....	116
Search wizard: Job .....	119
Search for text or accession number matches (Entrez) .....	120
Jobs panel .....	123
Table view.....	125
Pairwise view .....	127
Text view .....	129
Shared tools in the Table, Text and Pairwise views .....	131
Filter sequence search results .....	132
Filter text search results .....	134
Access ID dialog .....	135
<b>Customize Appearance and Behavior .....</b>	<b>136</b>
Customize the appearance of views .....	137
Available color schemes .....	141
Customize tables in the views .....	145
Descriptions of table data columns .....	147
Customize the interface layout .....	156
Consensus parameters .....	159
Specify coverage parameters .....	161
Specify the naming convention for paired reads .....	162
Specify where to store temporary files.....	163
Opt in or out of showing update notifications .....	164
Opt in or out of usage logging .....	165
<b>Export Data .....</b>	<b>166</b>
<b>Save, Close and Exit .....</b>	<b>169</b>
<b>Appendix.....</b>	<b>170</b>

Troubleshoot the error “A different version is running” .....	171
IUPAC codes .....	172
Supported file types .....	174
Installed Lasergene file locations .....	175
Research references.....	177


# SeqMan Ultra Overview

---

SeqMan Ultra serves as both a workflow launcher for the SeqMan NGen assembler, and as the place to perform downstream analysis after assembly. In the latter role, SeqMan Ultra is used to analyze the consensus, features, and other components using a wide range of views and reports. BAM alignment files are not editable, but SeqMan Ultra can be used to edit *.sqd* formatted assemblies.

Assembly information is summarized in SeqMan Ultra's tabular reports and graphical views. The Strategy view graphically summarizes the position and orientation of every constituent sequence in a contig or contig scaffold and allows you to easily assess the coverage in your assembly. If coverage seems unsatisfactory, you can add more sequences and reassemble, or you can use the primer walking feature to drive the closure of gaps or to fill in low coverage areas. The Alignment view gives you a more detailed picture of the assembly, and allows you to edit constituent sequences, override the called consensus, review underlying trace or flowgram data, add new features, restore previously trimmed data, and adjust alignments. You may also evaluate putative variants identified by SeqMan NGen, which uses an exceptionally accurate variant-calling algorithm. You can also force contigs to join, or split a contig into two or more segments. Once you are satisfied with your assembly, you may save your project, export the data, or merge contigs with those imported from previous assembly projects.

## About this User Guide:

- For help **INSTALLING** Lasergene, see our separate [Installation Guide](#).
- To **PRINT** the current page of the User Guide, click the printer icon in the top right corner (  ).
- To download a **PDF** of the entire User Guide, scroll to the bottom of the table of contents on the left, and press **Download as PDF**.

# SeqMan Ultra Tutorials

---

Each of the following tutorials begin in SeqMan Ultra's "Welcome" window, proceed seamlessly to the SeqMan NGen wizard, and end with downstream analysis in SeqMan Ultra.

The first few steps of each tutorial feature a direct download link for that tutorial's data set. You will need to extract the .zip file before following along with the tutorial steps.

Tutorial number	Topic
1	<a href="#">Whole genome reference-guided workflow</a>
2	<a href="#">Whole genome <i>de novo</i> workflow with mate pair data</a>
3	<a href="#">Analysis of a whole genome <i>de novo</i> assembly</a>
4	<a href="#">De novo assembly using Sanger data</a>
5	<a href="#">RNA-Seq <i>de novo</i> transcriptome workflow</a>
6	<a href="#">Gap closure workflow</a>

# Whole genome reference-guided workflow

---

In this tutorial, you will create a reference-guided assembly using SeqMan NGen and then analyze the results using SeqMan Ultra. The time required for the assembly component is approximately 2-5 minutes.

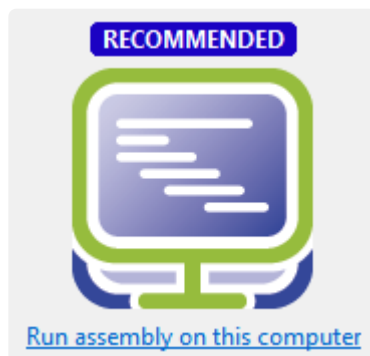
## Running a reference-guided assembly in SeqMan NGen:

1. Download [T1\\_Whole\\_Genome\\_Ref.zip](#) (1.4 GB) and extract the contents to any convenient location (e.g., your desktop). The folder contains the following sequences:
  - Reference sequence *DH10B\_NC010473.gbk*
  - Paired end sample sequences *SRR1284938\_1.fastq* and *SRR1284938\_2.fastq*
2. Launch SeqMan Ultra and choose **New Assembly** on the left. On the right, click on the **Genomics** workflow named **Variant analysis and resequencing**. This causes SeqMan NGen to open at the Workflow screen.
3. Choose the **NGS-Based** workflow named **Whole genome**.
4. In the Reference Sequence screen, use the **Add** button to add the sequence *DH10B\_NC010473.gbk*. Then click **Next**.
5. In the Input Sequences screen, change the **Experiment setup** to **Single sample**. Press the **Add** button to add the paired reads *SRR1284938\_1.fastq* and *SRR1284938\_2.fastq*. Then click **Next**.
6. In the Assembly Options dialog, click **Next**.
7. In the Analysis Options screen, change the **Variant detection mode** to **Haploid** (since this is a bacterial organism) and the **SNP filter stringency** to **High**. Then click **Next**.
8. In the Assembly Output screen, type “Templated E coli” into the **Project Name** text box. This name will be assigned to all output files, including the finished assembly. Use the **Browse** button to specify a **Project Folder** for your assembly output files. For local users, an alternative way to select a location is to drag and drop a folder from the file explorer onto the **Project folder** row. The folder you choose must be in a writable location. Then click **Next**.
9. In the **Run Assembly Project** screen, note that the **Estimated coverage** is “375X.”

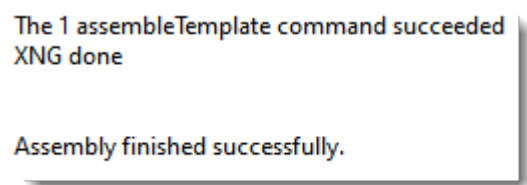
Input information	
Total length reference sequences:	4.7 MBases
Estimated input read bases:	1.8 GBases
Estimated coverage:	<span style="border: 2px solid red; padding: 2px;">⚠ 375X</span>

A coverage of 50-100X is ideal, and additional coverage only serves to increase the assembly time. There are several wizard parameters you can change to reduce coverage; here, you'll reduce the **Maximum total reads**.

10. Click **Assembly Options** in the left margin to return to that screen. Check the **Maximum total reads** box and type in 2000000 (2 followed by six zeros).
11. Click **Run Assembly Project** in the left margin to return to that screen. Note that the **Estimated coverage** is now "64X." Lower down the screen, observe that the recommendation is to perform a local assembly. Press **Run assembly on this computer**.




12. Wait until being informed that assembly has finished (approx. 2-5 minutes), then click **Next**.

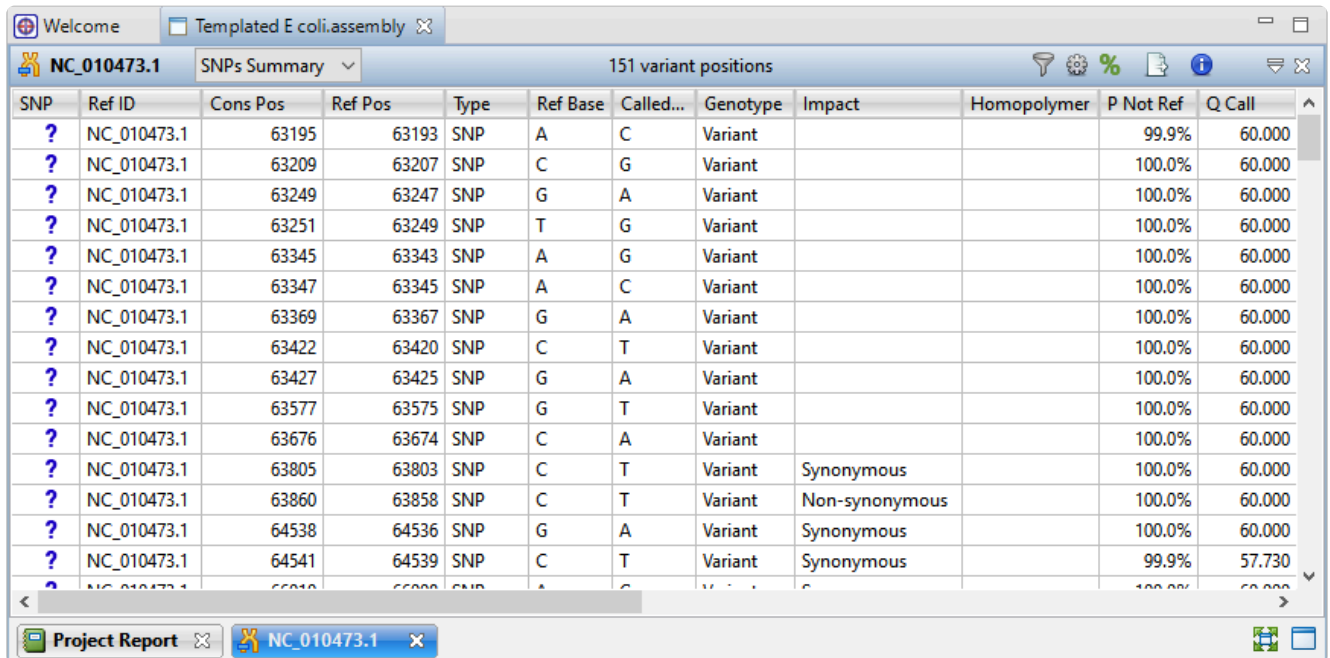


13. From the Assembly Summary screen, click **Open assembly** to launch the results in SeqMan Ultra.
14. Close the SeqMan NGen project by clicking the **Finish** button and confirming **Yes**.

#### Analyzing variants and structural variations in SeqMan Ultra:


1. SeqMan Ultra's Explorer panel (**Explorer** tab on the upper right of the window) contains a single contig containing the bacterial chromosome. The contig is already selected by default.

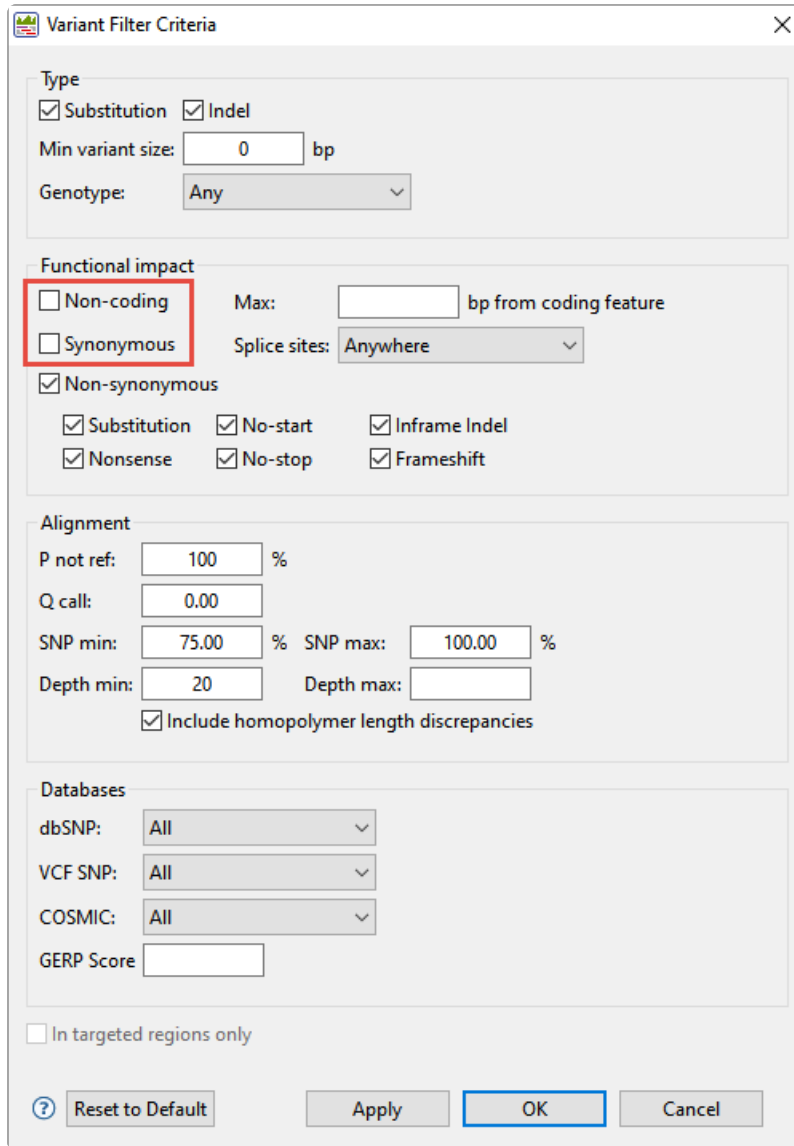
- Press the **Show table of variants** tool (  ) to the right of the Explorer panel to reveal 151 variant positions. This number is displayed in the view header.



SNP	Ref ID	Cons Pos	Ref Pos	Type	Ref Base	Called...	Genotype	Impact	Homopolymer	P Not Ref	Q Call
?	NC_010473.1	63195	63193	SNP	A	C	Variant			99.9%	60.000
?	NC_010473.1	63209	63207	SNP	C	G	Variant			100.0%	60.000
?	NC_010473.1	63249	63247	SNP	G	A	Variant			100.0%	60.000
?	NC_010473.1	63251	63249	SNP	T	G	Variant			100.0%	60.000
?	NC_010473.1	63345	63343	SNP	A	G	Variant			100.0%	60.000
?	NC_010473.1	63347	63345	SNP	A	C	Variant			100.0%	60.000
?	NC_010473.1	63369	63367	SNP	G	A	Variant			100.0%	60.000
?	NC_010473.1	63422	63420	SNP	C	T	Variant			100.0%	60.000
?	NC_010473.1	63427	63425	SNP	G	A	Variant			100.0%	60.000
?	NC_010473.1	63577	63575	SNP	G	T	Variant			100.0%	60.000
?	NC_010473.1	63676	63674	SNP	C	A	Variant			100.0%	60.000
?	NC_010473.1	63805	63803	SNP	C	T	Variant	Synonymous		100.0%	60.000
?	NC_010473.1	63860	63858	SNP	C	T	Variant	Non-synonymous		100.0%	60.000
?	NC_010473.1	64538	64536	SNP	G	A	Variant	Synonymous		100.0%	60.000
?	NC_010473.1	64541	64539	SNP	C	T	Variant	Synonymous		99.9%	57.730

Imagine that you wish to identify those variants which are most likely to alter function of genes in *E. coli*. You will therefore want to focus on variations predicted to have a non-synonymous impact on the *E. coli* genome. In the next step, you will filter out some of the variants and only leave those predicted to cause a coding change.

- To show only variants that cause coding changes, click the **Filter all variant tables** tool (  ) in the upper right of the view. In the Variant Filter Criteria dialog, remove the checkmarks next to **Non-coding** and **Synonymous** and click **OK**.

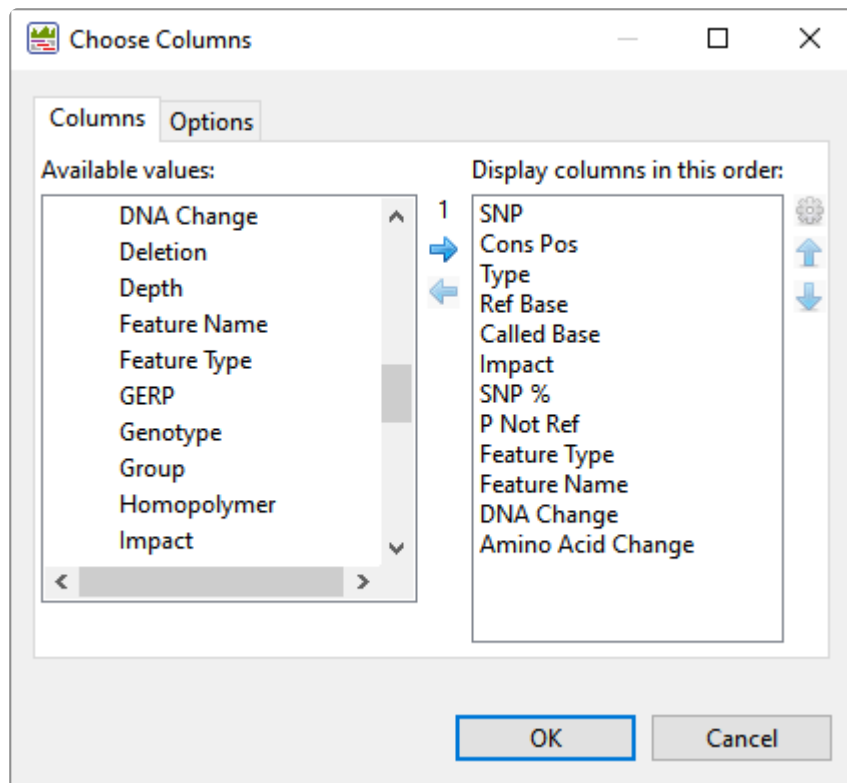


The image shows the 'Variant Filter Criteria' dialog box with the following settings:

- Type:**
  - ☒ Substitution
  - ☒ Indel
  - Min variant size:  bp
  - Genotype:
- Functional impact:**
  - ☐ Non-coding (highlighted with a red box)
  - ☐ Synonymous (highlighted with a red box)
  - ☒ Non-synonymous
    - ☒ Substitution
    - ☒ No-start
    - ☒ Inframe Indel
    - ☒ Nonsense
    - ☒ No-stop
    - ☒ Frameshift
  - Max:  bp from coding feature
  - Splice sites:
- Alignment:**
  - P not ref:  %
  - Q call:
  - SNP min:  %
  - SNP max:  %
  - Depth min:
  - Depth max:
  - ☒ Include homopolymer length discrepancies
- Databases:**
  - dbSNP:
  - VCF SNP:
  - COSMIC:
  - GERP Score:
- ☐ In targeted regions only
- Buttons:

The remaining 22 variants all have a non-synonymous impact or consist of an in-frame deletion.

- Use the scrollbar on the bottom of the view to scroll right, and observe the many data columns available in this report. While each column is useful in certain circumstances, many will not be needed for this tutorial workflow.
- To remove unneeded columns from the report, click the **Change alignment options** tool (⚙️). Using the image below as a guide, move all of the wanted columns “up” into the position shown below using the up arrow. Then select the remaining, unwanted, columns and press the left arrow to dismiss them. Press **OK**.



Note that the view is now much more compact and may be easier to view in its entirety, depending on your monitor size.

Now turn your attention to the black arrows in some of the **SNP** column cells. These represent “coalesced variants.” In *.assembly* projects like this one (but not *.sqd* projects), variants in adjacent columns are coalesced into a single insertion or deletion if they are of the same type, and if at least 80% of the reads with the called variant in one column have a variant in the adjacent column.

- Click any of the triangles to open the coalesced multiple-base indels and reveal individual variants. After clicking a triangle, information on each position of the insertion/deletion is displayed in a separate row.

NC_010473.1 SNPs Summary 22 variant positions											
0 confirmed 0 rejected 22 putative 0 mixed 158 filtered											
SNP	Cons ...	Type	Ref Base	Called Base	Impact	SNP %	P Not Ref	Feature Type	Feature Name	DNA Change	Amino Acid Change
?	63860	SNP	C	T	Non-synonymous	100.0%	100.0%	CDS	mraZ	c.121C>T	p.H41Y
?	218679	SNP	T	A	Non-synonymous	100.0%	100.0%	CDS	yafJ	c.239T>A	p.L80Q
▼ ?	418879	SNP [2]	AA	GG	Non-synonymous	100.0%	100.0%	CDS	hha	c.13_14delinsCC	p.F5P
?	418879	SNP	A	G	Non-synonymous	100.0%	100.0%	CDS	hha	c.14T>C	p.F5S
?	418880	SNP	A	G	Non-synonymous	100.0%	100.0%	CDS	hha	c.13T>C	p.F5L
?	751890	SNP	A	G	Non-synonymous	100.0%	100.0%	CDS	nagD	c.289T>C	p.Y97H
?	844877	SNP	A	G	Non-synonymous	100.0%	100.0%	CDS	galE	c.368T>C	p.F123S
▶ ?	862407	SNP [2]	AG	CA	Non-synonymous	100.0%	100.0%	CDS	bioA	c.40_41delinsTG	p.L14W
?	971162	SNP	A	G	Non-synonymous	100.0%	100.0%	CDS	ybjD	c.1494A>G	p.I498M

- Look at the **Impact** column for the variant in the first row (Cons Pos = 63860), and note that a non-synonymous mutation is predicted.

- Double-click on the row to launch the Alignment view with the variant highlighted.

The Alignment View displays the consensus sequence, and the alignment of all sequences making up a contig, at the nucleotide level. The consensus sequence for the selected contig is displayed at the top of the window, immediately beneath the ruler. Constituent sequences are displayed in the bottom section.

- Click to either side of the highlighted column to temporarily clear the selection. This will allow for a better view of the color-coding used by SeqMan Ultra.

NC\_010473.1 Position=17

Ruler

Consensus

Coverage

NC\_010473.1(1>4686377) CATTGACATTTCATCACCCGTGCCTGCT

ID\_1698742\_f C

ID\_655022\_r CATT

ID\_820868\_r CATT

ID\_713240\_r CATTGACA

ID\_1852302\_r CATTGACATTAT

ID\_914064\_r CATTGACATTATCAC

ID\_1636180\_f CATTGACATTATCAC

ID\_115384\_r CATTGACATT

ID\_1775994\_f CATTGACATTATCACCC

ID\_461262\_f CATTGACATTATCACCCG

ID\_1254702\_f CATTGACATTATCACCCGTGCCT

ID\_1471690\_f CATTGACATTATCACCCGTGCCTGCT

ID\_352602\_r CATTGACATTATCACCCGTGACCACT

ID\_461714\_f CATTGACATTATCACCCGTGCCTGCT

ID\_514280\_f CATTGACATTATCACCCGTGCCTGCT

NC\_010473.1 x

NC\_010473.1 SNPs Summary 1 selected rows; 22 variant positions


0 confirmed 0 rejected 22 putative 0 mixed 158 filtered

SNP	Cons Pos	Type	Ref Base	Called...	Impact	SNP %
?	63860	SNP	C	T	Non-synonymous	100.0%
?	218679	SNP	T	A	Non-synonymous	100.0%

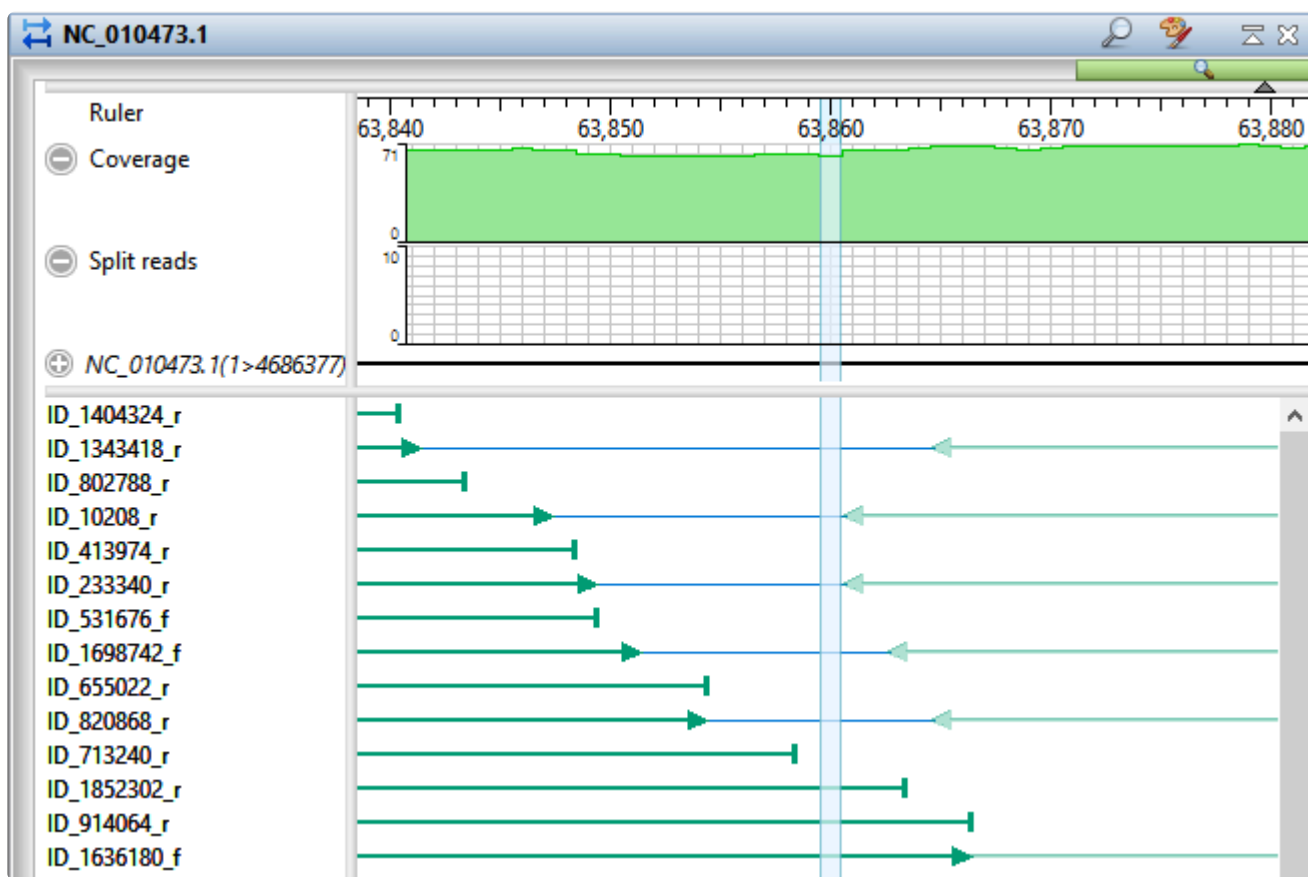
Note that the **Reference** base at the position of interest is **C**, while all of the samples contain **\*T** at that position. The yellow highlighting used for variants is specified in the Alignment area of the Style panel and can be changed, if desired.

- To view the variant in the Strategy view, select **View > Strategy > Show Strategy View**. Because two

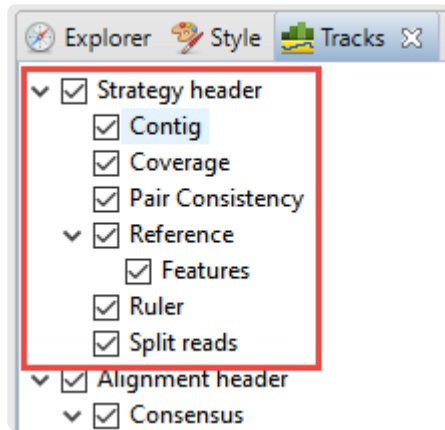
views are already open, the Strategy view appears in a pop-up window. The view graphically summarizes the position and orientation of every constituent sequence in the selected contig.

11. Zoom in as far as possible by dragging the green Zoom slider (  ) all the way to the right.
12. Click on the top row of the Variants view table to again move the cursor to the variant of interest.

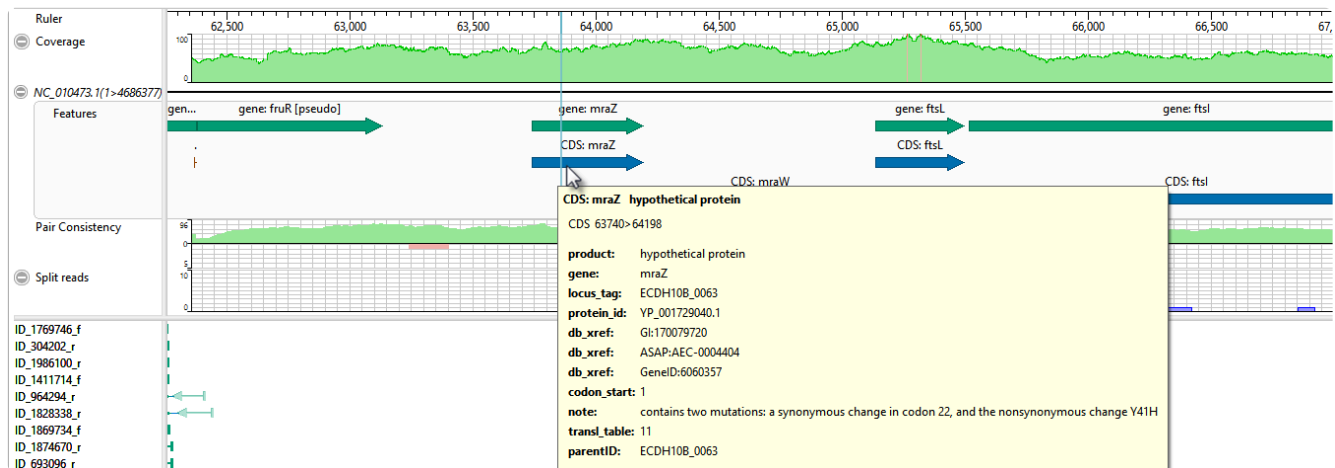
In the lower part of the window, observe the green arrows. Arrows are used to graphically represent the individual sequences in the contig. The point of the arrowhead represents the 3' end and the tail represents the 5' end. The green coloration of arrows in this region denotes consistent, correctly-specified paired reads.



13. To add data tracks to enhance the Strategy view, click on the Tracks tab and check all boxes in the "Strategy header" section.



14. Returning to the floating Strategy view window, click any plus signs appearing next to the header tracks to reveal the “hidden” tracks.
15. Move the horizontal zoom slider (at the top) until you can read the feature names, as shown below. Hover over any part of a histogram to display information about that location.



- **Coverage** – Represents the coverage within the assembly, with green representing above-threshold coverage and red representing areas of low coverage.
- **Split reads** – Shows the location and depth of split reads.
- **Features** – The putative variant is located within a labeled CDS called *mraZ*.
- **Pair Consistency** – Shows consistent reads in green and inconsistent in red. The deeper the histogram, the more recognized pairs of reads there are.

16. Close the floating Strategy view window by clicking the ‘x’ in the top corner. Then direct your view to the Variants view at the bottom of the main window.
17. Confirm the variant that you just investigated by clicking on the **SNP** (leftmost) column to change the

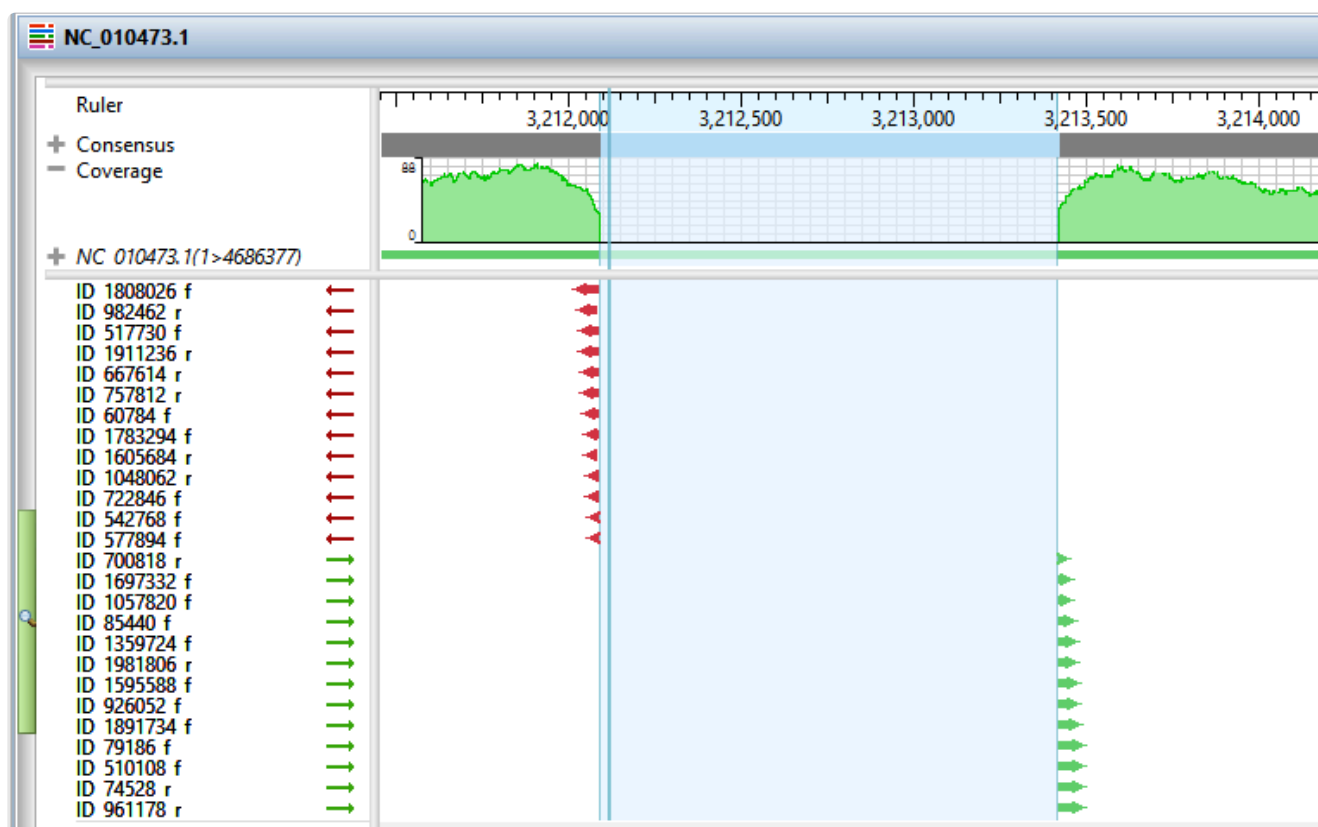
question mark to a checkmark. (If you had found supporting evidence was lacking, you could have clicked a second time to reject the variant and add an 'x'.)

Short indels are included in the Variants view. However, any indels long enough to inhibit assembly are instead gathered in another report called the Structural Variation Report.

18. To see the longer indels, choose **Contig > NC\_010473.1 Structural Variation**. Scroll down to the putative Indel at 3211929.

NC_010473.1		Contig Structural Variation				
Contig ID	Ref Pos	Length	Type	Splits	Cover...	Region Capture
NC_010473.1	3003961	1331	Del	29	34.0	no
NC_010473.1	3019404	56934	Indel	-----	62.0	no
NC_010473.1	3170620	1190	Del	44	44.0	no
NC_010473.1	3170626	1184	Indel	-----	44.0	no
NC_010473.1	3199470	1328	Indel	-----	.0	no
NC_010473.1	3211929	1328	Indel	-----	.0	no
NC_010473.1	3229073	44443	Indel	-----	63.0	no
NC_010473.1	3336619	20201	Indel	-----	62.0	no
NC_010473.1	3381608	11894	Indel	-----	63.0	no

19. Double-click on that table row to select it and center the Alignment view on that position. Use the horizontal zoom slider and vertical scrollbar and note that this area is devoid of reads.



This marks the end of this tutorial.

# Whole genome de novo workflow with mate pair data

---

In this tutorial, you will use SeqMan NGen to de novo assemble an *E. coli* K12 MG155 data set composed of Illumina MiSeq (2×300) paired end reads.

The default stringency setting is designed to produce longer contigs; however, some may contain false joins. You'll learn how to manually examine contigs for false joins and split them into two contigs. Finally, you'll realign all the contigs, including the newly split ones, using SeqMan Ultra.

## Assembling the data in SeqMan NGen:

1. Download [T2\\_Whole\\_Genome\\_MatePair.zip](#) (231 MB) and extract it to any convenient location (e.g., your computer's desktop). The folder contains the following paired end reads:  
  

```
Ecoli_S1_L001_R1_001.fastq
```

```
Ecoli_S1_L001_R2_001.fastq
```
2. Launch SeqMan Ultra and choose **New Assembly** on the left. On the right, click on the **Genomics** workflow named **De novo genome assembly and editing**. This causes SeqMan NGen to open at the Workflow screen.
3. On the right of the screen, under **NGS-Based**, click the workflow named **De novo assembly**.
4. In the Input Sequences screen, press the **Add** button to add the two paired reads whose names begin with *Ecoli\_S1\_L001\_R*. Then click **Next**.
5. In the Preassembly Options screen, change the **Maximum total reads** to 800000 (8 followed by five zeros). Click **Next**.
6. In the Assembly Options dialog, type in an **Estimated genome length** of 4600000 (46 followed by five zeros). Click **Next**.
7. In the Assembly Output screen, type "De novo mate pair" into the **Project Name** text box. This name will be assigned to all output files, including the finished assembly. Use the **Browse** button to specify a **Project Folder** for your assembly output files. For local users, an alternative way to select a location is to drag and drop a folder from the file explorer onto the **Project folder** row. The folder you choose must be in a writable location. Then click **Next**.
8. In the **Run Assembly Project** screen, press **Run assembly on this computer**.
9. Wait until being informed that assembly has finished (approx. 1 hour), then click **Next**.

10. From the Assembly Summary screen, click **Open assembly** to launch the results in SeqMan Ultra.
11. Close the SeqMan NGen project by clicking the **Finish** button and confirming **Yes**.

### Correcting misjoined contigs and realigning data in SeqMan Ultra:

After following the steps above, your *.sqd* assembly file is now open in SeqMan Ultra. If analyzing your own data in real life, you would keep the current file open and proceed with the analysis. For purposes of this tutorial, however, you'll close this assembly and open a similar one that is included in your tutorial data. Why? No two assemblies will be completely identical, due to randomization of the order in which contigs are created by the assembler. Because this tutorial will refer to specific contigs, it will be easier to follow along if you are using a pre-made assembly. Therefore, you will close the current assembly in the first step, below.

1. Use **File > Close Editor** to close the current file, *De novo mate pair.sqd*.
2. Open the pre-made assembly by using **File > Open** to open the demo data file *Sample de novo mate pair.sqd*.
3. Observe the information in the Project Overview on your screen. Briefly note the high **Contig N50** value of 149 kb. This means that half of the genome sequence is located in contigs equal to or longer than 149 kb.

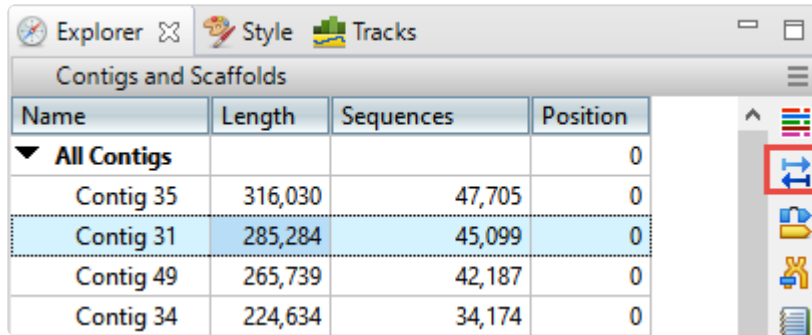
#### **Sample de novo mate pair.sqd**

C:\Users\yildizs\Desktop\T2\_Whole\_Genome\_MatePair

Contig N50:	149.0 kb
Largest contig size:	316,030 bp
Number of contigs:	66
Reads assembled:	720,452

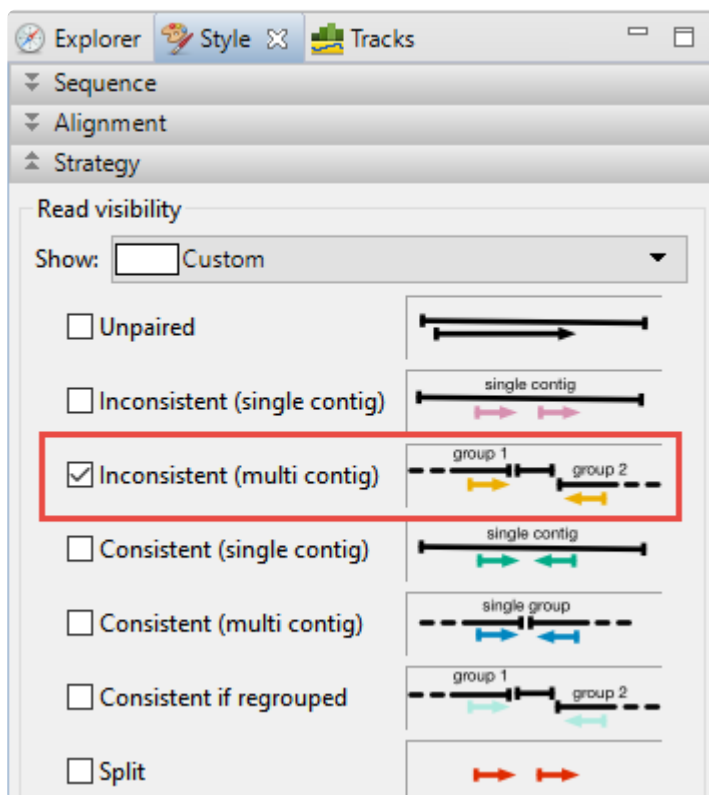
Using a high stringency in SeqMan NGen would likely have resulted in about ten times as many contigs. However, a tradeoff is that some contigs might have been joined when they should not have been. You will address that issue in the next few steps. In “real life,” you would check all of the contigs for false joins (or at least the contigs longer than 100k) and split them where needed. For purposes of this tutorial, you will only examine two contigs: 31 and 38.

4. In the Explorer panel on the right, select Contig 31, which is one of the longer contigs in the project. Then press the **Show strategy view** tool.

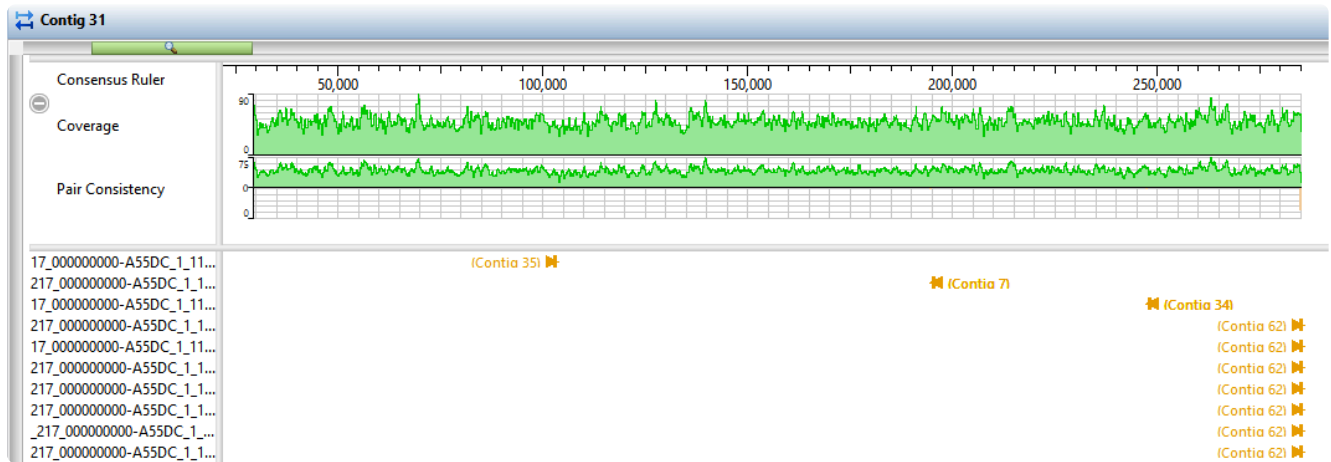


Name	Length	Sequences	Position
▼ All Contigs			0
Contig 35	316,030	47,705	0
Contig 31	285,284	45,099	0
Contig 49	265,739	42,187	0
Contig 34	224,634	34,174	0

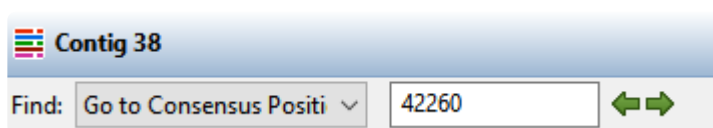
5. You will be looking for locations along the contig that are flagged for inconsistent pairs. To easily see inconsistent pairs, click the **Style** tab on the right. In the **Strategy** section, uncheck all boxes except **Inconsistent (multi-contig)**.



6. In the Strategy view on the left, slide the green horizontal zoom slider to the left so the entire contig is visible. Note that the **Pair Consistency** histogram near the top does not show any areas of orange, meaning that there are no areas of inconsistency. Looking at the orange arrows at the bottom, ignore the ones at the right edge; these indicate that Contig 31 is likely adjacent to Contig 62. Instead, look for orange arrows that are not on the edges. In this case, there are only three orange arrows that represent possible inconsistent pairs. The very small number indicates that these three are likely artifacts. Contig 31 thus exemplifies a correctly-built contig without false joins.



7. Click the **Explorer** tab and select Contig 38. Click the **Show strategy view** tool, then click the **Show Alignment view** tool just above it.
8. In the Strategy view (upper view), slide the green horizontal “zoom slider” all the way to the left so you can see the entire sequence. In the Alignment view (lower view), slide the zoom slider all the way to the right for a detailed view.
9. In the Strategy view (upper view), look at the area around 42,000. Both the **Pair Consistency** histogram and the large number of arrows denote that this is a likely misjoin. The arrows facing left have pairs in Contig 47, while the arrows facing right have pairs in Contig 55. Contig 38 represents a contig with at least one false join. (There is another candidate at the very left of the contig).
10. Near the upper right corner of the Alignment view (lower view), click the **Search alignment** tool (🔍). On the left, use the **Find** menu to select **Go to Consensus Position**. Then type **42260** into the search bar and press either green arrow to navigate to the position.



11. The position 42,260 is now selected, but the cursor needs to be between two nucleotides. In the Alignment view, click just to the right or left of 42,260 to remove the selection
12. Choose **Contig > Split At Insertion**. A new contig is added to the Explorer panel. The new contig is the only one in the table with a **Position** other than zero.

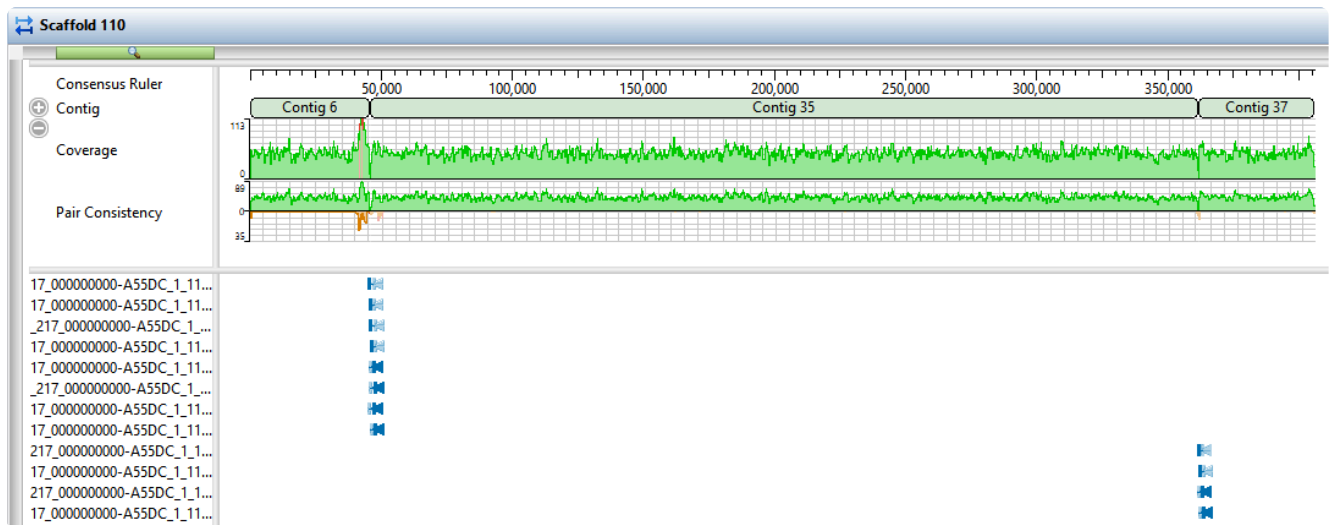
In a real-world project, we recommend going through each contig as above. For this tutorial, however, splitting Contig 38 is sufficient. In the next steps, you will create contig scaffolds.

13. In the Explorer panel, select the upper-most row, **All Contigs**. Right-click on the selection, choose **Order Contigs into Scaffolds** and press **OK**. SeqMan Ultra creates 23 scaffolds numbered 100 to

230.

14. Click on **Scaffold 110**. Then right-click on the selection and choose **Show Strategy View**. Move the horizontal zoom slider to the left until you can see all three contigs, which appear as light green bars at the top of the view.
15. To display consistent pairs click on the **Style** tab. In the **Strategy** section, uncheck all boxes except **Consistent (multi-contig)**.

The blue arrows where two contigs meet provide visual verification that the scaffold is based on good evidence from the paired reads.



For this tutorial, we'll assume that each of the other scaffolds has also been built correctly.

16. In the Explorer panel, select the **Scaffold 100** row. Choose **Contig > Align Contigs End to End** and press **OK**. Repeat for all other Scaffold rows. Most scaffolds will collapse into 1-2 contigs. Some scaffolds (e.g., Scaffold 170) will not collapse at all, and will give a message to that effect. In these case, simply move on to the next scaffold on the list.
17. One at a time, drag each contig that remains in a scaffold and drop it anywhere above **Scaffold 100**. Then use **Shift+click** to select all of the now empty scaffolds and press the **Delete** key on your keyboard.
18. Choose **Project > Project Overview** and observe how the steps above have improved the assembly, even though you only split one false join.

## Sample de novo mate pair.sqd

C:\Users\yildizs\Desktop

Contig N50: 204.0 kb  
Largest contig size: 361,504 bp  
Number of contigs: 44  
Reads assembled: 720,452

Note that the **Contig N50** has increased from **149kb** to **204kb**, meaning the typical contig is now almost 37% longer than before. The longest contig is over 361,000 bp in length. The number of contigs decreased from 66 to 44. Splitting off additional false joins between Steps 8-9 would likely have improved the assembly even more.

This marks the end of this tutorial.

# Analysis of a whole genome de novo assembly

DNASTAR Lasergene lets you set up a *de novo* assembly with ease. One of the outputs is an editable *.sqd* file that can be opened and edited in SeqMan Ultra. In SeqMan Ultra, you can evaluate the assembled contigs, edit them, organize them into scaffolds, and close any gaps in those scaffolds.

This tutorial uses a de novo whole genome assembly that was assembled from two MiSeq 2×300 paired-end read files from *E.coli* K12 MG1655. SeqMan NGen was used to assemble 2.5M reads from this data set, producing an assembly with a large contigN50 of 203Kbases in 51 contigs. Due to the size of the reads (28 GB combined), this tutorial will begin at the stage of downstream analysis in SeqMan Ultra.

1. Download [T3\\_Whole\\_Genome\\_DeNovo.zip](#) (332 MB) and extract it to any convenient location (i.e., your desktop). The data set consists of a single SeqMan Ultra project named *E.coli K12 MG1655 MiSeq de novo.sqd*.
2. Launch SeqMan Ultra and use **File > Open** to open *E.coli K12 MG1655 MiSeq de novo.sqd*. Note that the Project Overview shows that the Contig N50 is 204.0 kb, which is quite large. Also observe that the assembly resulted in 51 contigs.


## E.coli K12 MG1655 MiSeq de novo.sqd

C:\Users\yildizs\Desktop\T3\_Whole\_Genome\_DeNovo

Contig N50: 204.0 kb  
 Largest contig size: 361,321 bp  
 Number of contigs: 51  
 Reads assembled: 2,182,744

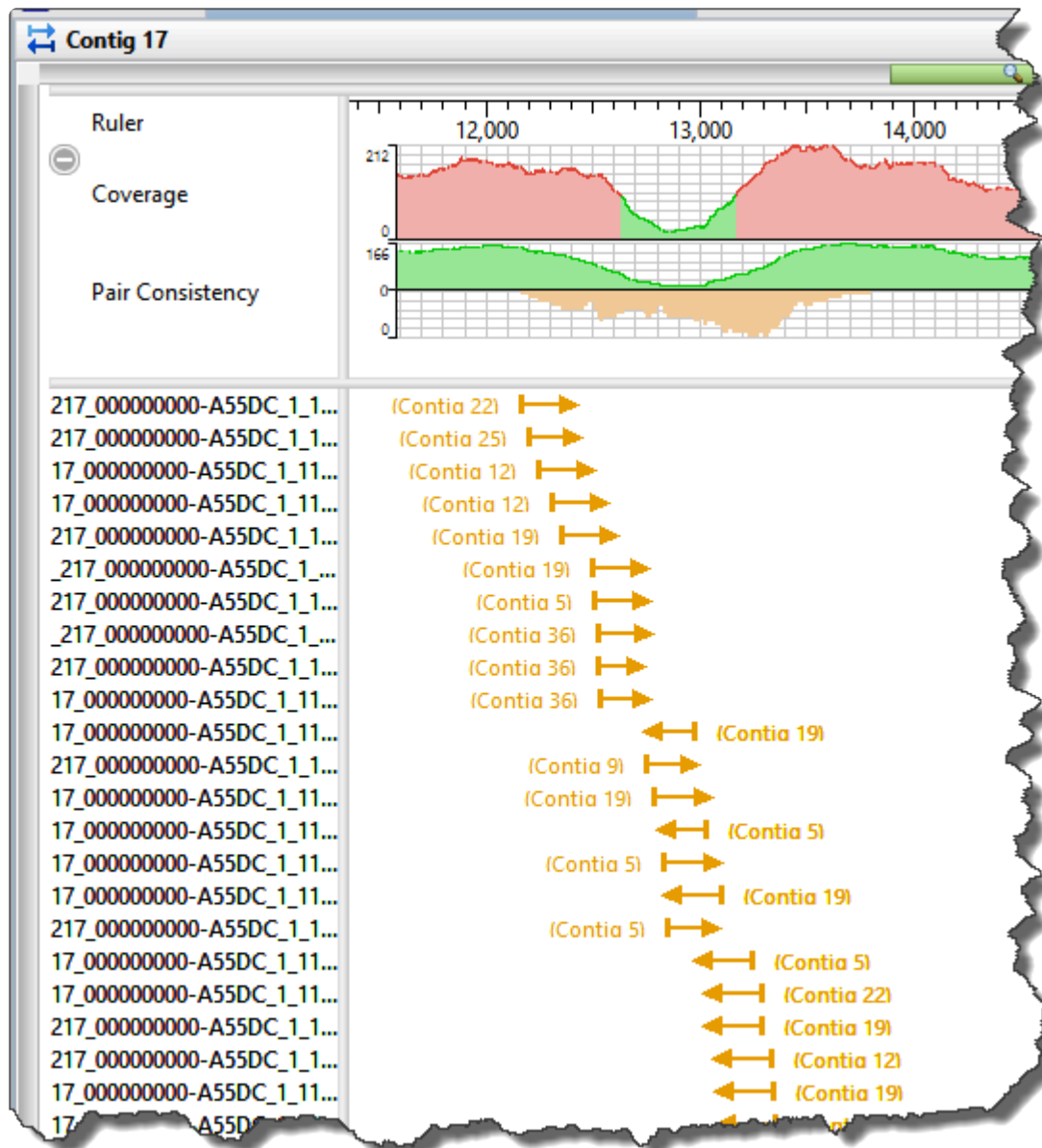
3. The header of this section contains a drop-down menu. Use the menu to select **Project Details**. This section has additional statistics followed by the complete script that was used by SeqMan NGen to assemble the reads. In the upper section of the report, observe that there are an average of 42798 sequences per contig and that that average depth of coverage is 135.

In the next few steps, you will check for potential assembly errors and then correct them. Assume that you have already performed these steps for Contigs 1-16 and have found no errors.

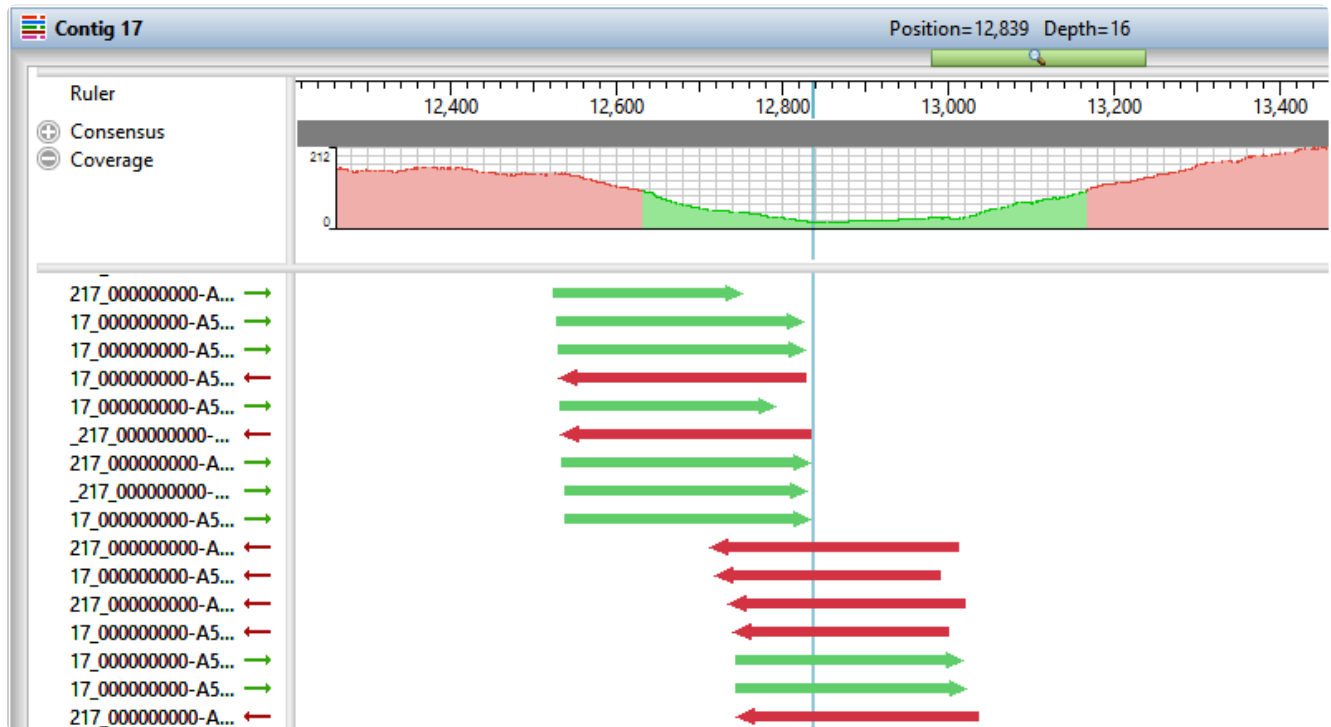
4. In the Explorer panel on the right, select **Contig 17** and press the **Show strategy view** tool () on the right.
5. To limit the display to show only arrows for inconsistent pairs, click the **Style** tab on the right. In the Strategy area, use the **Show** menu to select **Inconsistent**. The orange arrows represent pairs in the

middle of contigs that are not matched up with their mate; the numbers associated with these pairs indicate the contig where their “mate” resides.

- Use the horizontal zoom slider and scroll until you are centered on the area around 13,000. Observe the large number of orange (inconsistent) arrows in this section, as well as orange coloration in the **Pair Consistency** graph in the header. This represents a possible mis-assembled contig. Note that Contigs 5, 19 and 22 are referenced by both left- and right-facing arrows, indicating the region is likely a repeat. This is a good candidate for a contig split.



- Open the Alignment view for this contig by right-clicking in any white portion and choose **Show in Alignment view**. Use the horizontal zoom slider and the horizontal scroll bar to navigate to the same area visible in the Strategy view. Insert the cursor in the low pair-consistency region; the thinnest area of green in the Coverage histogram.



8. Choose **Contig > Split at Insertion** to split the contig into two contigs.
9. Return to the Explorer panel by clicking the **Explorer** tab on the right.
10. Select the uppermost row, **All Contigs** then right-click and choose **Order Contigs into Scaffolds**. When prompted, confirm that you wish to perform scaffolding. SeqMan Ultra uses the pair information at the ends of the contigs to order them into scaffolds.

Fourteen scaffolds are created, numbered from 100-230. The reason that microbial genomes do not assemble into a single contig (when using short reads) is a high prevalence of repetitive elements, like transposons that occur throughout the genome.

11. Now that the contigs have been ordered into scaffolds, you can now merge adjacent overlapping contigs. To do this, select Scaffold 100 and choose **Contig > Align Contigs End to End**. Keep the default settings and press **OK**. Repeat this procedure for all fourteen scaffolds. On occasion, you may get a message saying that the alignment didn't work. In this case, just move on to the next scaffold.
12. Use **Ctrl/Cmd+click** to select each contig that is in a scaffold. Then drag and drop the entire selection anywhere above **Scaffold 100**. Finally, select all of the now-empty scaffolds and press the **Delete** key on your keyboard to remove them.
13. To see the effects of these edits and realignments, use **Project > Project Overview**.

## E.coli K12 MG1655 MiSeq de novo.sqd

C:\Users\yildizs\Desktop\T3\_Whole\_Genome\_DeNovo

Contig N50: 290.0 kb  
Largest contig size: 496,706 bp  
Number of contigs: 36  
Reads assembled: 2,182,744

Note that the **Contig N50** has increased from 204 kb to 290 kb, and that the number of contigs has decreased from 51 to 36.

In a real-world situation, you could continue merging contigs and closing gaps via one or more of the following:

- Continue to make contig edits, create scaffolds and merge contigs to close additional gaps, as above.
- [Perform a BLAST search](#) on contig ends to determine genome coordinates, then manually create new scaffolds and attempt additional end-to-end alignments.
- Resolve remaining gaps by adding new data (e.g., Sanger reads) and using the [gap closure workflow](#).




This marks the end of this tutorial.

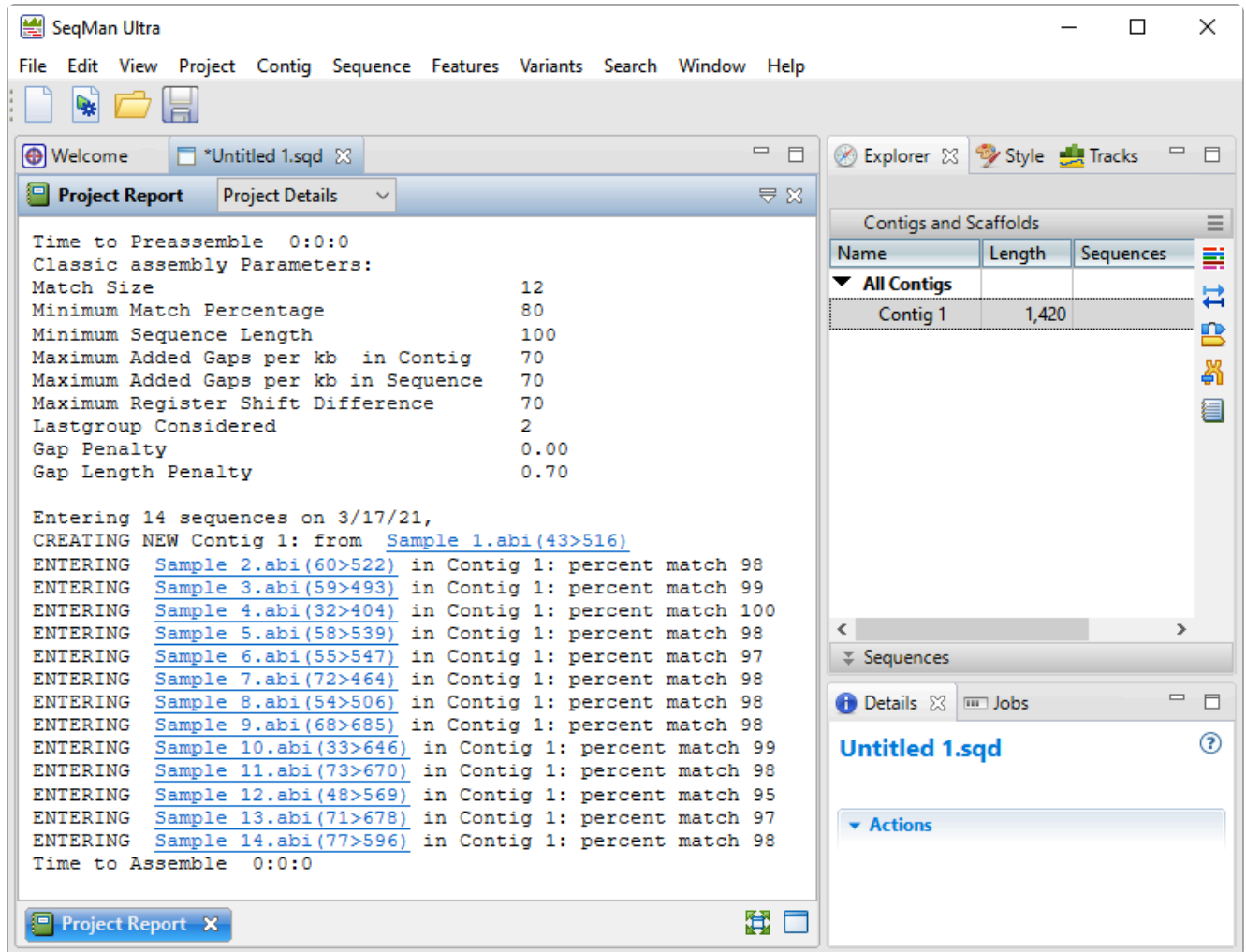
# De novo assembly using Sanger data

---

In this tutorial, you will *de novo* assemble fourteen short trace sequences from PE Applied Biosystems, Inc. and then analyze the resulting contig in SeqMan Ultra. Scroll to the bottom of this topic for a short video showing a slightly different version of this tutorial using the same data set.

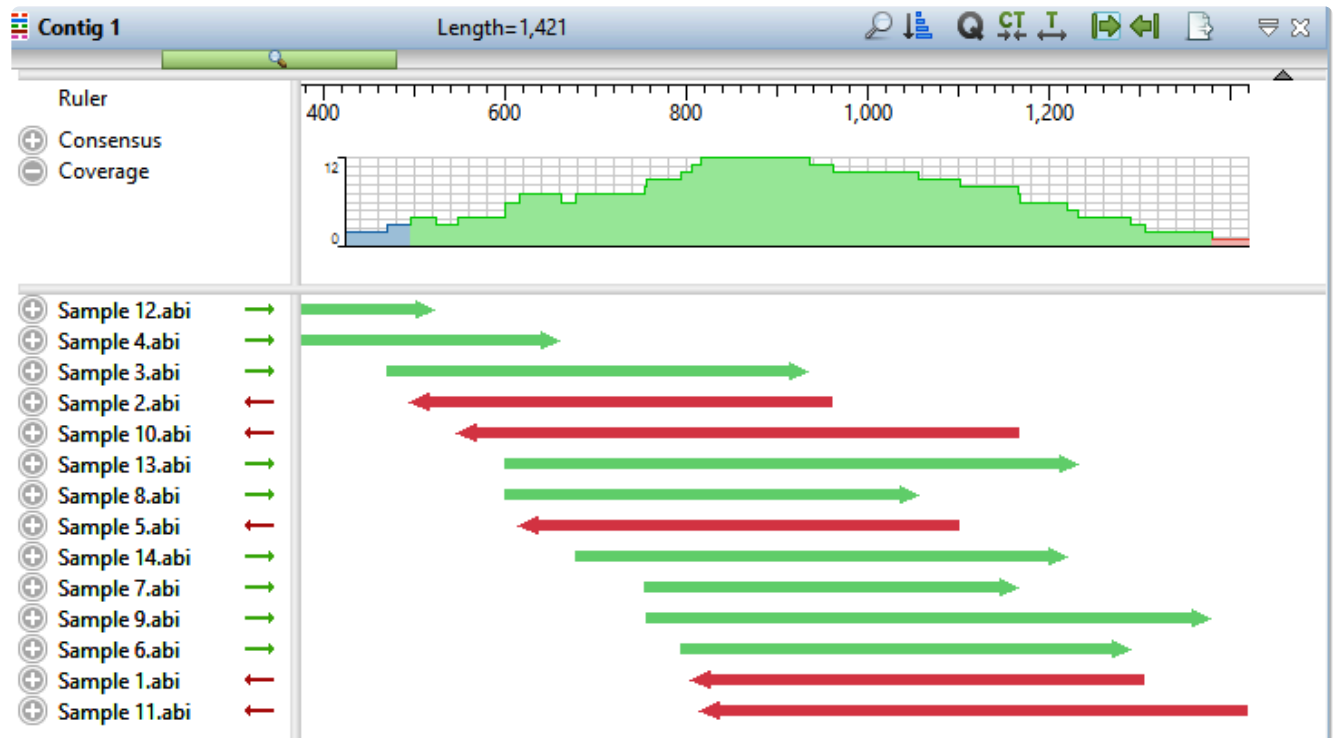
## Assembling Sanger trace reads:

1. Download [T4\\_Sanger\\_DeNovo.zip](#) (1 MB) and extract it to any convenient location (i.e., your desktop). The data set consists of a folder of *Janus vectors* and fourteen *.abi* sample sequences.
2. Launch SeqMan Ultra. On the left of the window, choose the blue bar named **New Assembly**.
3. Under **Molecular biology**, click on **New Sanger/ABI assembly**. The “Unassembled” window opens.
4. Press the **Add** button near the top left ( **Add**), select the 14 *.abi* sequences and press **Open**.
5. In the header area, leave **Quality Trim** checked. This tells SeqMan Ultra to trim read ends based on trace quality evaluation. To remove Janus vector, add a checkmark next to **Vector Trim** and click on the file path just to its right. Navigate to and select the folder *Janus vectors* then press **Select Folder**.
6. Press the **Trim Now** button ( **Trim Now**). In the **Vector** column in the table, observe that the vector *InvJanus* was discovered and removed.
7. Press the **Assemble** button ( **Assemble**) to begin the assembly. The assembly should take a few seconds, at the most.



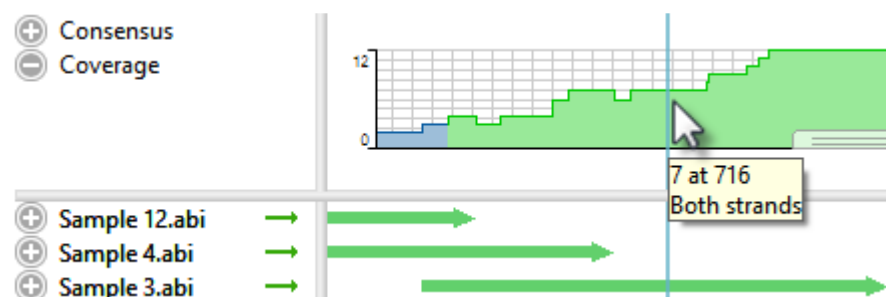
### Examining the alignment:



1. In the Explorer tab on the right, note that the assembly resulted in one contig.
2. Double-click on **Contig 1** to open it in the Alignment view. Move the green horizontal zoom slider (near the top of the view) to the left until you can see the entire contig.



The Coverage graph has areas of blue, green and red. Blue indicates single-direction coverage, while red shows single-read coverage. Green denotes coverage on both strands; the height of the histogram corresponds to the depth of coverage.

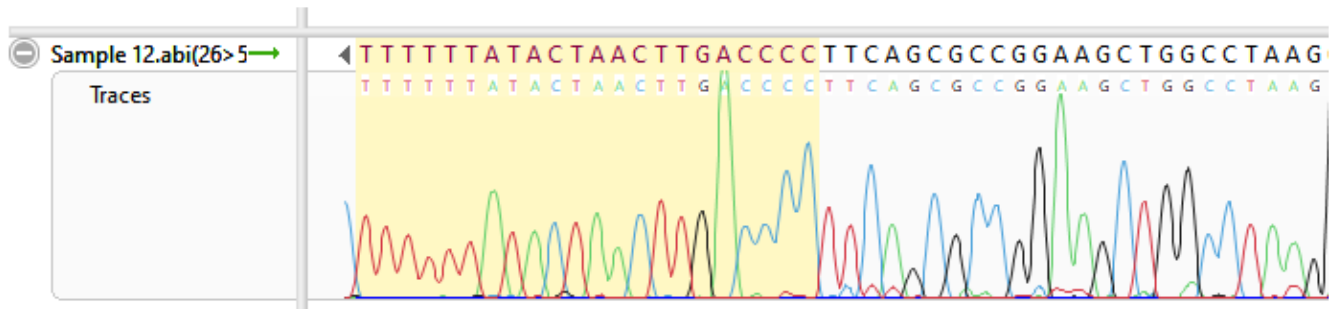
3. Hover the mouse over different parts of this histogram to see tooltips showing the coverage at a given position and whether or not it meets threshold requirements.



4. To zoom in to view details, click the **Restore default zoom** tool (  ) at the top right of the view.
5. To locate conflicts, shown by default with yellow highlighting, click the **Search alignment** tool (  ). Use the **Find** menu to choose **Conflict**, then press the green arrow keys to navigate from one conflict to the next.
6. To view the trace data, right-click on any sample name on the left and choose **Expand All**.
7. Scroll all the way to the left so that only Sample 12.abi is visible.

During assembly, SeqMan NGen trimmed ends for the constituent sequences based on trace data quality and presence of vector. Although sufficient data remained to assemble the sequences into a single contig, there are cases when restoring some of the trimmed data may allow SeqMan Ultra to join multiple contigs into a single contig. You may also wish to restore data in order to verify the consensus in a low-coverage area.

To reveal the trimmed trace data, grab the black triangle to the left of the sequence and drag it to the left. Trimmed sequence appears with a yellow background by default. Conflicts between the restored data and the consensus are shown via red text.



8. To unmask the trimmed regions at both ends of the contig, use **Contig > Extend Trimmed Ends > Extend 3' and 5' Sequence Ends**.

- Trimmed portions of notably poor quality (e.g., misshapen and overlapping peaks) were removed because the average peak quality fell below the acceptable stringency threshold.
- Data removed due to vector contamination is characterized by normal peak quality in combination with a high number of conflicts. In this example, such regions were removed because they originated from the Janus vector. Regardless of its deceptively high peak quality, it is not recommended that you restore vector data.

If you thought the trimmed ends merited being kept in the alignment (they don't in this case), you could have used the command **Contig > Extend Trimmed Ends > Extend and Align 3' and 5' Sequence Ends** to both extend the ends and reassemble the reads.

The following video illustrates the tutorial workflow using the same data set but slightly different steps.

This marks the end of this tutorial.

# RNA-Seq de novo transcriptome workflow

---

In this tutorial, you will *de novo* assemble an abbreviated set of paired end RNA-Seq sequences from *Saccharomyces cerevisiae* (yeast) from [Nookaew I et al., 2012](#). This workflow uses an abbreviated yeast data set with about 1 million reads per file.

With other applications, *de novo* assembly of RNA-Seq data can potentially result in thousands of unlabeled contigs representing the expressed transcripts. By contrast, SeqMan NGen automatically attempts to group contigs from the same gene, and then name and annotate them based on the best match to a collection of annotated reference sequences (the “Transcript Annotation Database”) extracted from data on NCBI’s [RefSeq](#) website. Results from this workflow are non-quantitative.

## Running the transcriptome assembly in SeqMan NGen:

In this part of the tutorial, you will use SeqMan NGen to *de novo* assemble and annotate the RNA-Seq data.

1. Download [T5\\_RNA-Seq\\_DeNovo\\_Transcriptome.zip](#) (147 MB) and extract it to any convenient location (i.e., your desktop). The tutorial data consist of the paired-end reads *Yeast\_RNASeq\_1Mreads\_1.fastq* and *Yeast\_RNASeq\_1Mreads\_2.fastq*.
2. Launch SeqMan Ultra and choose **New Assembly** on the left. On the right, click on the **Transcriptomics** workflow named **De novo transcriptome assembly and annotation**. This causes SeqMan NGen to open at the Workflow screen.
3. Choose the **De Novo Assembly** workflow named **De novo transcriptome**.
4. In the Set Contaminant screen, take the opportunity to verify that you are logged in by looking at the key icon in the bottom left corner. If there is a green check mark, click **Next**. If there is a yellow triangle, click the icon and enter the same login credentials you use for the DNASTAR website. Once you return to the Set Contaminant screen, click **Next**.
5. In the Input Sequences screen, press **Add** and add the *Yeast\_RNASeq\_1Mreads\_1.fastq* and *Yeast\_RNASeq\_1Mreads\_2.fastq* files. Click **Next**.
6. In the Transcript Annotation Database screen, click the **Download Database** button. Choose **RefSeq Fungi** and press **Select**. Then click **Next**.

Reference Annotation Databases		
Reference Annotation Database ▲	Ref-Seq Version	Download Size
Complete RefSeq	94.0	14.4 GB
RefSeq Archaea	94.0	353.4 kB
RefSeq Bacteria	94.0	7.1 MB
RefSeq Fungi	94.0	1.3 GB
RefSeq Invertebrate	94.0	2.4 GB
RefSeq Mitochondrion	94.0	9.4 kB
RefSeq Plant	94.0	2.4 GB
RefSeq Plasmid	94.0	4.7 kB
RefSeq Plastid	94.0	9.3 kB
RefSeq Protozoa	94.0	473.8 MB
RefSeq Vertebrate (Mammalian)	94.0	3.8 GB
RefSeq Vertebrate (Other)	94.0	3.9 GB
RefSeq Viral	94.0	93.2 MB

7. In the Assembly Options screen, click **Next**.
8. In the Assembly Output screen, type “Transcriptome” into the **Project Name** text box, then use the **Browse** button to specify a **Project Folder** for your assembly output files. Click **Next**.
9. In the Run Assembly Project screen, note that:
  - The estimated disk requirement of 2.1 TB is based on the total length of the fungal Transcript Annotation Database, which is 4.2 GB: larger than a human genome. That estimate is based on reference guided genome assemblies that have fixed 50X coverage, not reference guided transcriptome assemblies, which have highly variable coverage. The assembly in this tutorial has extremely low coverage and uses far less disk space than what is estimated here.
  - Cloud Assembly is not offered for the *de novo* transcriptome workflow because most data sets exceed the 48 hour time limit.

Click the link “Run assembly on this computer.” The assembly will take approximately one hour on a standard laptop.

10. Wait until being informed that assembly has finished, then click **Next**.

```
;xng 'C:\Users\Public\Documents\
Project.Transcriptome\dbSearch-
```

Assembly process has finished

11. In the Assembly Summary screen, note the button **View assembled transcripts**. In the future, this button will allow you to open transcriptome results in SeqMan Ultra. As of version 17, however, *.transcriptome* results can only be analyzed in SeqMan Pro.
12. Click **Finish** to close SeqMan NGen and press **Yes** when prompted.

### Viewing transcripts in SeqMan Pro:

During the assembly process, the *de novo* transcriptome assembly output was saved to a package called *Transcript Project.Transcriptome*. Any assembled transcripts with a database match exceeding the specified thresholds were termed “Identified Transcripts,” while assembled transcripts that did not have a database match were called “Novel Transcripts.” This part of the tutorial shows how to load the annotated transcripts into the SeqMan Pro application for downstream analysis.

1. Launch SeqMan Pro and drag and drop the result file *Transcriptome.Transcriptome* from your file explorer onto the SeqMan Pro window.
2. Observe that the ensuing All Transcripts window contains two tabs. Each tab’s heading shows the total number of transcripts in the table, and the number currently selected. The tables in the two tabs support a wide variety of sortable columns which can be displayed or hidden, as desired.

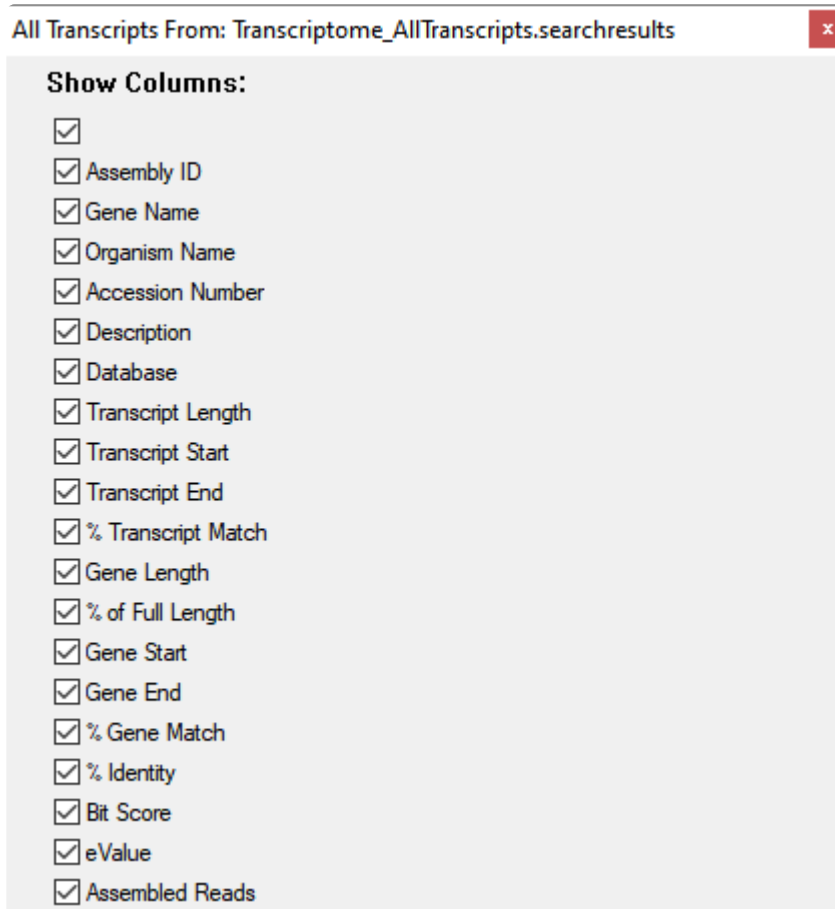
The **Identified Transcripts** tab is active, by default. You should see over 1300 **Total Identified Transcripts**. Since you haven’t yet made any selections, the number of **Selected Identified Transcripts** is zero.

Identified Transcripts		Novel Transcripts			
Total Identified Transcripts:		1332			
Selected Identified Transcripts:		0			
Assembly ID	Gene Name	Organism Name	Accession Number		
AAH1_NM_001182979.1_co_1167_1	AAH1	Saccharomyces cerevisiae	NM_001182979.1	S288c	
AAT2_NM_001181914.1_co_891_1	AAT2	Saccharomyces cerevisiae	NM_001181914.1	S288c	
ABF2_NM_001182570.1_co_1056_1	ABF2	Saccharomyces cerevisiae	NM_001182570.1	S288c	
ABP140_NM_001183658.6_co_1312_1	ABP140	Saccharomyces cerevisiae	NM_001183658.6	S288c	
ACB1_NM_001181166.3_co_694_1	ACB1	Saccharomyces cerevisiae	NM_001181166.3	S288c	
ACC1_NM_001183193.1_co_1211_1	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288c	
ACC1_NM_001183193.1_co_1211_4	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288c	
ACO1_NM_001182192.1_co_964_1	ACO1	Saccharomyces cerevisiae	NM_001182192.1	S288c	

3. Click on the **Novel Transcripts** tab. You should see approximately 50-70 **Total Novel Transcripts**. This table lists the assembled contigs that did not have any match to the Transcript Annotation Database that met the search criteria thresholds and therefore, were not labeled with any match information. Note that this table contains only three columns.

Identified Transcripts		
Novel Transcripts		
<b>Total Novel Transcripts: 56</b> <b>Selected Novel Transcripts: 0</b>		
Assembly ID	Transcript Length	Assembled Reads
CRH1_NM_001181318.1_co_735_1	1626	627
PPN1_NM_001180760.3_co_614_2	1121	260
RPS22A_NM_001181623.1_co_821_2	534	562
cl_1168_2	262	350
cl_1982_1	963	724
cl_2307_1	311	906
cl_2855_1	354	192
cl_3239_4	537	311
cl_3300_1	453	109
cl_3650_2	326	180
cl_4563_1	385	103

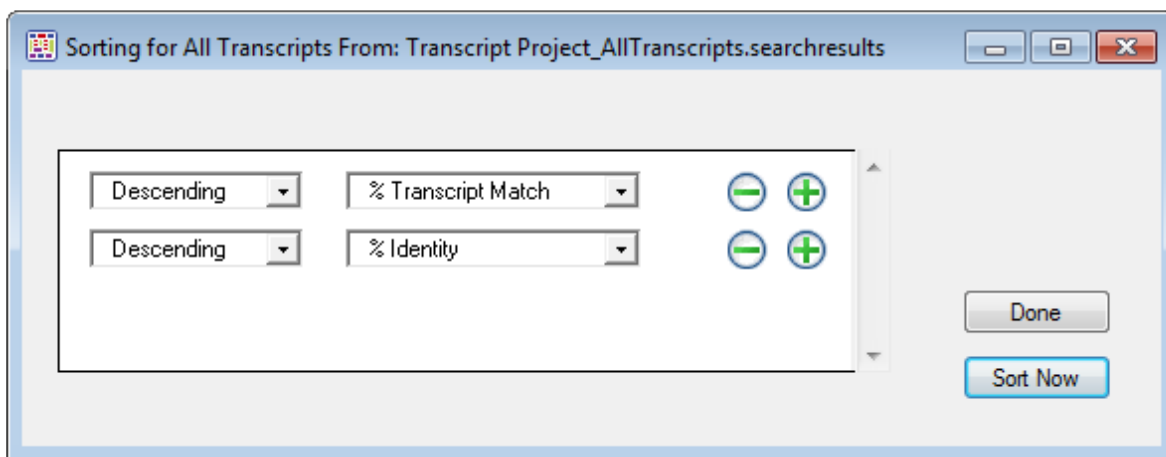
4. Return to the **Identified Transcripts** tab and experiment with the following:
- To show or hide columns - Right-click and choose **Show/Hide Column**, then check or uncheck boxes. Each column is [described in detail](#) in the SeqMan Pro help.



- To move a column - Use the mouse to drag and drop it in the desired location.
- To sort data in alphabetical or numerical order - Click on the column header that you wish to sort. Note that the resulting groups are also shown in different colors to help visually differentiate between them.

Identified Transcripts		Novel Transcripts			
Total Identified Transcripts:		1322			
Selected Identified Transcripts:		0			
Gene Name	Description	% Transcript Match	% Gene Match	% Identity	
TOS1	S288c Tos1p	78.17 %	100.00 %	100.00 %	
RIB5	S288c riboflavin synthase	92.80 %	70.15 %	100.00 %	
ARO4	S288c 3-deoxy-7-phosphoheptulonate synthase ARO4	87.91 %	100.00 %	100.00 %	
TIM17	S288c protein transporter TIM17	65.98 %	100.00 %	100.00 %	
	S288c hypothetical protein partial mRNA	99.67 %	46.57 %	100.00 %	
APE3	S288c Ape3p	96.30 %	100.00 %	100.00 %	
HYR1	S288c peroxiredoxin HYR1	88.97 %	100.00 %	100.00 %	
YSY6	S288c Ysy6p	67.35 %	100.00 %	100.00 %	
RPL2B	S288c ribosomal 60S subunit protein L2B	63.38 %	23.53 %	100.00 %	
RPA34	S288c DNA-directed RNA polymerase I subunit RPA34	93.34 %	97.86 %	100.00 %	
UMP1	S288c Ump1p	76.80 %	100.00 %	100.00 %	
HIS4	S288c trifunctional histidinol d...sphoribosyl-ATP diphosphatase	92.95 %	100.00 %	100.00 %	
RPS21B	S288c ribosomal 40S subunit protein S21B	55.23 %	100.00 %	100.00 %	
GBP2	S288c Gbp2p	100.00 %	83.64 %	100.00 %	
PGI1	S288c glucose-6-phosphate isomerase	89.56 %	100.00 %	100.00 %	
GDT1	S288c putative ribosome biosynthesis protein GDT1	86.37 %	100.00 %	100.00 %	
RRP7	S288c Rrp7p	92.62 %	47.76 %	100.00 %	
SNL1	S288c Snl1p	97.74 %	90.00 %	100.00 %	
TIF4631	S288c Tif4631p	87.08 %	81.60 %	99.96 %	
CDC48	S288c AAA family ATPase CDC48	95.29 %	100.00 %	99.96 %	
MNN1	S288c alpha-1,3-mannosyltransferase MNN1	90.76 %	100.00 %	99.96 %	
HOM3	S288c aspartate kinase	87.36 %	100.00 %	99.94 %	
	S288c RDN37-2 miscRNA	97.60 %	26.16 %	99.94 %	
ADE17	S288c bifunctional phosphorib...rase/IMP cyclohydrolase ADE17	89.48 %	100.00 %	99.94 %	
PAB1	S288c polyadenylate-binding protein	85.29 %	100.00 %	99.94 %	
SSZ1	S288c Ssz1p	94.12 %	100.00 %	99.94 %	
ATR1	S288c Atr1p	91.31 %	100.00 %	99.94 %	
MET2	S288c homoserine O-acetyltransferase	86.39 %	100.00 %	99.93 %	
COR1	S288c ubiquinol-cytochrome-c reductase subunit COR1	74.86 %	100.00 %	99.93 %	
GAS5	S288c 1,3-beta-glucanosyltransferase	92.03 %	100.00 %	99.93 %	
PFK2	S288c 6-phosphofructokinase subunit beta	92.75 %	100.00 %	99.93 %	
TIM50	S288c protein translocase subunit TIM50	92.81 %	98.32 %	99.93 %	

- To open individual contigs in SeqMan Pro for visualization and editing, double-click on a row of interest to navigate to the corresponding contig assembly. The appropriate .sqd file is loaded with the Alignment view of the selected contig displayed. All the usual visualization and editing tools in SeqMan Pro are available.
- To set stringency thresholds, return to the Identified Transcripts tab and choose **View > Sort**. Set up the dialog to match the image below. To add the second row, click on the plus sign in the original row. Press **Sort Now**.




- To save the sub-set of transcripts that met the stringency thresholds:
  1. In the Identified Transcripts table, click the **%Transcript Match** header to sort the column in decreasing order.
  2. Select all rows where **Transcript Match** is  $\geq 99.00$ , noting that the **Identity** is also  $\geq 99$  for those rows. In the header, you will see approximately 130-160 selected transcripts.
  3. Choose **File > Save Selected Transcripts for**. In the Save dialog, designate a name and location for the output file, and then click **Save**. Two files will be created, each with a different extension (*.fas* and *.searchresults*).
  4. (optional) To see what is contained in each of these file types, open the files in any suitable text editor.

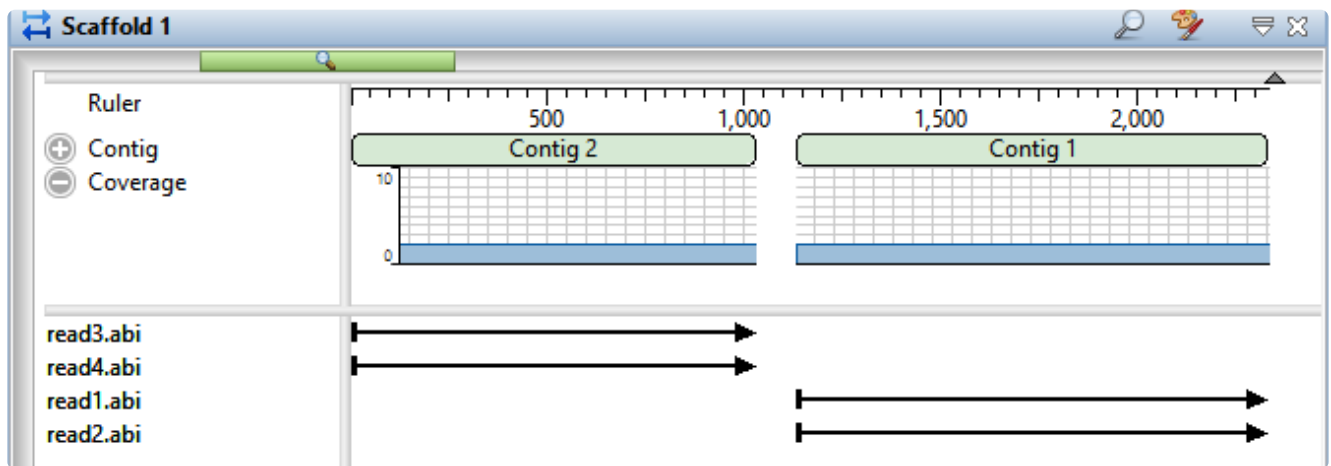
This marks the end of this tutorial.

# Gap closure workflow

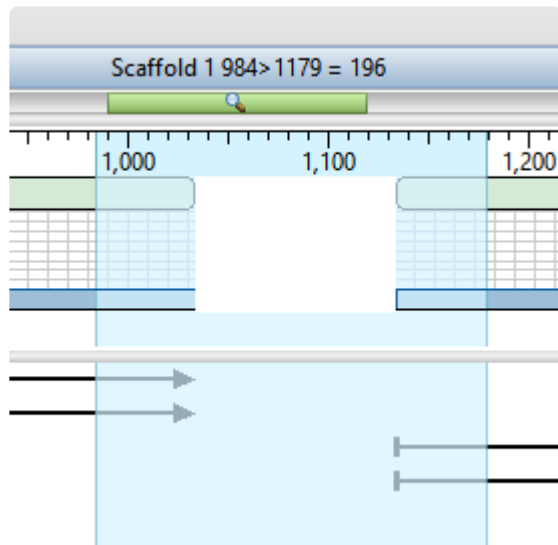
In this tutorial, you will use an *.abi* read file to close a gap between two contigs.

1. Download [T6\\_Gap\\_Closure.zip](#) (45 kb) and extract it to any convenient location (i.e., your desktop). The data set consists of one very small SeqMan Ultra project and one *.abi* file.
2. Launch SeqMan Ultra and use **File > Open** to open the file *De novo assembly.sqd*. This assembly contains two unscaffolded contigs
3. From the Explorer panel, use **Shift+click** to elect both contigs. Then right-click on the selection and choose **New Scaffold**.
4. Select the scaffold ("Scaffold 1") and press the **Show Strategy view** tool () to the right of the selection.

In the Strategy view, note that each contig is comprised of two reads and that there is a clear gap between the two contigs.



5. Drag the cursor across the ruler area to select a range that spans the gap between the two contigs. There is no need to make a precise selection.



6. Right-click on the selection and choose **Add Sequences to Close Gap**.
7. Click the **Add** button and add the sequence *Gap filling read.abi*. Note that there are changeable **Alignment options**, but leave them at the defaults for purposes of this tutorial.

Name	Limits	Type
Gap filling read.abi		Trace

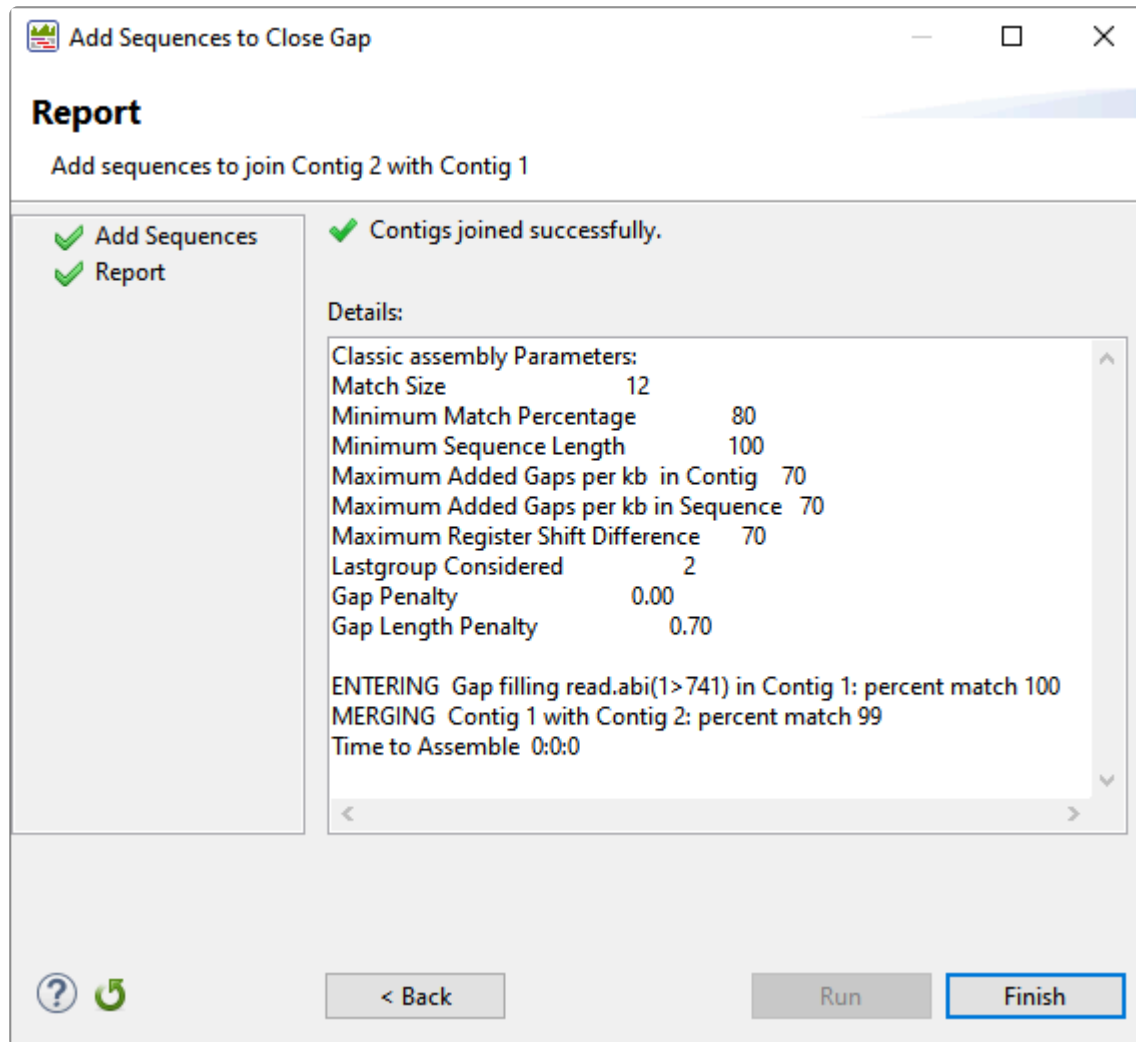
Alignment options:

Minimum match percentage: 80.0 %

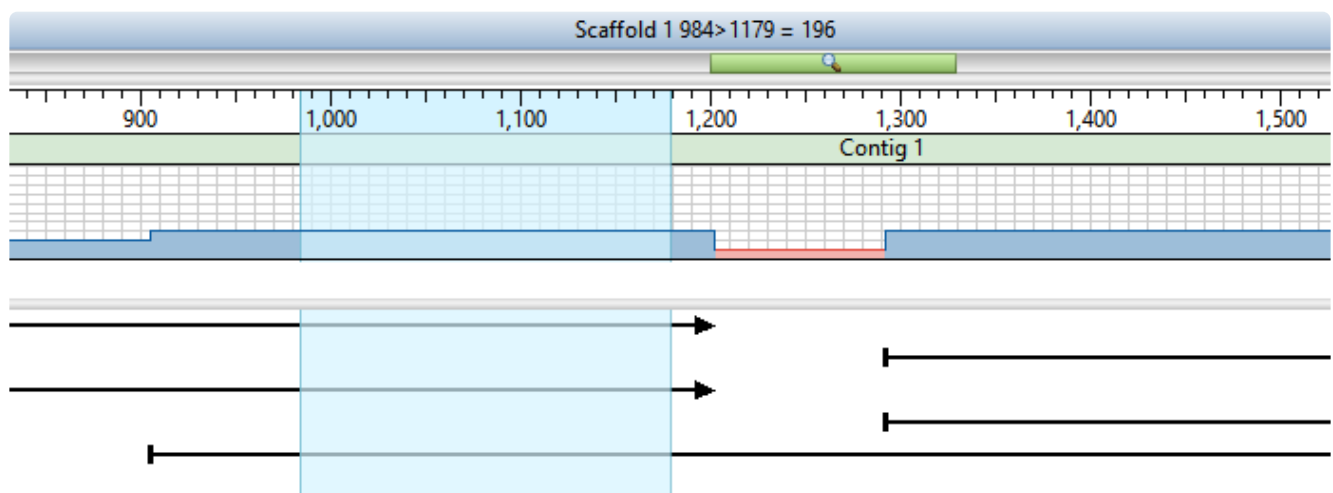
Match size: 12 bp

Gap penalty: 0.00

8. Click **Run**. At the bottom of the report, observe the text: **MERGING Contig 1 with Contig 2: percent match 99 (or 100)**.



9. Click **Finish**
10. Return to the Strategy view and observe that the gap is now closed. In the image below, note that the selection is the same as in Step 5.



This marks the end of this tutorial.

# Reference-guided alignment of Sanger/ABI data

---

The following video illustrates how to align (assemble) Sanger/ABI data to a reference sequence. Data files are not provided for this tutorial.

# Assembly Projects

---

Does the SeqMan Ultra window appear “blank”? Wondering how to get started? Or perhaps you have a project open, but want to open a different project.

To learn how to create, open or navigate between multiple open assemblies, see these topics:

- [Create a new assembly](#)
- [Open an existing assembly](#) and then [get started](#) with analysis
- [Navigate between open assemblies](#)

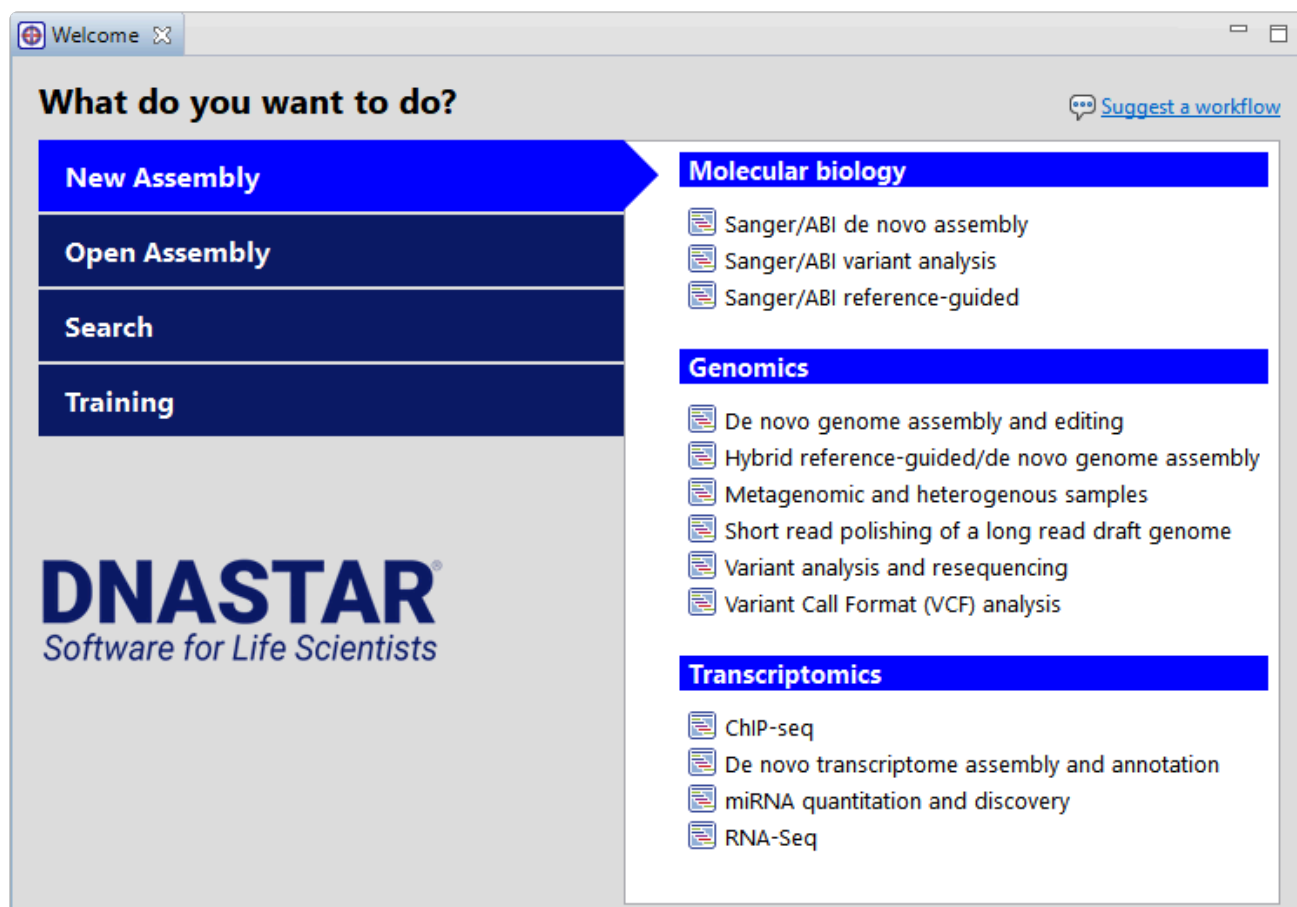
# Create a new assembly

The Welcome window opens when you launch SeqMan Ultra, and is used for creating or opening assemblies. As you add projects, the Welcome window will remain as the leftmost project tab unless you manually move or close it.

To set up a new assembly:

Beginning in Lasergene 17 (early 2020), all assemblies are performed within SeqMan NGen. SeqMan Ultra's Welcome window can be used as a launcher for a SeqMan NGen assembly workflow, as described below. You can also initiate an assembly through SeqMan NGen directly.

1. If the Welcome tab is not visible, choose **View > Welcome**.
2. Click on the Welcome tab and choose **New Assembly** on the left.



3. Click on the assembly type of interest. This launches the SeqMan NGen wizard at the [Workflow screen](#). The screen is pre-populated with options specific to that workflow.
4. Choose the workflow of interest on the right.

5. Proceed through the SeqMan NGen wizard, uploading data files and selecting the desired options and settings. When prompted, specify and name and location in which to save the assembly files.
6. Once the assembly is ready to run, SeqMan NGen will recommend you click one of two links, depending on the size of the project. One link will run the assembly on your local computer. The other link will run the assembly on the cloud. Cloud assemblies can be purchased from our [Academic](#) / [Commercial](#) pricing pages.
7. As soon as the assembly finishes, you may proceed directly to downstream analysis by pushing one of the provided buttons. The buttons, which depend on the workflow you followed, lead to ArrayStar, SeqMan Ultra and/or GenVision Pro.

See the [SeqMan NGen User Guide](#) for complete information on setting up and performing the assembly using the SeqMan NGen wizard.

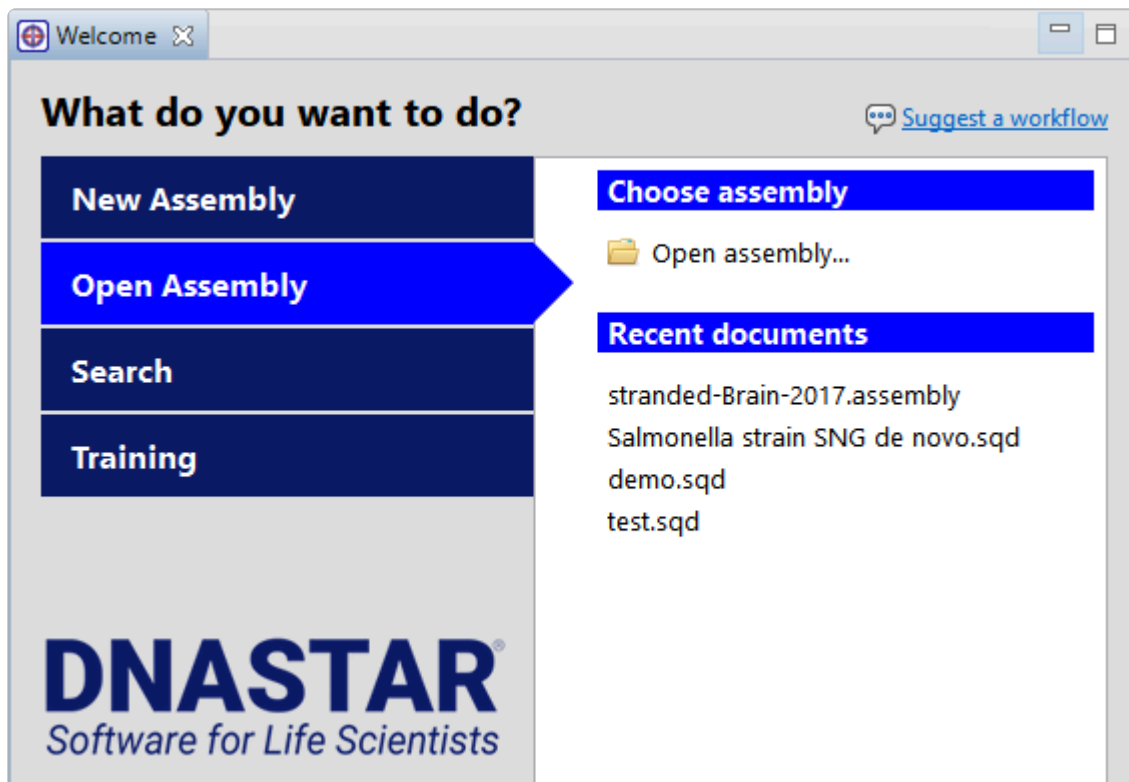
# Open an existing assembly

SeqMan Ultra lets you review and, in some cases, edit assemblies made with SeqMan NGen. Functionality varies depending upon the file extension of the assembly. SeqMan NGen assemblies saved in SeqMan Ultra format (.sqd) can be reviewed and edited. SeqMan NGen assemblies saved in BAM format (.assembly) can be reviewed, but not edited.

To open an existing SeqMan NGen assembly in SeqMan Ultra:

To open a .assembly, .sqd or Transcriptome project in SeqMan Ultra, do any of the following:

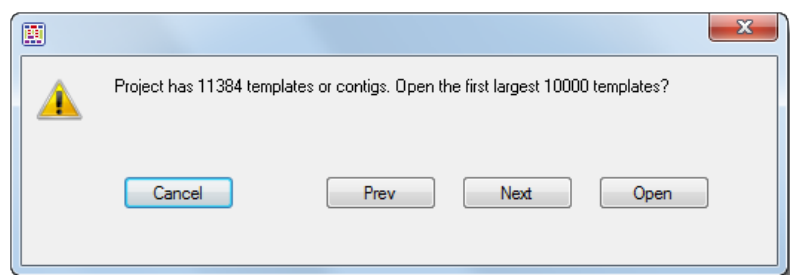
- Click on the **Welcome** tab. If the Welcome tab is not visible, choose **View > Welcome**. Then choose **Open Assembly** on the left. On the right side of the screen, look for the document in the **Recent documents** area. If it is there, click on its name. Otherwise, click on **Open assembly** on the right.



- With any window active, use the **File > Open** command.
- Drag & drop an assembly file/package onto the SeqMan Ultra window.
- Double-click on an .sqd or .ace file. Macintosh also supports double-clicking on an .assembly package.

If the .assembly contains more than 10,000 templates or contigs (most commonly seen in *de novo*

Transcriptome/RNA-Seq assemblies), you will receive the following warning: **Project has <number over 10,000> templates or contigs. Open the first largest 10,000 templates?**



There are three options for handling this situation:

- Choose **Open** to open only the 10,000 longest contigs in the assembly. Or use the **Next/Previous** buttons to navigate to the group of “second longest” contigs, and so on. When you find the group of interest, click **Open**.
- Open the *.assembly* in ArrayStar, locate the contig of interest and send it to SeqMan Ultra via ArrayStar's **File > Send to SeqMan Ultra** command.
- Depending on the size of your project, you may be able to force SeqMan Ultra to open all contigs. When you receive the warning message, hold down the **Alt** (Win) or **Option** (Mac) key and press **Open**. SeqMan Ultra will attempt to load all of the contigs into the project.

# Getting Started

---

Have you opened an assembly, but have no idea what to do next? You've come to the right place!

See the following topics to learn how to:

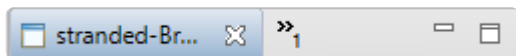
- [View information about the currently open assembly](#)
- [Work with contigs](#) (including descriptions of most of the views and panels), [data tracks](#), [features](#) or [variants](#)
- [Make a selection](#) and [view information about it](#)
- [Search for sequences online](#) using BLAST or Entrez
- [Customize appearance and behavior](#)
- [Export consensus sequences or tabular data](#)

# Navigate between open assemblies

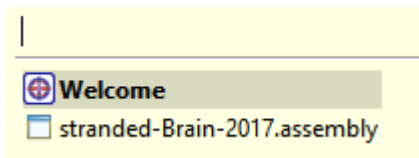
One instance of SeqMan Ultra may include multiple open projects. Each window is displayed in a separate tab.

One tab, the Welcome screen, is always present by default unless it has been manually closed. To make another project active, click on its tab.

If you cannot see all the window tabs in a project, press the >> icon that appears to the right of any visible window tab(s). This icon is displayed when multiple view windows are open and the main SeqMan Ultra window is too small to display all the tabs.



The window chooser appears as a yellow popup with a line for typing in text at the top, and a list of open projects below.



Use your mouse to click on the window you want to make active. If you do not have a mouse, make a selection using the keyboard arrow keys, then press **Enter**. To instead search for a project using a filter, type text into the text field and press **Enter**. The text field supports the “?” and “\*” wildcards.

# View information about the currently open assembly

---

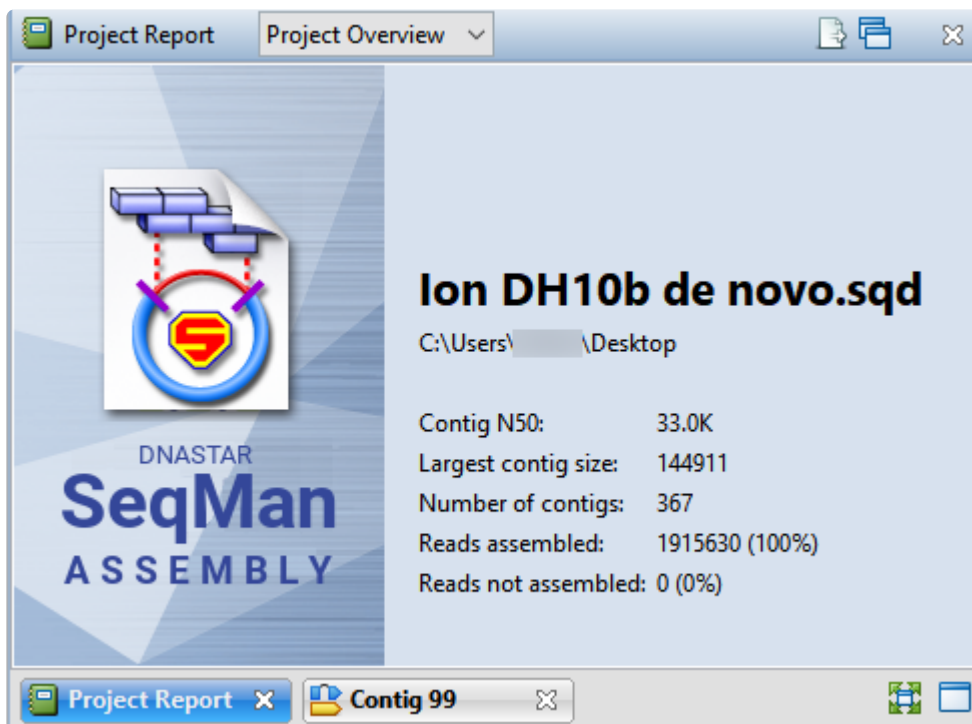
Information about the currently open assembly as a whole is shown in the Project Report view. This view comes in three versions:

- [Project Overview](#)
- [Project Details](#)
- [Trimming](#)

To access a version of the view, use any of the **View > Project Report** sub-commands. Once the report is open, you can choose another version of it using the drop-down menu on the left of the header. Note that different project types may allow only a subset of view types or may contain one or more empty Project Report views.

# Project Overview


Project Overview is one of three Project Report views that may be available, depending on the assembly type that is currently open. For information on accessing or performing tasks in this view, see [View information about the currently open assembly](#).



The available parameters differ depending on the type of assembly project.

Parameter	Description
<b>.sqd projects</b>	
<b>Contig N50</b>	N50 is the sequence length of the shortest contig in a group of contigs at 50% of the total genome length. In other words, N50 is the point of half of the mass of the distribution. The number of bases from all contigs longer than the N50 will be close to the number of bases from all contigs shorter than the N50.
<b>Largest contig size</b>	The length of the longest contig in bp.
<b>Number of contigs</b>	The number of contigs in the finished assembly.
<b>Reads assembled</b>	The number of reads that were used in the finished assembly.
<b>Reads not assembled</b>	The number of reads that were not included in the finished assembly.
<b>.assembly projects</b>	

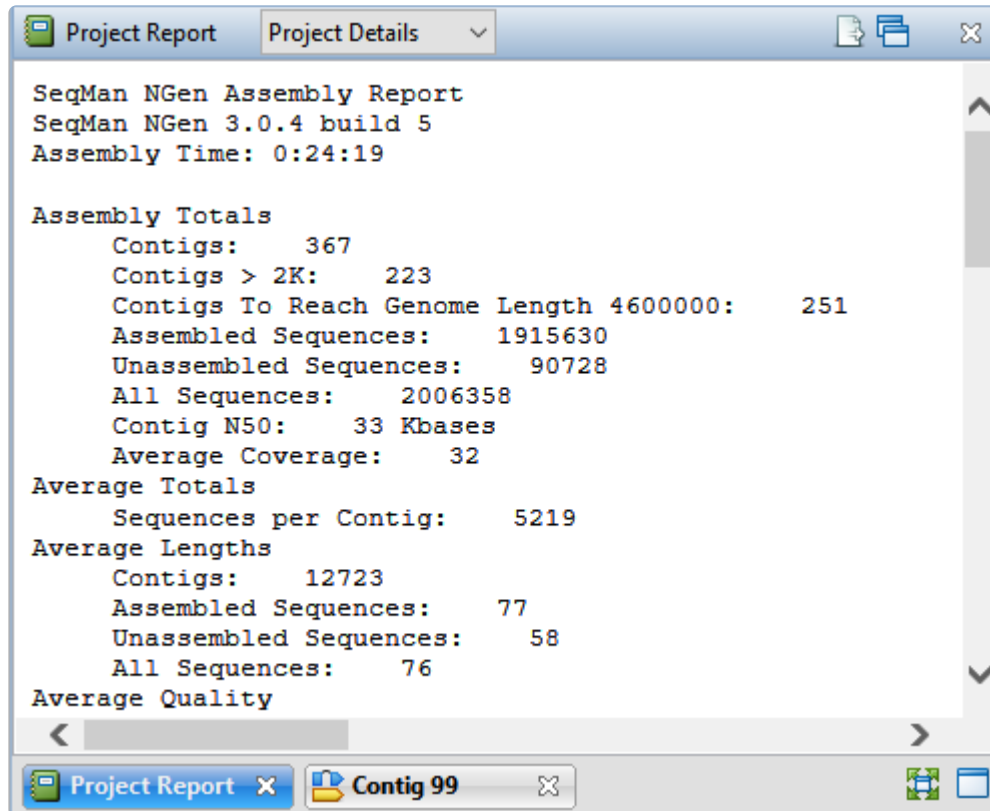
<b>Average coverage</b>	Average depth of coverage along the assembly.
<b>Number of references</b>	Number of reference sequences used.
<b>Aligned sequences</b>	Number and percentage of sequences used in the finished alignment.
<b>Number of variants</b>	Number of variants found.
<b>Variant assemblies</b>	For multi-sample assemblies only. The number of assemblies in the project.

To export data from this view, press the **Export data** tool (  ) in the top right of the view. This opens a save dialog in which you can save the tabular data in comma-separated (.csv) or tab-separated (.tab) format.

# Project Details


The Project Details tab summarizes assembly time, the parameters used in constructing a contig, and each stage in the progress of an assembly, as well as the median pair distance and standard deviation for each set of a given sized pair library in an assembly. Project Details is one of three [Project Report views](#) that may be available, depending on the assembly type that is currently open.

The image below shows a portion of one version of the Project Details view.



The table below describes some of the parameters with the most general interest.

Parameter	Description
<b>Total Reads Assembled</b>	Contains information about reads that made it into the assembly.
<b>Unassembled Sequences</b>	Lists reasons why certain reads were not included.
<b>Coalesced 'n' identical reads</b>	Shows the number of reads that were treated as duplicates.
<b>Found SNP Cnt (incl. indel lengths)</b>	Shows the number of SNPs that were found.
<b>Assembly Parameters</b>	Shows the SNP-calling parameters specified in the SeqMan NGen wizard for this assembly.
<b>SCRIPT</b>	Shows the entire assembly script used in SeqMan NGen when creating this assembly.

To export data from this view, press the **Export data** tool (  ) in the top right of the view. This opens a save dialog in which you can save the tabular data in comma-separated (.csv) or tab-separated (.tab) format.

# Trimming


Trimming is one of three [Project Report](#) views and is only available for de novo assemblies of Sanger trace data. When present, the report provides information about quality trimming and vector removal done as part of the SeqMan NGen assembly.

Project Report Trimming					
SCAN FOR VECTOR					
Janus					
TRIM ENDS					
MEDIUM TRIM PARAMETERS					
Trace: Threshold = 12					
Qual: Threshold = 12					
Non-Trace: Window Size = 50					
Maximum Ns = 2					
NAME	AVERAGE QUALITY	AMOUNT TRIMMED	TRIMMED LENGTH	PRE-TRIM LENGTH	
"Sample 14.abi" (68>615)	26	275	548	823	
"Sample 1.abi" (43>538)	33	245	496	741	
"Sample 2.abi" (53>523)	28	389	471	860	
"Sample 3.abi" (56>519)	32	318	464	782	
"Sample 4.abi" (32>405)	34	443	374	817	
"Sample 5.abi" (31>540)	30	283	510	793	
"Sample 6.abi" (54>548)	28	337	495	832	
"Sample 7.abi" (52>483)	26	375	432	807	
"Sample 8.abi" (45>507)	30	389	463	852	
"Sample 9.abi" (58>686)	26	141	629	770	
"Sample 10.abi" (33>667)	29	67	635	702	
"Sample 11.abi" (63>690)	28	209	628	837	
"Sample 12.abi" (48>570)	30	269	523	792	
"Sample 13.abi" (55>694)	26	168	640	808	
14 Sequences	Average	29	279	522	801

The Trimming report lists the name of each sequence, its average quality, amount trimmed, and the trimmed and pre-trimmed lengths.

Data Type	Description
Trace Threshold	
Qual Threshold	
Non-Trace Window Size and Max Ns	SeqMan NGen trims data from sequence ends based on the number of ambiguous residues (Ns) within each sliding window. If the first number was 50 and the second number was 2, for example, this would show that SeqMan NGen trimmed sequence ends to leave two or fewer ambiguous bases (Ns) in the first and last 50 bases in the sequence.

<b>Name</b>	Name of the read file followed by parentheses showing the position of the sequence on the consensus.
<b>Average Quality</b>	The average quality of bases that remained after trimming.
<b>Amount Trimmed</b>	Number of bases removed from the read file, in bp.

To export data from this view, press the **Export data** tool (  ) in the top right of the view. This opens a save dialog in which you can save the tabular data in comma-separated (.csv) or tab-separated (.tab) format.


# Contigs

---

The following topics contain information on viewing, editing and searching within contigs.





- [View contig information in the Explorer panel](#)
- [Display contig information in a graphical or tabular view](#)
- [Search within a contig](#)
- [Edit sequences](#). This topic applies only to *.sqd* projects

# View and select contigs in the Explorer panel

Basic contig information can be viewed in the Explorer panel, represented by a compass icon (  ). This panel includes a table showing all of the contigs, their lengths, constituent sequences and positions. Below these are the names of sequences that were not placed in any contigs. If scaffolds exist, they are also included in the table.

To open the Explorer panel, do any of the following:

- Click on the **Explorer** tab
- Choose **View > Explorer**
- Press **Ctrl+Alt+E** (Windows) or **Option+Cmd+E** (Macintosh)

Explorer    			
Name	Length	Sequences	Position▼
▼ All Contigs			0
NC_000001	248,961...	2,331,301	0
NC_000002	242,197...	1,487,731	0
NC_000003	198,298...	1,221,578	0
NC_000004	190,216...	739,050	0
NC_000005	181,541...	1,056,056	0
NC_000006	170,808...	1,018,661	0
NC_000007	159,348...	1,074,285	0
NC_000008	145,140...	798,841	0
NC_000009	138,396...	981,230	0
NC_000010	133,799...	838,837	0

The table's data columns are described in the table below.

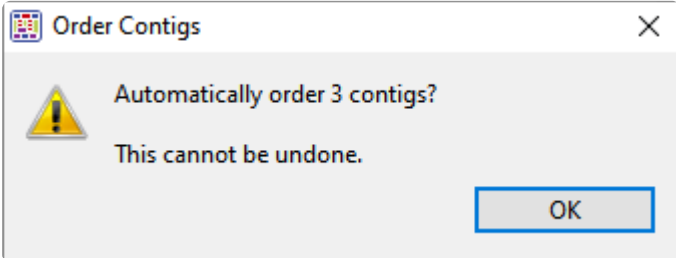
- To sort contigs, click once or twice on any of the column headers to sort by that value in ascending/descending order. If scaffolds are present, clicking on any column header will first sort by the position of the scaffold, and then sort the contigs within each scaffold. |
- To edit the **Name** or **Position** fields, hold the mouse cursor down on the field for one second, then release. Then overwrite or append to the existing name.

Column header	Description
<b>Name</b>	The name of the contig. Some contigs may have coordinates displayed in parentheses next to the name. These are always based on the original, ungapped reference sequence. The purpose of the coordinates is to show the relative positioning of the contig with respect to the original template. When two contigs are joined, the new contig retains the name and original coordinates of the 5' contig. When a contig is split using the <b>Contig &gt; Split at Insertion</b> command, the new 5'

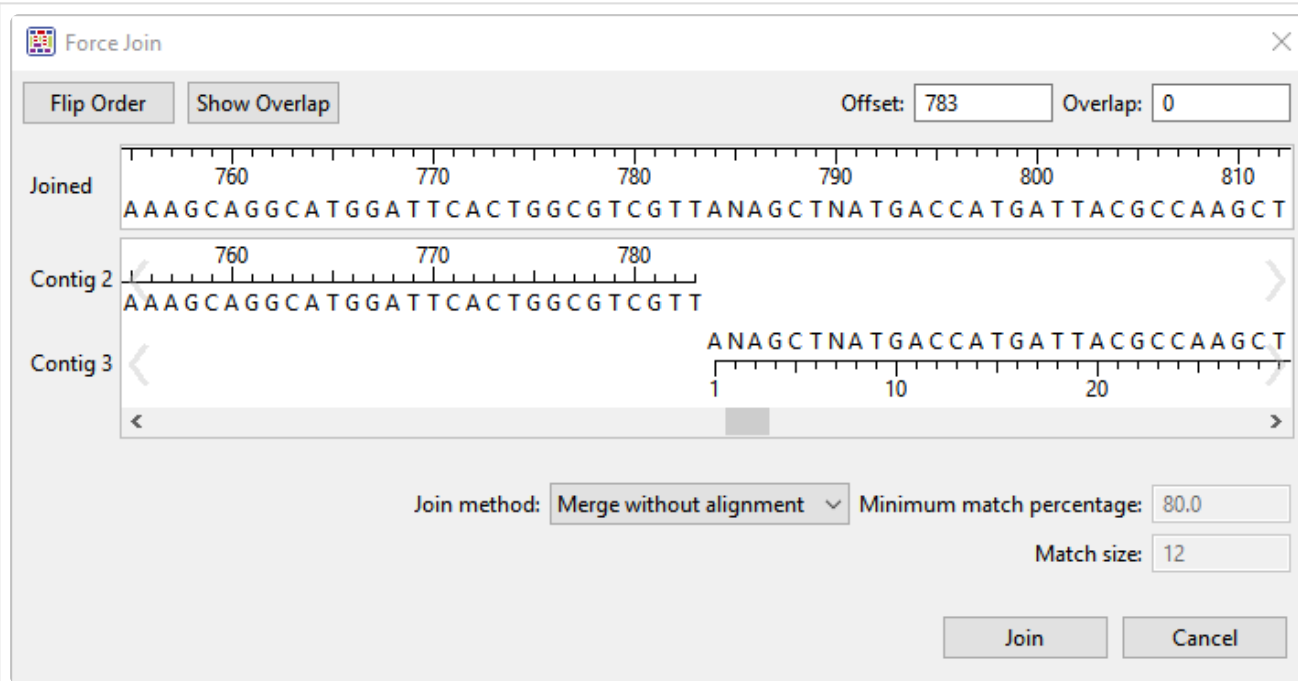
	contig keeps the original name and coordinates. The new 3' contig is given the identical name appended with "[2]".
<b>Length</b>	The length of the individual contig, including gaps and/or edits. Note that subtracting one <b>Name</b> coordinate from the other may yield a result that does not match the number shown in the <b>Length</b> column. This is because the <b>Name</b> coordinates are based on the ungapped and unedited reference sequence, while the <b>Length</b> value includes gaps and edits.
<b>Sequences</b>	The number of sequences in the contig.
<b>Position</b>	The position of each ordered contig or "scaffold. A "0" indicates that contigs or scaffolds have not been ordered.

# Organize contigs and scaffolds

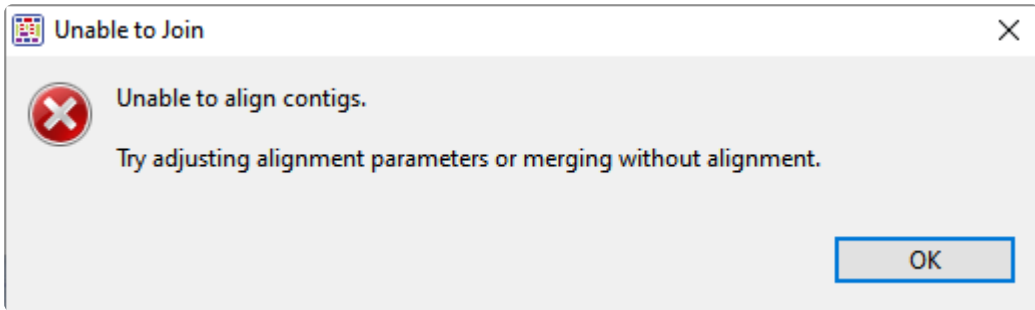
If your project is in .sqd format, you can edit and delete contigs; and create, edit and remove contig scaffolds as shown in the table below. Note that most of the following tasks require you to first select one or more contigs or scaffolds in the [Explorer panel](#).

Task	How To																																
Delete a contig	Select the contig in the Explorer panel, then use <b>Contig &gt; Delete Contig</b> .																																
Place contigs on a scaffold	<p>A “contig scaffold” is a collection of contigs that are related based on their known spatial relationships rather than their sequence similarity. SeqMan NGen uses paired end sequence data to put contigs into scaffolds. If you have independent evidence that two or more contigs are related, you may wish to scaffold the contigs manually. In the Explorer panel, use <b>Shift+click</b> or <b>Ctrl/Cmd+click</b> to select the contigs you wish to place together on a scaffold. Next, select <b>Contig &gt; New Scaffold</b> or right-click on a contig and choose <b>New Scaffold</b>. The contigs are now placed together in a scaffold at the bottom of the table.</p> <table><thead><tr><th>Name</th><th>Length</th><th>Sequences</th><th>Position</th></tr></thead><tbody><tr><td>▼ Unscaffolded Contigs</td><td></td><td></td><td>0</td></tr><tr><td>Contig 5</td><td>583</td><td>2</td><td>0</td></tr><tr><td>Contig 4</td><td>598</td><td>2</td><td>0</td></tr><tr><td>Contig 3</td><td>674</td><td>2</td><td>0</td></tr><tr><td>▼ Scaffold 1</td><td></td><td></td><td>0</td></tr><tr><td>Contig 2</td><td>783</td><td>3</td><td>0</td></tr><tr><td>Contig 1</td><td>698</td><td>3</td><td>0</td></tr></tbody></table> <p>Once at least one scaffold has been created, all other contigs are placed in the <b>Unscaffolded Contigs</b> scaffold. All positions in the <b>Unscaffolded Contigs</b> scaffold are initially set to 0. Note that the order of contigs in the <b>Unscaffolded Contigs</b> scaffold does not imply that adjacent contigs should be joined.</p>	Name	Length	Sequences	Position	▼ Unscaffolded Contigs			0	Contig 5	583	2	0	Contig 4	598	2	0	Contig 3	674	2	0	▼ Scaffold 1			0	Contig 2	783	3	0	Contig 1	698	3	0
Name	Length	Sequences	Position																														
▼ Unscaffolded Contigs			0																														
Contig 5	583	2	0																														
Contig 4	598	2	0																														
Contig 3	674	2	0																														
▼ Scaffold 1			0																														
Contig 2	783	3	0																														
Contig 1	698	3	0																														
Order contigs in a scaffold automatically	<p>In the Explorer panel, select the contigs using <b>Shift+click</b> or <b>Ctrl/Cmd+click</b>. Contigs must already be in the same scaffold. Next, choose <b>Contig &gt; Order Contigs</b> or right-click on the selection and choose <b>Order Contigs</b>. A warning appears.</p> <div><p>The dialog box titled "Order Contigs" contains a yellow warning triangle icon and the text "Automatically order 3 contigs? This cannot be undone." with an "OK" button.</p></div> <p>To continue with contig ordering, press <b>OK</b>. SeqMan Ultra will evaluate or reevaluate all contigs, including those already in scaffold. Note that changing the <b>Position</b> of a contig can influence whether pairs in different contigs are consistent.</p>																																
Order contigs in a	Contigs in a scaffold can be manually ordered to reflect known spatial arrangements between pairs. In the Explorer panel, edit the <b>Position</b> cell for each contig in the scaffold by clicking on it and typing in an																																

<b>scaffold manually</b>	integer representing the order in which the contig should be placed. To view the contigs in order, click once or twice on the <b>Position</b> header. Note that changing the <b>Position</b> of a contig can influence whether pairs in different contigs are consistent.
<b>Reverse complement contigs</b>	Select the contig(s) in the Explorer panel, then choose <b>Contig &gt; Reverse Complement (Contig Name)</b> .
<b>Split a contig</b>	To split a contig manually into two separate contigs at a specified location, select the contig in the Explorer panel. Use the toolbar on the right to open the <a href="#">Alignment view</a> , then place the cursor in the consensus at the position where you wish to split the contig. Finally, select <b>Contig &gt; Split at Insertion</b> . Reads that begin to the left of the insertion point remain in the same contig; reads that begin to the right will be assigned to a new contig. The name of the new contig will consist of the original contig name with %d appended to it. The contig position of the newly created contig is set to the position of the original contig plus the offset of the split location.
<b>Place unlocated contigs based on paired-end information</b>	Select the contigs in the Explorer panel, then use <b>Contig &gt; Place Unlocated Contigs</b> . SeqMan Ultra will use paired end information to place contigs from the unscaffolded contigs pool into a scaffold based on consistent pairs. In contrast to the <b>Contigs &gt; Order contigs</b> command, SeqMan Ultra will only attempt to place contigs from the unscaffolded pool.
<b>Join two or more adjacent contigs</b>	Select the contigs in the Explorer panel, then use <b>Contig &gt; Align contigs end-to-end</b> . This option is useful when you know the relative order of the contigs, as it prevents mis-joining of distant contigs that happen to have the same sequences at the contig ends. Realigning contigs is faster than reassembling them, as realigning only considers ~1 kilobase at the end of each contig.
<b>Replace a portion of the consensus sequence with a portion of a constituent sequence</b>	In the Alignment view, highlight the region of the constituent sequence you wish to use. Select <b>Contig &gt; Force Consensus</b> . After using this command, the selected bases from the constituent sequence will appear underlined. To revert to the original consensus, select the underlined range and use <b>Contig &gt; Compute Consensus</b> . It is not necessary to be exact, as long as all the forced bases you want to undo are included in the highlighted area. Once you use this command, underlining will be removed.
<b>Merge two contigs</b>	The Force Join dialog allows you to force merge two contigs into a single contig either manually or semi-automatically. If, for example, two contigs overlap by only 1 bp, they cannot be aligned using any alignment algorithm. You could use this interface to specify the 1 bp overlap and then force two contigs to merge. Conversely, If the overlap is 12 or more bp you can use this interface to align the ends semi-automatically.



1. Select two contigs from the Explorer panel and choose **Contig > Force Join Contigs**. The Force Join editor will appear, displaying the two selected contigs from 5'-3'.
2. Before joining the contigs, decide what order and orientation the contigs should be in. If you want to join the first selected contig to the 3' end of the second one, click the **Flip Order** button.
3. Produce an overlap by dragging the sequence ends or by entering **Offset** and **Overlap** values. For example, if the first sequence was 650 bp long and you wanted the second sequence to overlap the first by 20 bp, you would enter an **Offset** of 630 (i.e. the coordinate of the first sequence where the second sequence begins) and an **Overlap** of 20.
4. (optional) To center the overlapping area for a better view, press **Show Overlap**.
5. Finish by doing either of the following:
  - To align contigs without regard for similarity, select **Merge without alignment** from the **Join method** drop-down menu, then click **Join**.
  - To align the contigs only if they meet certain criteria, select **Align Overlap region** from the **Join method** drop-down menu. Enter the desired thresholds for **Minimum match percentage** and **Match size** values. Click **Join**. SeqMan Ultra will join the two contigs only if they meet the specified criteria. Otherwise, you will receive the message "Unable to align contigs."

	 <p>The dialog box titled "Unable to Join" contains a red "X" icon and the text: "Unable to align contigs. Try adjusting alignment parameters or merging without alignment." An "OK" button is located at the bottom right.</p>
<b>Restore the trimmed end(s) of a contig automatically</b>	<p>SeqMan Ultra can automatically restore up to 100 bases of the trimmed ends of sequences located at the beginning and ends of the selected contigs. The sequence will not be extended past any vector trim sites. Restoring trimmed ends allows you to evaluate whether some of the lower-quality trimmed data may provide evidence that two contigs may be joined. To restore ends automatically, select the contig in the Explorer panel, then choose <b>Contig &gt; Extend Contig Ends &gt; (Subcommand)</b>. Choose between: <b>Extend 3' Sequences ends</b>, <b>Extend 3' and 5' Sequences ends</b> and <b>Extend and Align 3' and 5' Sequences ends</b>. In each case, SeqMan Ultra will extend the entire trimmed region. Depending on how much was previously trimmed, the effect can be minimal or it can expose many new bases to include in the alignment.</p>
<b>Restore the trimmed end(s) of a contig manually</b>	<p>Select a contig in the Explorer panel and open it in the <a href="#">Alignment view</a>. Scroll to the 3' or 5' end of the sequence. Click the plus sign next to the sequence name to show the <b>Traces</b> data <a href="#">track</a>. Grab the small dark gray box to the left/right of the final base and drag your mouse to the side to reveal trimmed data.</p>

# Close a gap between contigs

---

In most cases, *de novo* sequence assembly will yield multiple contigs rather than a “perfect” single contig. DNASTAR offers several methods for filling in (AKA “closing”) the gaps between contigs:

- SeqMan NGen supports a **Hybrid reference-guided/de novo genome assembly** workflow. For information on this method, see the [SeqMan NGen help](#).
- SeqMan Ultra offers three ways to close gaps, as shown in the following video. One of these methods is described step-by-step in the section below the video.

To close a gap between two contigs in SeqMan Ultra:

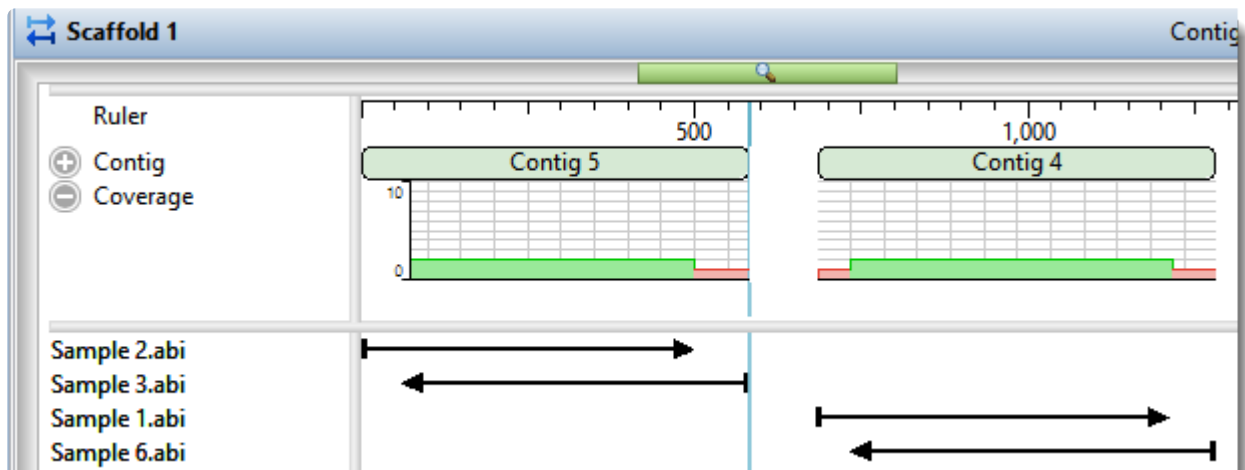
The following steps describe the most popular method for closing gaps using SeqMan Ultra. For other methods, see the video above.

Before you begin, note that this method works only for *.sqd* files (not *.assembly* files) that contain mate-pair data.

1. Open an *.sqd* assembly project in SeqMan Ultra.
2. Order and scaffold contigs using either of these methods:
  - Automatically: Choose **Contig > Order Contigs**. SeqMan Ultra will use mate-pair data to create one or more scaffolds consisting of multiple contigs already pre-numbered in increasing order.

- Manually: Use mate-pair data in the [Strategy view](#) to discover the relationship between contigs. When you find two or more contigs that you believe go together, select them in the [Explorer panel](#) and choose **Contig > New Scaffold**. Manually type their relative position numbers into the **Position** column. Choose any novel number (i.e. a number not already used in the Explorer panel) for the first contig, and larger numbers for each successive contig.

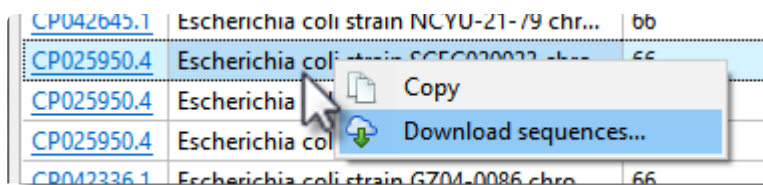
- Select the scaffold in the Explorer panel and open the Strategy view.



- Find the end/beginning of two contigs you wish to join. You can skip from one contig to the next by setting the [search](#) type to **Contig** and using the blue arrow keys.



- When you find a gap you would like to close, use your mouse to select a range of sequence near the end of either contig.
- Choose **Search > Search** to initiate a BLAST search of that contig end. Choose the desired settings and press **Run Now** or **Run**.
- Once the job finishes, click the link in the **Status** column of the [Jobs panel](#) to open the search results.
- Click on the best match, then right-click and choose **Download Sequences**.



- In the Explorer panel, select the two contigs on either side of the gap and choose **Contig > Add**

## Sequences to Close Gap.

**Close Gap Between Contigs**

**Add Sequences**

Add sequences to join Contig 5 with Contig 32

Name	Limits	Type

Alignment options:

Minimum match percentage:  %

Match size:  bp

Gap penalty:

Buttons: Add..., Remove, Remove All, Run, Cancel

10. Use the **Add** button to upload the saved match, adjust the alignment options as desired, and click **Run**.

A report shows details of the gap closing attempt, including whether or not it appeared to be successful.

11. Once you complete a successful gap closure, open the Strategy view to view the closed gap.
12. If there are additional gaps, follow the same steps to close each of them.
13. Once you have closed as many gaps as possible, [export the consensus](#) to use in setting up a new templated assembly in SeqMan NGen.

# Display contig information in a graphical or tabular view

SeqMan Ultra's data display views allow you to perform in-depth analysis on a contig created as part of a SeqMan NGen assembly.

These views can be accessed using commands in the **View** menu, but the fastest way to access them is to use the toolbar to the right of the [Explorer panel](#). Available tools vary depending on the type of assembly.







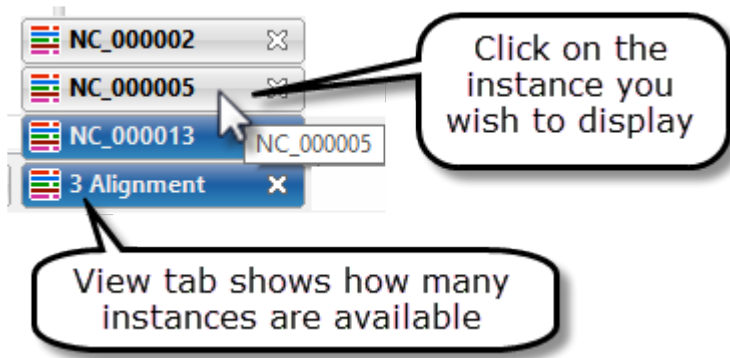
 **Note:** Once two views are open, subsequent views will open in separate moveable and resizable windows. To reattach the views, see [Customize the interface layout](#).

Image	Tool Name	Shortcut to Menu Command	Description
	Show alignment view of selected contigs	View > Alignment	Causes one <a href="#">Alignment view</a> to open for each selected contig.
	Show strategy view of selected contigs or scaffolds	View > Strategy	Causes one <a href="#">Strategy view</a> to open for each selected contig or scaffold.
	Show table of features in selected contigs	View > Features	Causes one <a href="#">Features View</a> to open for each selected contig.
	Show table of SNPs in selected contig	View > Variants	This option is only available for reference-guided .assembly projects and causes one <a href="#">Variant View</a> to open for each selected contig.
	Show report view of selected contigs	View > Contig Report	Causes one <a href="#">Contig Report View</a> to open for each selected contig. This view comes in three versions: <a href="#">Contig Information</a> , <a href="#">Contig Coverage</a> , <a href="#">Contig Structural Variation</a> . By default, the <a href="#">Contig Information</a> tab is shown.


SeqMan Ultra allows multiple instances of a view to be created, some of which may share a common view tab. As shown in the image below, the view tab shows how many views of that type are in that tab folder. To display a different instance, hover over the view tab until you see a popup showing each instance by name, then click on the desired instance.

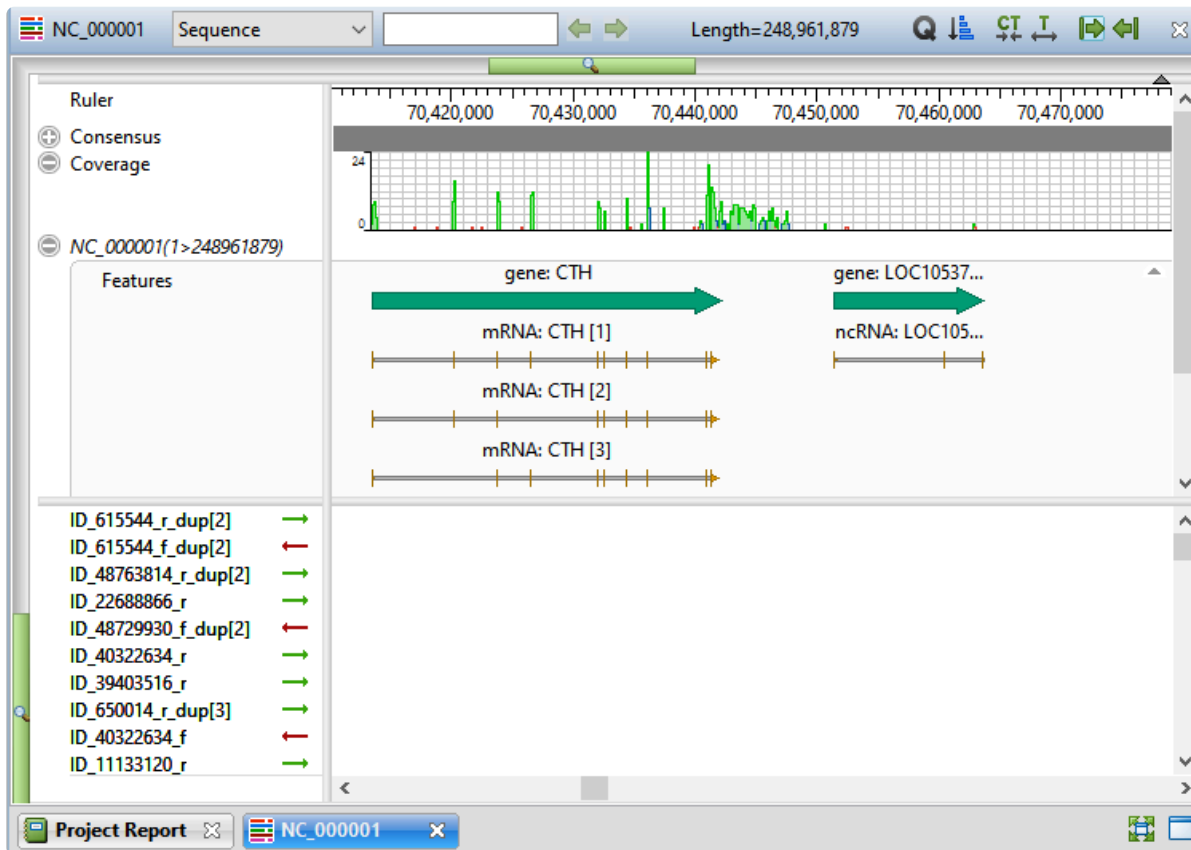




To learn how to show, hide, close, detach, reattach views or to make them full screen, see [Customize the Interface Layout](#).

# Alignment View

The Alignment view displays the consensus sequence and the alignment of all sequences making up a contig at the nucleotide level. The Alignment view also allows you to edit contigs and contig scaffolds.

To access the Alignment view, choose a contig from the [Explorer panel](#) and either use the **Show Alignment view of selected contigs** tool (  ) or choose **View > Alignment > Show Alignment View**.



Constituent sequences are shown in the lower portion of the view. To see the sequence orientation, examine the arrow to the left of each sequence string. A right arrow (  ) indicates that SeqMan NGen used the input orientation for the alignment. A left arrow (  ) shows that SeqMan NGen used the complementary strand.

To learn the length/range of a sequence, look at the sequence name in the left margin of the view. The sequence range is shown in parentheses.

To learn more about using this view, see the following topics:

- [Alignment view toolbar](#)
- [View data tracks in the Alignment view](#)
- [View multi-sample assemblies in the Alignment view](#)




- [Edit sequences](#)
- [Search within a contig](#)
- [Make a selection](#)
- [Customize appearance and behavior](#)

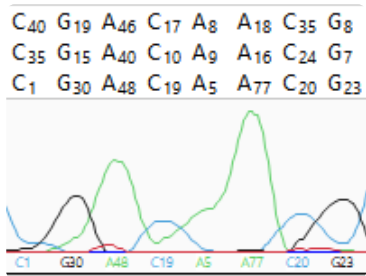





#### U's vs. T's in the Alignment view:

If you are viewing a SeqMan NGen assembly in which DNA sequencing reads were assembled against an RNA reference sequence, the reference row in the Alignment view will contain U's where the read rows have T's. Though not actually mismatches, the T's will appear in the mismatch [color scheme](#) in the Alignment view. By contrast, the T/U equivalency is recognized in the [Variant view](#), and T's there are not colored as variants.

# Alignment view toolbar

The tools located in the upper right of the [Alignment view](#) are used to accomplish the following tasks.

Task	How to
To display the search tools	To display the <a href="#">search</a> tools in the upper left of the Alignment view, use the <b>Search alignment</b> tool (  ).
To sort constituent sequences	<p>To sort read files (the files in the left margin of the view) according to the selected criteria, use the <b>Sort reads</b> tool (  ).</p> <ul style="list-style-type: none"> <li>• <b>By Name</b> – Sorts read files in alphanumeric order by name. This type of sorting is normally only used for Sanger reads, which are often user-named.</li> <li>• <b>By Position</b> – Sorts read files by their position along the consensus.</li> <li>• <b>By Variant</b> – Brings all the reads containing variant bases to the top of the alignment stack. When you choose this option, SeqMan Ultra automatically loads and calculates the SNPs. Note that if you are working with a large <i>.sqd</i> project, this calculation can take a considerable amount of time.</li> </ul>
To hide/show quality scores	<p>Use the <b>Quality scores</b> tool (  ). This tool only functions when a) trace data was used in the assembly.</p> <ul style="list-style-type: none"> <li>• <b>Hide scores</b> – Causes quality scores to be hidden from view.</li> <li>• <b>Show scores</b> – Causes quality scores to be displayed as a subscript next to the base. When bases with very high quality scores (&gt;99% of the 95th percentile) are displayed, two asterisks (**) will be shown in place of the usual numerical subscripts. Example: G<sub>52</sub> A<sub>84</sub> G<sub>**</sub> C<sub>85</sub>. The quality score of a peak is calculated based on the shape and height of each peak and is adjusted relative to the maximum height of any peak in the entire sequence. Taller, sharper peaks receive the highest quality scores. The heights of any underlying peaks are subtracted from the highest peak's score. Illumina data will typically have higher scores than Sanger data. For <i>.fas</i> files, SeqMan NGen automatically checks for <i>.qual</i> files with the same name in the same folder. If found, the quality values are displayed. Note that gaps are not assigned quality scores.</li> </ul>

	 <ul style="list-style-type: none"> <li>• <b>Show averaged scores</b> – Displays quality scores averaged over a defined window. This mode is useful for examining trimming. Note that gaps are not assigned quality scores.</li> </ul>
<b>To adjust the zoom level</b>	Use the <b>Compact text</b> tool (  ) to adjust the zoom level to maximize the number of bases while keeping the text legible. Use the <b>Restore default zoom</b> tool (  ) to adjust the zoom level to the (wider) default setting, making the text more comfortable to read.
<b>See how the sequence is anchored</b>	Prior to correcting an .sqd alignment with trace data (only), use the <b>Anchor editing left (insert right)</b> tool (  ) to see how the sequence is anchored. If the left anchor is active, residues left of the insertion point are held in place. If you insert any characters, the downstream sequence slides to the right. If you delete any bases, the downstream sequence fills in the gap by sliding left. Use the <b>Anchor editing right (insert left)</b> tool (  ) in a similar way.
<b>Export data</b>	Use the <b>Export data</b> tool (  ) . This tool acts as a shortcut to the <b>File &gt; Export Data &gt; Consensus</b> command and opens the same popup dialog. See <a href="#">Export Data</a> for information on using the dialog.

✿ **Note:** Right-click command options are available by right-clicking on a sequence rather than the name of the sequence.

# View data tracks in the Alignment view

Tracks are graphical data displays that run along the length of your sequences in the [Strategy](#) and [Alignment](#) views. To learn how to apply, show or hide tracks, or to edit track options, see [Work with Data Tracks](#). The following table shows all the tracks that can be displayed within the Alignment view.


Tracks panel category & track name	Image																																										
Alignment header or footer > Consensus	<div><div>Consensus</div><div>TACGCGCCATGTCGAGG - ATACGG - TCAACACCAGCAATA</div></div>																																										
Alignment header or footer > Consensus > Consensus translation	<div><div>Consensus translation</div><div><table><tr><td>1</td><td>T</td><td>R</td><td>H</td><td>V</td><td>E</td><td>D</td><td>T</td><td>V</td><td>N</td><td>T</td><td>S</td><td>N</td><td>K</td></tr><tr><td>2</td><td>R</td><td>A</td><td>M</td><td>S</td><td>R</td><td>I</td><td>R</td><td>S</td><td>T</td><td>P</td><td>A</td><td>I</td><td>N</td></tr><tr><td>3</td><td>A</td><td>P</td><td>C</td><td>R</td><td>G</td><td>Y</td><td>G</td><td>Q</td><td>H</td><td>Q</td><td>Q</td><td>•</td><td>T</td></tr></table></div><div><ul style="list-style-type: none"><li>Stop codons are represented by dots.</li><li>Translations are not displayed for incomplete codons.</li><li>Translation frames 4-6 are counted from the right end (the 3' end of the bottom strand).</li></ul></div></div>	1	T	R	H	V	E	D	T	V	N	T	S	N	K	2	R	A	M	S	R	I	R	S	T	P	A	I	N	3	A	P	C	R	G	Y	G	Q	H	Q	Q	•	T
1	T	R	H	V	E	D	T	V	N	T	S	N	K																														
2	R	A	M	S	R	I	R	S	T	P	A	I	N																														
3	A	P	C	R	G	Y	G	Q	H	Q	Q	•	T																														
Alignment header or footer > Coverage	<div><div>Coverage</div><div></div></div>																																										
Alignment header or footer > Reference	<div><div>NC_012920(1&gt;17260)</div><div>AAAGTGAAGTGTATCCGACATCTGGTTTCCTACTTCAGGGTCATAAAGCCTAAA</div></div>																																										
Alignment header or footer > Reference > Features	<div><div>Features</div><div></div></div>																																										
Alignment header or footer > Ruler	<div><div>A gap-aware sequence ruler.</div><div></div></div>																																										

Alignment reads	<div><div>AAC1D:822:306 AAC1D:914:1052 IB8EW:562:24 IB8EW:636:1225 RKGS3:950:649</div><div><div>←</div><div>←</div><div>←</div><div>→</div><div>←</div></div><div>TACGCGCCAT TACGCGCCATGTCGAGG - ATACGG - TCAACACCAGCA TACGCGCCATGTCGAGG - ATACGG - TCAACACCAGCA TACGCGCCATGTCGAGG - ATACGG - TCAACACCAGCA TACGCGCCATGTCGAGG - ATACGG - TCAACACCAGCA</div></div>
Alignment reads > Features	
Alignment reads > Traces	<div>This option is only available for .sqd projects that used Sanger/ABI data.</div> <div><div>Traces</div><div><div>G A G C C G C T A T C C G G T C A G T T C T T C C T G G C T G C T C A A C C G C C</div></div></div>

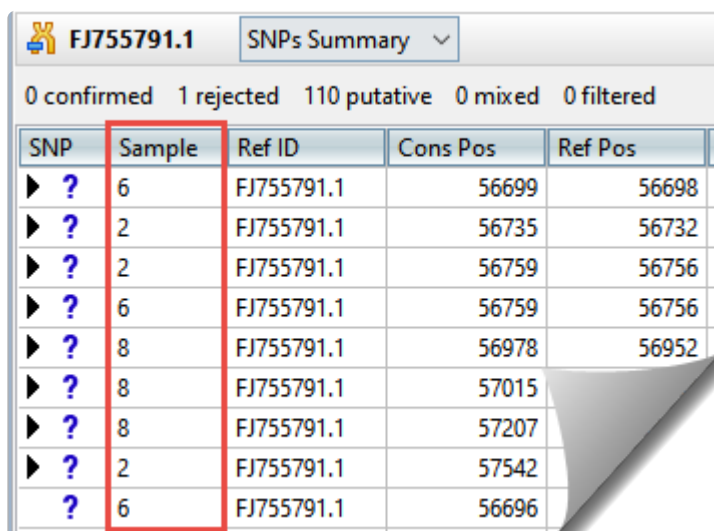
# View multi-sample assemblies in the Alignment view

If—when setting up the SeqMan NGen assembly—you chose **Multi-sample** or **Multi-sample with replicates** in the [Input sequences](#) screen, SeqMan Ultra displays a special version of the [Alignment view](#). This version shows a “pseudo-consensus” where you can compare different strains, varieties, etc., and easily see which samples have variants.

To view multi-sample assemblies in the Alignment view:

1. Open a multi-sample assembly.
2. In the [Explorer panel](#), select the contig of interest.
3. Press the **Show table of variants in selected contigs** tool () to launch the [Variants view](#).

Note that the **Sample** column contains the **Experiment** or **Replicate** name you specified on SeqMan NGen's Input sequences screen.

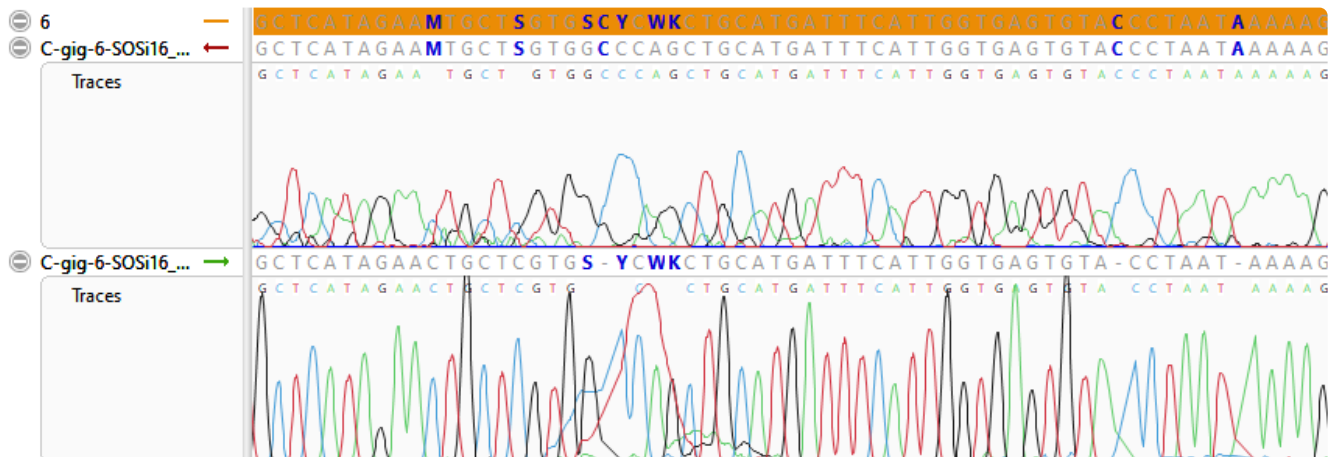


SNP	Sample	Ref ID	Cons Pos	Ref Pos
▶ ?	6	FJ755791.1	56699	56698
▶ ?	2	FJ755791.1	56735	56732
▶ ?	2	FJ755791.1	56759	56756
▶ ?	6	FJ755791.1	56759	56756
▶ ?	8	FJ755791.1	56978	56952
▶ ?	8	FJ755791.1	57015	
▶ ?	8	FJ755791.1	57207	
▶ ?	2	FJ755791.1	57542	
▶ ?	6	FJ755791.1	56696	

4. Double-click on a variant of interest to launch the [Alignment view](#) at that position. The view shows SNPs in blue text on an orange background, allowing you to easily see which samples have variants.

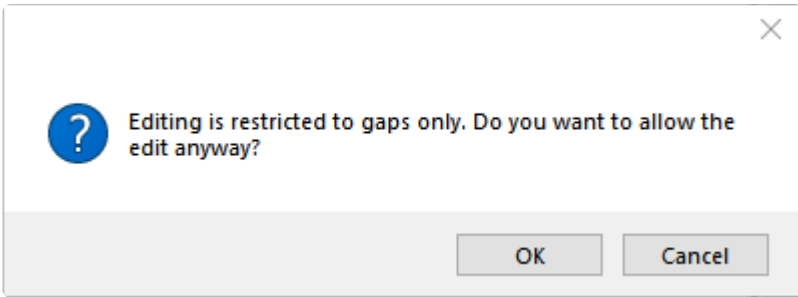


- If desired, click the plus signs to expand the samples and view the individual read sequences. You can also expand the read sequences in the same way to see trace data, if trace files were used in the assembly.



# Edit sequences in the Alignment view


In the [Alignment view](#), you may edit constituent or consensus sequences in an `_sqd_` project by removing or adding nucleotides or gaps. Only the editing of gaps is supported.

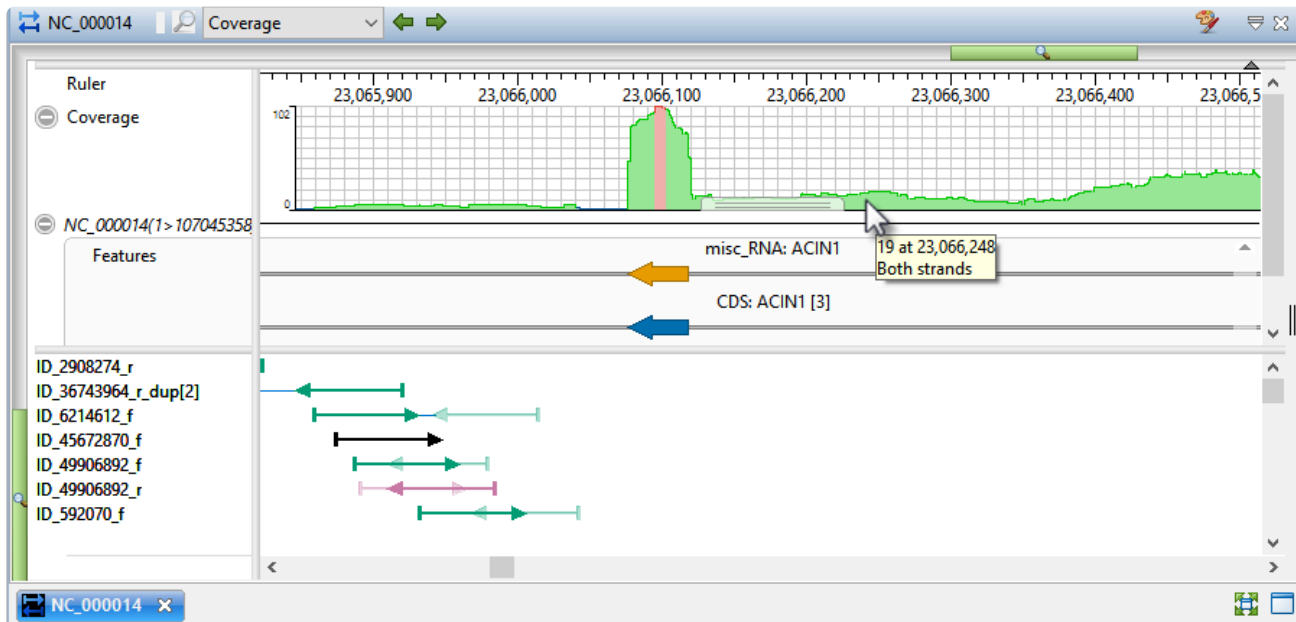
Task	How To
<b>Remove or add gaps</b>	Insert the cursor into an individual sequence and type one or more dashes (-) or delete the same. As you do this, the Consensus sequence will be recalculated per your changes.
<b>Remove or add nucleotides</b>	<p>If you attempt to type in or delete sequence other than gaps, you will receive the following warning:</p>  <p>If you still wish to proceed, click <b>OK</b>. You may then insert the cursor into an individual sequence and type in an <a href="#">IUPAC nucleic acid code</a> or delete the same. As with gaps, the Consensus sequence will be recalculated. When you add new nucleotides, they will appear in lower case to signify a manual addition.</p>
<b>Cut or copy a selected portion of the sequence</b>	Select some or all of the sequence, then right-click and choose <b>Cut</b> or <b>Copy</b> .

Note that changes made to a constituent sequence only directly affect that sequence, but changes made to the consensus sequence affect the consensus and every constituent sequence in the contig. Note that the original sequence files stored on your computer will not be affected by these procedures.

# Strategy View



The Strategy view graphically summarizes the position and orientation of every constituent sequence in a contig or contig scaffold.

To access the Strategy view, choose a contig from the [Explorer panel](#) and either use the **Show Strategy view of selected contigs or scaffolds** tool () or choose **View > Strategy > Show Strategy View**.



Tools in the Strategy view:

The following tools are located in the top right of the Strategy view window.


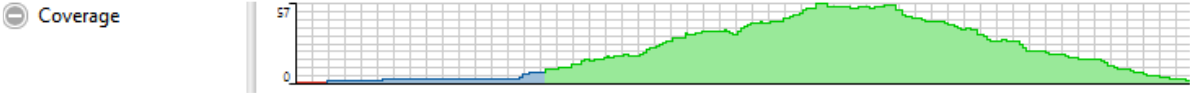
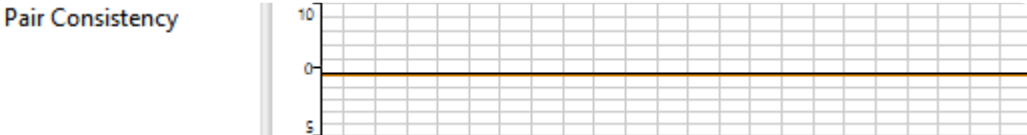
- To display the [search](#) tools in the upper left of the Strategy view, use the **Search alignment** tool () .
- To open the [Style panel](#), where you can customize the appearance of the view, use the **Show Style panel** tool () .

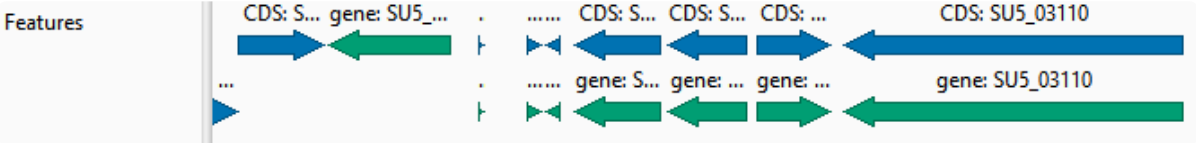
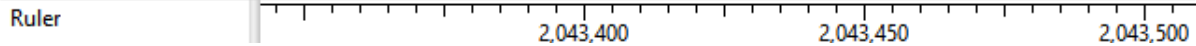
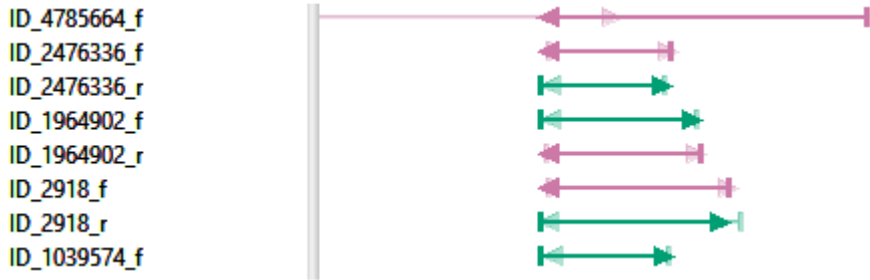
To learn more about using this view, see the following topics:

- [View data tracks in the Strategy view](#)
- [Search within a contig](#)
- [Make a selection](#)
- [Customize appearance and behavior](#)

# View data tracks in the Strategy view

Tracks are graphical data displays that run along the length of your sequences in the [Strategy](#) and [Alignment](#) views. To learn how to apply, show or hide tracks, or to edit track options, see [Work with Data Tracks](#). The following table shows all the tracks that can be displayed within the Strategy view.

Tracks panel category & track name	Image
Strategy header > Contig	<p>Applying this track only has an effect if you are displaying the Strategy view for a contig <a href="#">scaffold</a>. If a scaffold Strategy view is open, the track shows bars denoting the placement and name of each contig.</p> 
Strategy header > Coverage	<p>This histogram represents the depth of sequences throughout the assembly. The numbers to the left of the histogram indicate the actual number of sequences represented. Hover over any part of the histogram to display information about that location.</p>  <p>Based on the <a href="#">Strategy Viewing and Coverage</a> parameters, the histogram will appear:</p> <ul style="list-style-type: none"> <li>• thick green where coverage meets both the <b>Coverage threshold</b> and <b>Minimum number on each strand</b> criteria</li> <li>• thin green where both strands are covered but <b>Coverage threshold</b> is not met</li> <li>• thin blue where coverage is on one strand only</li> <li>• red where coverage is by a single read only</li> </ul>
Strategy header > Pair Consistency	<p>The Pair Consistency histogram displays a summary of either split reads (for non-paired data) or good versus bad paired end sequence data content (for paired reads). Hover over any part of the histogram to display information about that location.</p>  <p>The deeper the Pair Consistency histogram, the more recognized pair of reads there are. Note that SeqMan Ultra cannot distinguish whether inconsistent pairs are due to naming errors or assembly errors.</p> <ul style="list-style-type: none"> <li>• Green above the baseline indicates the presence of pairs that are consistent with the current assembly.</li> </ul>

	<ul style="list-style-type: none"><li>• Red below the baseline indicates the presence of pairs that are inconsistent with the current assembly due to relative orientation, distance apart, or both.</li><li>• Gold below baseline spans regions corresponding to paired reads in different contigs whose assembly locations or orientations are inconsistent with pair specifier parameters, even if the contigs were rescaffolded or reordered.</li><li>• Pink below the baseline represents split reads in the same contig whose locations or orientation inconsistent with pair specifier parameters.</li></ul>
Strategy header > Reference > Features	
Strategy header > Ruler	<p>A gap-aware sequence ruler.</p> 
Strategy header > Split reads	
(Sequences)	<p>The lower half of the Strategy view displays arrows representing the individual sequences in the contig, including color-coded information about paired-end data.</p>  <p><u>To see the location of the second member of a pair</u>, hover over the first member for a tooltip with coordinates and its mate pair.</p> <p><u>To move to the location of the second member of a pair</u>, right-click on the first member and select <b>Show Mate Pair</b>.</p> <p>The point of the arrowhead represents the 3' end and the tail represents the 5' end. Solid arrows pointing right represent sequences assigned to the top strand of the consensus sequence. Dash-lined arrows pointing left represent sequences assigned to the bottom (complementary) strand of the consensus sequence. The arrow color uses the color coding shown in the table below.</p> <p>Paired end sequence data are forward and reverse sequence reads originating at opposite ends of the same DNA fragment. Even though such a pair of reads may not overlap by sequence similarity, they belong in the same contig. If the reads are assembled into different contigs, the knowledge they are linked may help in joining the contigs together though it may instead indicate that the reads are misassembled or misnamed. Paired end data can also be used for within-contig assembly evaluation. If the two reads for a given pair are not in opposite</p>

orientations, for instance, an assembly problem may be indicated. Finally, if the distance between the two reads is known, this information can also be used to confirm whether or not the assembly is correct. When using paired end sequence data to evaluate assemblies, be sure to allow for the possibility that there may be errors in some of the sequence names.

Certain actions may affect the current classification (coloring) of the sequences in the Strategy view. For example, deleting a contig or scaffold in the Explorer panel may remove one of a paired end pair. The color of its pair sequence would then be updated to black. Other actions that affect coloring include: adding or deleting sequences from your project or reverse complementing, aligning, force joining, splitting, or ordering contigs.


#### Color-coding for reads:

Color	Description
<b>Black</b>	Represents reads for which no paired data information is available or recognized.
<b>Green</b>	Represents paired reads in the same contig whose assembly locations are consistent with the <a href="#">pair specifier parameters</a> .
<b>Dark Blue</b>	Represents paired reads in different contigs in the scaffold whose assembly locations are consistent with pair specifier parameters.
<b>Pale Blue</b>	Represents paired reads in different contigs (in the same or different scaffolds) whose assembly locations could be consistent with pair specifier parameters if the contigs were rescaffolded or reordered. The number appearing next to the pale blue arrows indicates which contig the other member of the pair is in.
<b>Orange</b>	Represents paired reads in different contigs whose assembly locations or orientations are inconsistent with pair specifier parameters, even if the contigs were rescaffolded or reordered.
<b>Red</b>	Represents paired reads in the same contig whose locations or orientation are inconsistent with pair specifier parameters.
<b>Pink</b>	Represents split reads in the same contig whose locations or orientation are inconsistent with pair specifier parameters.

# Contig Report View


---

The three types of Contig Report views display detailed information about a contig. To access the Contig Report view, choose one or more contigs from the [Explorer panel](#) and do any of the following:

- From the toolbar on the right, press the **Show report view of selected contigs** tool (.
- Choose **View > Contig Report**.
- Right-click on the selected contig and choose **Contig Info Report**.

The Contig Report comes in three variations. To choose which variation of the report to display, use the drop-down menu on the upper left of the Contig Report view. Note that different types of assemblies may allow only a subset of view types or may contain one or more empty Contig Report views.

- [Contig Information](#)
- [Contig Coverage](#)
- [Contig Structural Variation](#)

To export data from the view, press the **Export data** tool () in the top right of the view. This opens a save dialog in which you can save the tabular data in comma-separated (.csv) or tab-separated (.tab) format.

# Contig Information

Contig Information is one of three Contig Report views that may be available, depending on the assembly type that is currently open. For information on accessing or performing tasks in this view, see [Contig Report View](#).

The information in this view depends on the type of assembly (i.e., *.sqd* vs. *.assembly* ; templated vs. non-templated). Some parameter descriptions are shown in the table below:

Parameter	Description
<b>Contig Length</b>	The length of the consensus sequence. This length reflects gaps inserted into the individual sequences and any editing you have done to the contig.
<b>Average Length/Sequence</b>	The quotient of <b>Total Sequence Length</b> and the number of sequences in a contig.
<b>Total Sequence Length</b>	The sum of all individual sequence lengths. This length reflects gaps inserted into individual sequences.
<b>Top Strand</b>	The number of sequences assigned to the top strand of the consensus sequence.
<b>Bottom Strand</b>	The number of sequences assigned to the bottom strand of the consensus sequence.
<b>Total</b>	The total number of sequences used to construct a contig.
<b>Sequences ordered by position</b>	Summarizes the offset position and length for each sequence in a contig.
<b>Coverage report by position</b>	Summarizes coverage in four categories: only sequenced once, sequenced only on one strand, sequenced on both strands but below threshold, and sequenced on both strands and equal to or above threshold. Each region is ordered by its position in the contig, from left to right. The threshold value is set in <a href="#">Coverage</a> parameters.
<b>Coverage Summary (in bases)</b>	Summarizes the number and frequency of bases in each coverage category.

Double click on hyperlinked text to open the [Alignment view](#) for the contig in which that sequence appears, with the position or range already selected.

# Contig Coverage

Contig Coverage is one of three Contig Report views that may be available, depending on the assembly type that is currently open. For information on accessing or performing tasks in this view, see [Contig Report View](#).

This tabular view contains the following columns:

Column Name	Description
<b>Contig ID</b>	The name of the contig.
<b>Type</b>	The type of coverage for the particular region.
<b>Feature</b>	If a feature on the reference sequence intersects the area of coverage, the feature name will be listed in this column. Note that when a Contig Coverage report is created from features, the features are used to confine the coverage analysis to the areas that have features. In this case, the <b>Feature</b> column shows which feature was used to constrain the analysis area.
<b>Location</b>	The range of sequence corresponding to the region of coverage.
<b>Length</b>	The length of the range of sequence corresponding to the region of coverage.
<b>Depth</b>	The minimum depth of coverage in the specified region. Note that adjacent areas with identical types of coverage will be combined. Note that the “Depth” column may not match the depth of coverage seen in the lower portion of the <a href="#">Alignment view</a> . For instance, the Contig Coverage report may show a depth of “3” for a particular range even though you can see 5 or more sequences in the same range of the Alignment view. The apparent discrepancy occurs because the Coverage Report shows the minimum depth of coverage for a location, while the Alignment view shows all sequences that contribute to the coverage.
<b>SNP Count</b>	The number of variants in the region. This column is only present if the sequences were assembled to a reference/template.

# Contig Structural Variation

In addition to variants and small insertions and deletions, genetic variation can also involve large scale rearrangements. These rearrangements may include large insertions and deletions, inversions, and translocations — collectively known as *structural variations* (SV's). During a templated data assembly with DNASTAR's SeqMan NGen, potential insertions and deletions are coded as such in the assembly output. When the assembly project is opened in SeqMan Ultra, the encoded SV information can be viewed in tabular format in the Contig Structural Variation report. For information on accessing or performing tasks in this view, see [Contig Report View](#).

This tabular view contains the following columns:


Column Name	Description
<b>Contig ID</b>	The name of the contig in which the SV was found.
<b>Contig Pos</b>	The position where the SV was detected in gapped coordinates. For insertions, a single coordinate is given. For deletions, the first and last coordinates of the deleted region are given.
<b>Ref Pos</b>	the position where the SV was detected in ungapped reference coordinates.
<b>Length</b>	Total length of a deletion in bases.
<b>Type</b>	The type of SV: Ins (insertion), Del (deletion), or Indel (substitution).
<b>Split Read Count</b>	The number of "split reads" defining the breakpoints of a deletion.
<b>Library</b>	A number corresponding to the library with information in that row. This column is present only for assemblies with multiple mate pair libraries of different insert sizes. General information on each library can be seen on a per contig basis using the <b>Contig &gt; Contig Info</b> command.
<b>Pair Dist</b>	The median distance between mate pair reads spanning the SV. For deletions, this distance is approximately the average distance between all pairs in the assembly plus the size of the deleted segment. For insertions smaller than the insert size of the mate pair library, this distance is approximately the average distance between all pairs in the assembly minus the size of the inserted segment. For insertions larger than the insert size of the mate pair library, no distance is reported because no pairs will span the insertion. For assemblies with multiple mate pair libraries of different insert sizes, a separate row for each library with spanning pairs is shown for each SV.
<b>Pair Count</b>	The number of mate pairs spanning the SV. For assemblies with multiple mate pair libraries of different insert sizes, a separate row for each library with spanning pairs is shown for each SV.
<b>Coverage</b>	The average depth of coverage across a putative deleted or indel region. The Coverage value is the mean depth of coverage over all columns, between the two edges of the deletion, in the original (unsplit) assembly. A "good" deletion will normally have much lower coverage than the flanking regions of the assembly. In cases where the deleted region is composed of non-repetitive sequence, the Coverage value will typically be zero or near zero. In cases where the deleted region is composed of repetitive sequence (e.g. an insertion sequence [IS] element), this value will typically be some fraction of the average coverage of the entire assembly. The exact value will depend on how many instances of the repeat are in the reference and the genome being sequenced. Note that clicking a row in the Contig Structural Variation report will take you to the

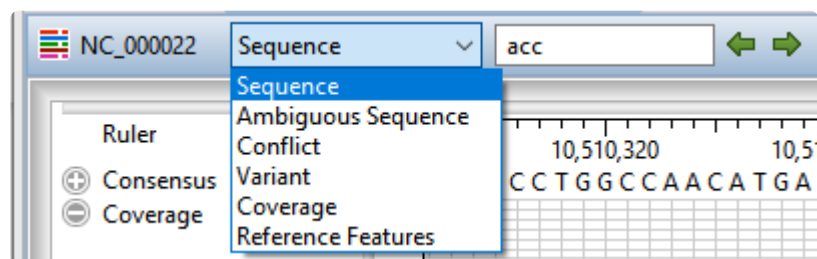
	<p>Alignment View column just prior to the left edge of the deletion. However, <b>Coverage</b> does not measure the depth of coverage of that column, as the column is excluded from the range over which the mean is taken.</p> <p><b><u>Coverage FAQ:</u></b></p> <p>Q: In the SV Report for a gap-closure workflow, why does the <b>Coverage</b> column display low values (e.g., 0-7) for the majority of reference positions?</p> <p>A: Coverage refers to the read depth in the original templated assembly and is calculated by summing the total number of aligned bases in that region and then dividing by the length of the region. Coverage is therefore expected to be low for deletions and indels. It is important to note that the Structural Variation Report is static and does not change after being created in SeqMan NGen.</p> <p>Q: Why doesn't the <a href="#">Alignment view</a> for an SV with a <b>Coverage</b> of <b>4</b> have four sequences of coverage?</p> <p>A: That's because the deleted region is an insertion sequence (IS) repeat that is present elsewhere in the reference and in the data set. Since this is repeated sequence, SeqMan NGen's assembler sees the match and places only a portion of the repeated reads in that spot. We can tell the region is deleted by the edges that form on both ends, the number of split reads defining the deletion endpoints, the relatively low depth in the deleted area and the lack of left and right joining pairs (pairs with one end outside the deletion and the other end in the putative deleted region). In cases where the deleted region is composed of unique sequence, the coverage will be zero or very near zero.</p>
<b>Feature</b>	<p>The feature(s) affected by the SV. For deletions affecting multiple features, the first and last feature are displayed separated by an ellipsis (e.g. <i>polB...ilvH</i>). To see a complete list of affected genes for a given SV, select the SV and use the <b>Features &gt; Show Feature Table</b> command. Note that the Contig Structural Variation report does not explicitly show features of type "DNA_SPLIT," as this feature type merely denotes the presence a structural variation.</p>



**Note:** For multi-sample data, structural variation is reported per MID-tag, per sample.

# Search within a contig

By default, the search tools are hidden. To display them, click on the **Search alignment** tool () in the upper right of the [Alignment](#) or [Strategy](#) view. This causes the search tools to be displayed on the top left of that view. The Alignment view lets you search by all variables in the table below, while the Strategy view only allows you to search for coverage and reference features.




1. Use the search bar drop-down menu to choose the item to search for.

Search Target	Description
<b>Sequence</b>	Performs a literal (perfect match) sequence search on the query typed into the text box at right. You can use IUB ambiguity codes in literal sequence searches, but these will match only the same ambiguity code in the sequence. For example, searching for “S”—the ambiguity code for G or C—will locate only S, and neither G nor C. This can be useful for locating specific heterozygotes. To search for gaps, enter a dash (-) into the text box.
<b>Ambiguous Sequence</b>	Performs a sequence search on the query typed into the text box at right. Unlike the <b>Sequence</b> option, <b>Ambiguous Sequence</b> recognizes IUB ambiguity codes in the query. If you enter the query GATS, for example, you would find GATC, GATG and GATS.
<b>Conflict</b>	Locates any base that differs from one or more bases in other sequences in the same column of the alignment. SeqMan Ultra will find and highlight single conflicting residue columns or consecutive groups of conflicting columns. These conflicts occur when the consensus cannot be called cleanly, so there are ambiguous calls in the consensus. <b>Conflict</b> is available for <i>de novo</i> assembly projects only.
<b>Variant</b>	Locates bases that have been identified by SeqMan NGen’s SNP-calling detector as variants. <b>Variant</b> is available for reference-guided assembly projects only.
<b>Coverage</b>	Locates areas with any amount of coverage, including depth=1.
<b>Reference Features</b>	Locates features on the consensus sequence.

2. If you choose either of the **Sequence** options, you will need to enter the query sequence (text string) in the text box before using the arrows. Otherwise, skip to the next step.

If you type a query that is “too wide” for the search box, you can hover over the search box to view


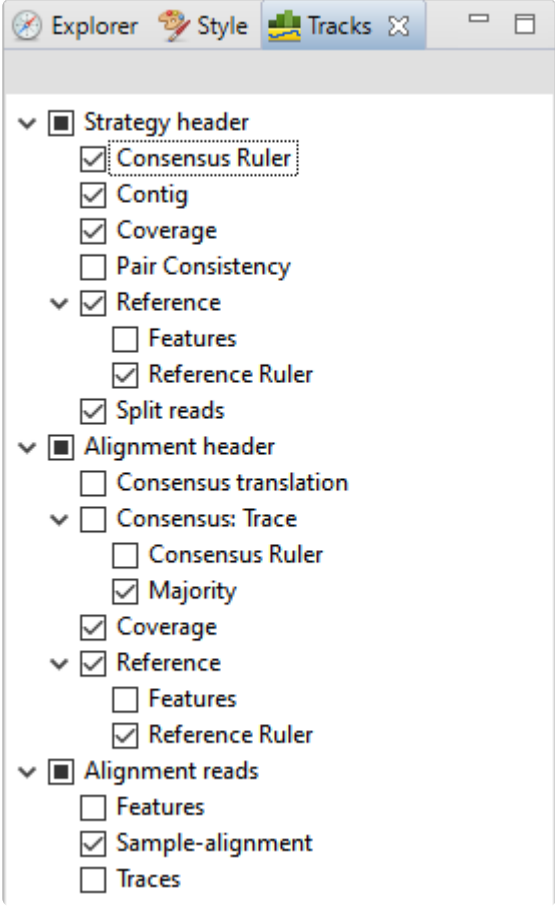
the entire query. Note that long query sequences may take considerable time to locate.



3. Use the right/left arrows (  ) to jump to the next/previous find. After reaching either end of a sequence, SeqMan Ultra wraps around and continues searching from the other end.

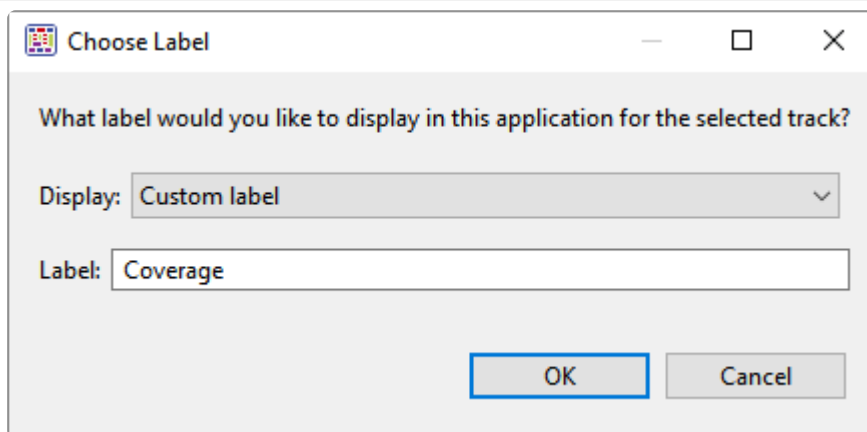
# Data Tracks

“Tracks” are rows of information that can be displayed or hidden, as desired, in the [Alignment](#) and [Strategy](#) views. Data for some tracks are imported automatically with the sequence data (e.g., feature tracks), while others are calculated with SeqMan Ultra.

The following table shows tasks that affect the display of tracks:

Task	How to...
To display a track	<p>To display a track in the Alignment and/or Strategy views:</p> <ol style="list-style-type: none"> <li>1. Open the Tracks panel by clicking on the <b>Tracks</b> tab ( Tracks). If the tab is not visible, reveal it using Choose <b>View &gt; Tracks &gt; Tracks</b> or <b>View &gt; Reset Panel Layout</b>.</li> </ol> <p>The uppermost part of the Tracks panel lets you specify which tracks are visible, and configure their options and appearance. Not all tracks will be available for all projects. For those tracks that are available, some will be displayed and other will be hidden, by default.</p>
	

	<ol style="list-style-type: none"> <li>Put check marks next to available tracks that you wish to display, and remove check marks next to tracks that you wish to hide. Tracks are added to the views in the same order in which their boxes are checked in the Tracks panel. This means that the position of a track may change when its box is unchecked and then checked again.</li> <li>In the Alignment or Strategy view(s) click on the plus sign next to the track name to reveal it.</li> </ol>
<b>To expand/collapse selected tracks</b>	Click the plus/minus signs (  /  ) to the left of the sequence name. In the <a href="#">Alignment view</a> , you can also right-click on a track and choose <b>Expand</b> or <b>Collapse</b> .
<b>To expand/collapse all tracks in the Alignment view</b>	Right-click anywhere in the <a href="#">Alignment view</a> and choose <b>Expand All</b> or “Collapse All”.
<b>To hide tracks</b>	<p>Do one of the following:</p> <ul style="list-style-type: none"> <li><u>To hide a specific track from all views</u>, uncheck the box next to the name of the track in the Tracks panel (see top row of this table).</li> <li><u>To hide a specific track in one view only</u>, right-click on the track within that view and choose <b>Hide Track</b>.</li> <li><u>To remove a selected track as if unchecking its box in the Tracks panel</u>, use <b>Edit &gt; Delete</b> or press the <b>Delete</b> key.</li> </ul>
<b>To change the order of tracks in a view</b>	Open a sequence row to view its track. Then drag a track label (e.g. “Features”) to the desired position among the other tracks and drop it there. Note that if you turn off some tracks in the Tracks panel and add them again later, they will appear in the order in which their boxes were checked.
<b>To change options for a track</b>	See <a href="#">Edit track options</a> .
<b>To change track height or spacing between tracks</b>	See <a href="#">Change track height or spacing between tracks</a> .
<b>To manually rename a single selected track</b>	Select a track in the <a href="#">Alignment</a> or <a href="#">Strategy</a> view, then right-click on it and choose <b>Rename</b> . Alternatively, select the track name and choose <b>Edit &gt; Rename</b> . The Choose Label dialog opens with the <b>Display</b> drop-down menu set to <b>Custom label</b> .



**Choose Label**

What label would you like to display in this application for the selected track?

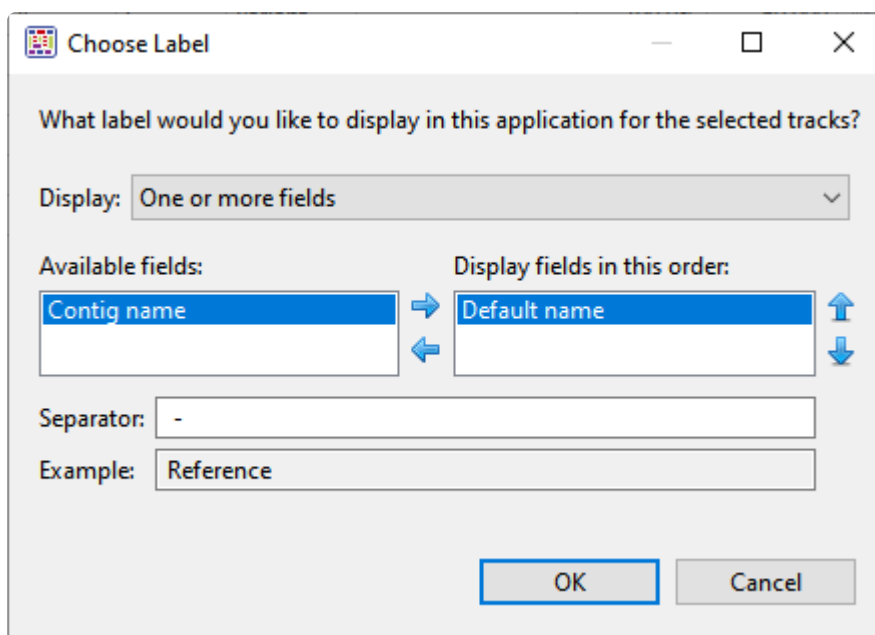
Display: Custom label

Label: Coverage

OK Cancel

Type the desired name into the **Label** text box, then click **OK**.

Select one or more tracks in the [Alignment](#) or [Strategy](#) view, then right-click on it and choose **Rename with Fields**. Alternatively, select the track name and choose **Edit > Rename with Fields**. Making no selection prior to using the main menu command is equivalent to selecting all tracks. The Choose Label dialog opens with the **Display** drop-down menu set to **One or more fields**.



**Choose Label**

What label would you like to display in this application for the selected tracks?

Display: One or more fields

Available fields: Contig name

Display fields in this order: Default name

Separator: -

Example: Reference

OK Cancel

To  
automatically  
rename  
tracks using  
specified  
data fields

The left pane shows available fields, and the right pane shows fields that will be displayed as part of the sequence name. Initially, **Default name** is the only field listed in the right pane. An example showing the appearance of the name using the specified field(s) appears in the **Example** box at the bottom of the dialog. Select and organize the fields you wish to display as part of the sequence name:

- To add or remove items, move them between panes using the left and right arrows (←→). Use the up and down arrows on the right (↑↓) to specify the order in which displayed fields should appear.
- By default, a hyphen (-) is included between each of the displayed fields. If desired,

type in a different **Separator**. Or remove all text in the **Separator** box to include nothing between the fields.

- The **Example** text box shows what the label will look like with the current selections. This text can be manually edited, if desired.

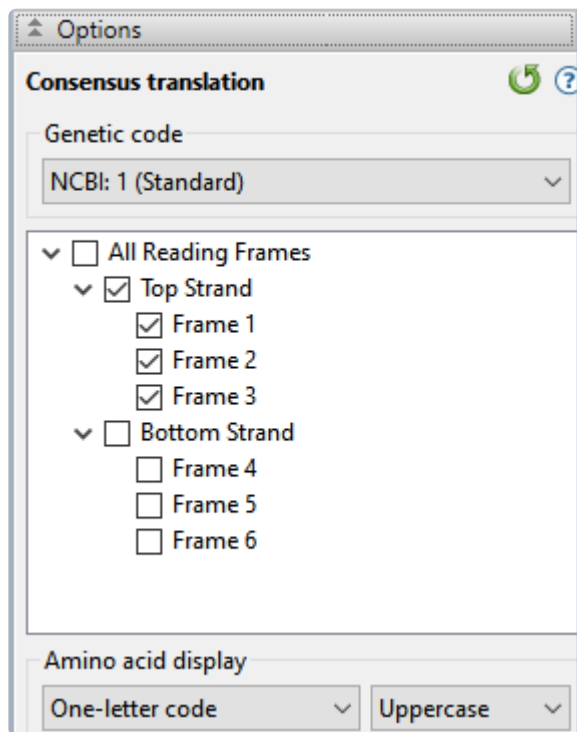
When you are satisfied with the choices made in this dialog, click **OK**

# Edit track options


To change options for a data track:

Currently, only the **Consensus translation** [track](#) has user-specifiable options. To access these, right-click on a consensus translation track in the view and choose **Show Track Options**; or click on the name

**Consensus translation** in the Tracks panel, then open the Options bar ().

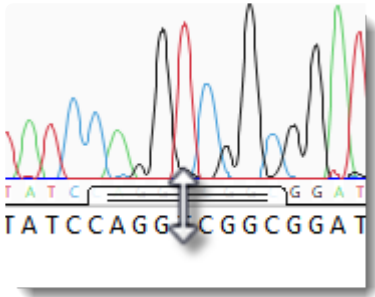


- Use the upper drop-down menu to choose from a variety of [NCBI](#) and Lasergene translation codes.
- Put checkmarks next to the reading frames you wish to display, and remove checkmarks from those you do not wish to display. Click **All Reading Frames** to show or hide all reading frames. Click **Top Strand** or **Bottom Strand** to show or hide all the reading frames on that strand.
- Use the left-hand drop-down menu to choose between **One-letter** or **Three-letter** amino acid abbreviations.
- Use the right-hand drop-down menu to choose between **Uppercase** and **Lowercase** letters.

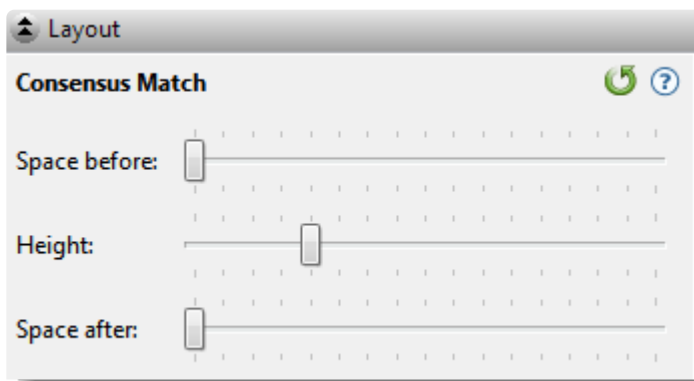
Click  if you wish to return to the default values. |

# Change track height or spacing between tracks

To increase/decrease the vertical height of a visible [data track](#) in a view, hover the mouse over the bottom of the track in either view, then grab the resize bar and drag it up or down.



Alternatively, to change both track height and the amount of space above and below the track, open the [Tracks panel](#) and select a track from the list by clicking on its name. Next, open the “Layout” section by clicking on the expand bar or by choosing **View > Tracks > Layout**.



- To control the amount of blank space above and/or below each of the selected track(s), use the **Space before/after** sliders.
- To control the heights of the selected tracks, use the **Height** slider. This option is enabled only when all selected tracks support user-scalable height.

If you want to return the sliders in this dialog to their default settings+, click the **Restore default settings** tool (🔄). |

# Features

SeqMan Ultra lets you view features in text, graphical and tabular displays.

## View feature information in graphical format:

You can use the [Tracks panel](#) to apply **Features** tracks to the [Strategy](#) and/or [Alignment](#) views. In this type of display, the feature type and name are displayed above each feature.

Hover over a feature to see a tooltip with detailed information about that feature.

The screenshot shows the SeqMan Ultra interface with a track view of genomic features. A tooltip is displayed over a tRNA feature, providing detailed information:

<b>tRNA: TRNR tRNA-Arg</b>	
tRNA 10919>10983	
<b>product:</b>	tRNA-Arg
<b>gene:</b>	TRNR
<b>gene_synonym:</b>	MTTR
<b>db_xref:</b>	GeneID:4573
<b>db_xref:</b>	HGNC:HGNC:7496
<b>db_xref:</b>	MIM:590005
<b>note:</b>	NAR: 0197
<b>anticodon:</b>	(pos:10435..10437,aa:Arg,seq:tcg)
<b>codon_recognized:</b>	CGA
<b>parentID:</b>	TRNR

An asterisk next to a feature name indicates that the feature was inferred by SeqMan NGen. For example, if you assembled data in SeqMan NGen using a template that contained a tRNA feature without a corresponding gene feature then “gene” would be inferred at that position and would be displayed as a feature in the view. However, the inferred gene would be marked with an asterisk.

## View feature information in text format:

The same information in the tooltip described above can also be viewed in the [Details panel](#). [Select a feature](#) in any view and choose **View > Details**.

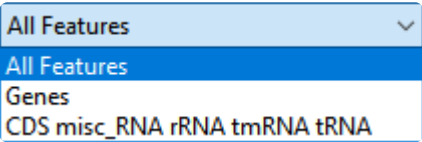


## View feature information in tabular format:

The Features view provides a table that allows you to view and manage all of the features within a selected contig. Checking and unchecking features in the view partially determines which features can be seen in the Alignment and Strategy views. (Feature display is also dependent on which [tracks](#) are visible).

To access the Features view, choose one or more contigs from the [Explorer panel](#). Then either choose **View > Features > (Contig Name)** or right-click on the selected contig(s) and choose **Show Feature Table**.

Show	Type	Name	Location	Sequence	Target Length	Coverage Depth	Coverage Strand Depth
<input type="checkbox"/>	source	Homo sapiens	1>156040895	NC_000023	~D	0.66	<na>
<input checked="" type="checkbox"/>	assembly_gap	10000	1>10000	NC_000023	~D	0	<na>
<input checked="" type="checkbox"/>	misc_RNA	PLCXD1	276324>276394,281482>...	NC_000023	2088	8.65	5.67
<input checked="" type="checkbox"/>	gene	PLCXD1	276324>303356	NC_000023	27033	1.56	1.04
<input checked="" type="checkbox"/>	mRNA	PLCXD1 [1]	276331>276394,284167>...	NC_000023	1168	8.30	5.51
<input checked="" type="checkbox"/>	mRNA	PLCXD1 [2]	276366>276394,281482>...	NC_000023	1336	7.78	5.13
<input checked="" type="checkbox"/>	mRNA	PLCXD1 [3]	281194>281256,281482>...	NC_000023	1370	7.66	5.05

If you select multiple contigs, a separate Features view will open for each of them. If you select “Unlocated Contigs,” individual tables will open for each member of the group. You may wish to avoid selecting “Unlocated Contigs” if you are working with a large number of such contigs, as with many transcriptome or multi-chromosome genome assemblies.

Task	How To
Choose which features appear in the Features view	<p>Use the drop-down menu on the left of the Features view header to show increasingly specific tables of features: <b>All Features</b>, <b>Genes</b> (only) and the subset of feature types chosen on the Input Reference Sequences page of the SeqMan NGen wizard (e.g. <b>CDSs</b> and <b>ncRNAs</b> only).</p> 
Choose which features appear in the Strategy view	<p>To choose which features should be shown or hidden in the Strategy view when Feature tracks are displayed, add or remove checkmarks from the features in the Features view. As a shortcut, you can use any of the commands in the <b>Features</b> menu. For example, to show all CDS features, you could select any CDS feature in the Features view, then choose <b>Features &gt; Show &gt; Show Only of Type CDS</b>.</p>
Add, rearrange, rename, or remove table columns	<p>Press the <b>Change options in this view</b> tool (  ) in the top right of the view. For instructions and a description of each column type, see <a href="#">Customize tables in the views</a>.</p>
Export data from the table	<p>Press the <b>Export data</b> tool (  ) in the top right of the view. This opens a save dialog in which you can save the tabular data in comma-separated (.csv) or tab-separated (.tab) format.</p>
View features for constituent sequences	<p>Display the Features <a href="#">track</a> in the <a href="#">Alignment</a> or <a href="#">Strategy</a> views. If you are unable to see an expected feature, it may have been on part of the sequence that was trimmed as vector.</p>
View features on the consensus	<p>Display the Features <a href="#">track</a> in the <a href="#">Strategy</a> view.</p>

<b>sequence</b>	
<b>Simultaneously select a feature in multiple views.</b>	When you select a feature in the Features view, the range of the feature will also be selected in the <a href="#">Alignment view</a> , for features on constituent sequences; or the <a href="#">Strategy view</a> , for features on the consensus sequence.

# Variants

The Variant view provides a table with a row for each variant position. This view is only available for reference-guided assemblies in *.assembly* format.

To access the Variant view, choose a contig from the [Explorer panel](#). Then either choose **View > Variants > (Contig Name)** or **Variants > Show Variant Table** or right-click on the selected contig and choose **Show Variant Table**. In later versions of SeqMan Ultra, you may be able to choose different versions of the view to display using the drop-down menu on the upper left of the view. Currently only available version is **SNPs Summary**.

SNP	Sample	Ref ID	Cons Pos	Ref Pos	Type	Ref Base	Called...	Genotype	Impact	Homopolymer	SNP %	Feature Type	Feature
?	6	FJ755791.1	56686	56686	SNP	C	A C	Variant	Non-synonymous		33.3%	CDS	SOS1B
?	6	FJ755791.1	56691	56691	SNP	C	G C	Variant	Non-synonymous		33.3%	CDS	SOS1B
?	6	FJ755791.1	56695	56695	SNP	G	C G	Variant	Non-synonymous		33.3%	CDS	SOS1B
?	6	FJ755791.1	56696	56695	Ins	-	C	Variant	Frameshift		50.0%	CDS	SOS1B
?	6	FJ755791.1	56697	56696	SNP	C	T C	Variant	Synonymous		33.3%	CDS	SOS1B
▶ ?	6	FJ755791.1	56699	56698	SNP [2]	AG	WK   AG	Variant	Non-synonymous		33.3%	CDS	SOS1B
?	6	FJ755791.1	56726	56724	Ins	-	C	Variant			50.0%	gene	SOS1B
?	6	FJ755791.1	56733	56730	Ins	-	A	Variant			50.0%	gene	SOS1B
▶ ?	2	FJ755791.1	56735	56732	SNP [2]	AA	WW	Variant			50.0%	gene	SOS1B
?	8	FJ755791.1	56741	56738	SNP	T	G	Variant			100.0%	gene	SOS1B
?	8	FJ755791.1	56748	56745	SNP	T	C	Variant			100.0%	gene	SOS1B

Each row in the **SNPs Summary** table summarizes information for all of the variant bases in an aligned column. If you are viewing a SeqMan NGen assembly, only variants meeting the SNP filter stringency (**High**, **Medium** or **Low**) that you specified in the SeqMan NGen wizard are displayed.

Task	How To
Open the Alignment view at the variant position	Double-click on any row in the table. This is especially useful in situations such as an amplicon assembled against a much longer reference. In this case, the <a href="#">Alignment view</a> may appear virtually empty when viewed on its own. However, double-clicking on a row in this table will open the Alignment view at a populated position.
Select a category of variants	Use the commands in the <b>Variants</b> menu to select a large group of variants at once. Regardless of the current selection, you can select <b>Variants &gt; All</b> or <b>Variants &gt; None</b> . If you have already made a selection in the table, the uppermost menu command will allow you to select all variants of the same type (e.g. <b>Variants &gt; Select &gt; Select All of Type 'CDS'</b> ). To select everything that is not currently selected, use <b>Variants &gt; Invert Selection</b> . Alternatively, you can use the right-click menu options of the same names.
Change the status of selected variants	<p>The leftmost column of the table (represented by a small triangle in the header) denotes the status of a variant. Initially, all variants are marked as putative.</p> <ul style="list-style-type: none"> <li>• ? – Putative (the initial status for all variants in the table)</li> </ul>

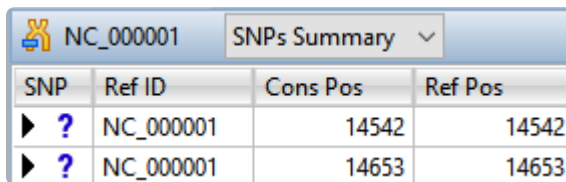
- ✓ – Confirmed

- ✗ – Rejected

To change the status of variants, select one or more rows, then right-click on the leftmost column and choose **Confirm (Reject, Question) Selected Variants**.

Alternatively, you can make the selection, then use the **Variant** menu commands of the same names.

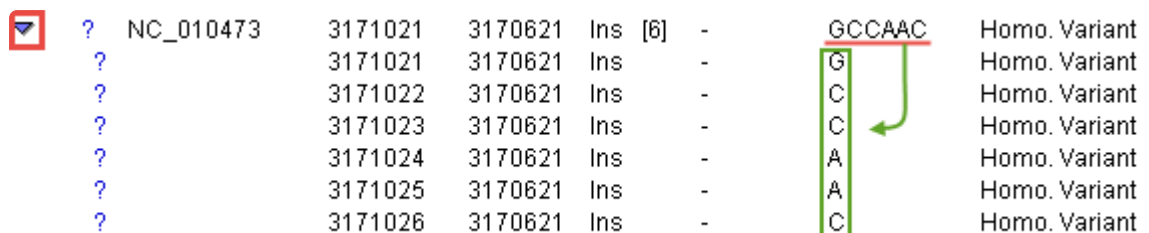
If there were multiple reference bases at the location of a variant, the leftmost column will feature a small triangle in that row. Click on the triangle to show the multiple reference bases.



SNP	Ref ID	Cons Pos	Ref Pos
▶ ?	NC_000001	14542	14542
▶ ?	NC_000001	14653	14653

#### Hide/show multiple references bases

In *.assembly* projects, variants in adjacent columns are coalesced into a single insertion or deletion if they are of the same type, and if at least 80% of the reads with the called variant in one column have a variant in the adjacent column. Each coalesced multiple-base insertion or deletion can be opened to reveal individual variants by clicking the corresponding triangle to the left of the SNP column. After clicking a triangle, information on each position of the insertion/deletion is displayed in a separate row.



▼ ?	NC_010473	3171021	3170621	Ins [6]	-	GCCAAC	Homo. Variant
?		3171021	3170621	Ins	-	G	Homo. Variant
?		3171022	3170621	Ins	-	C	Homo. Variant
?		3171023	3170621	Ins	-	C	Homo. Variant
?		3171024	3170621	Ins	-	A	Homo. Variant
?		3171025	3170621	Ins	-	A	Homo. Variant
?		3171026	3170621	Ins	-	C	Homo. Variant

#### Choose which items should appear in the view

Press the **Filter** (🔍) tool in the top right of the view. See [Filtering in the Variants view](#) for detailed information.

#### Add, rearrange, rename, or remove table columns




Press the **Change options in this view** tool (⚙️) in the top right of the view. For instructions and a description of each column type, see [Customize tables in the views](#).

#### Show counts as a percent


Press the **Show counts as percent** tool (📊) in the top right of the view. This affects columns named **Cnt**.

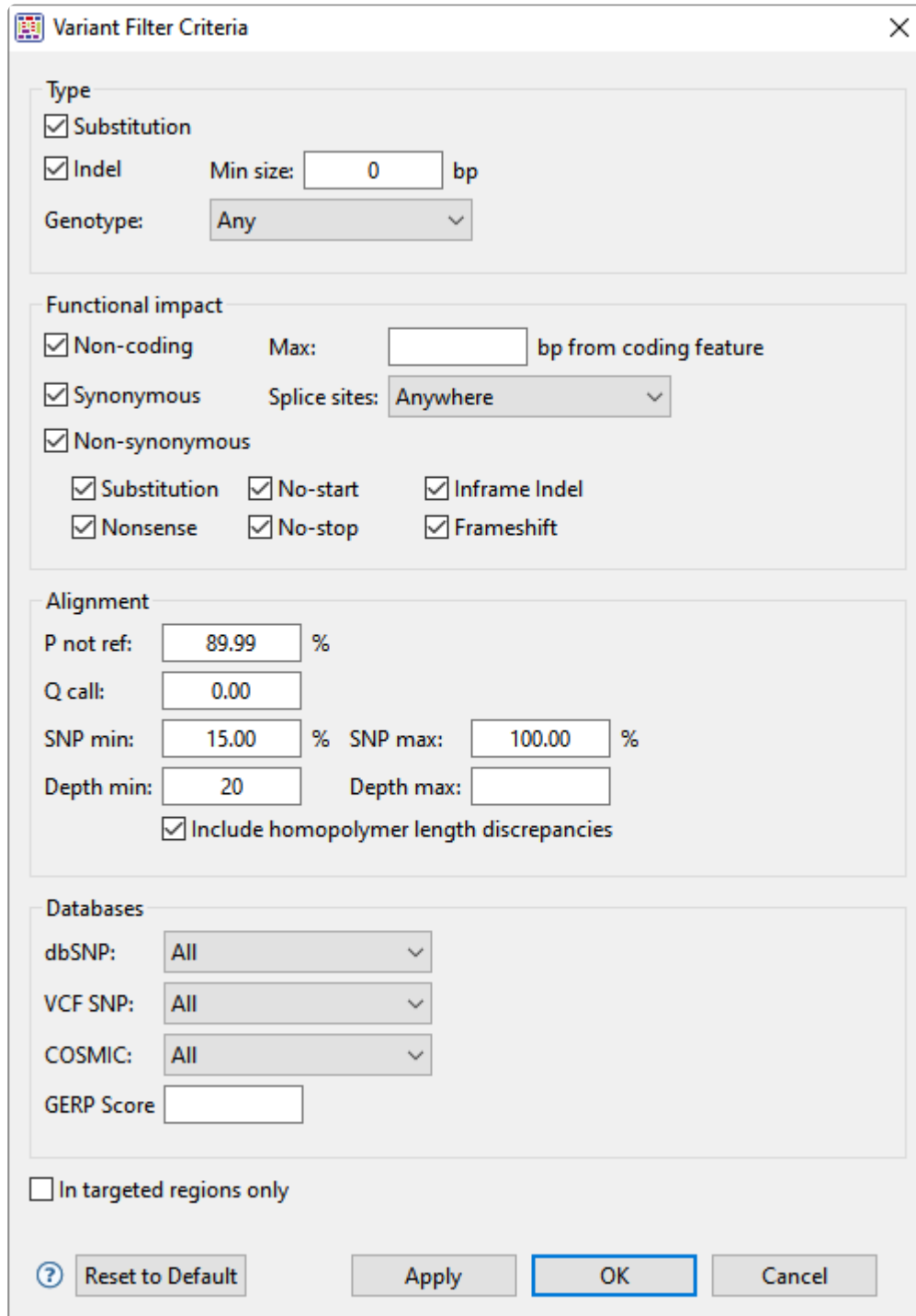
#### Export data from

Press the **Export data** tool (📄) in the top right of the view. This opens a save dialog in which you can save the tabular data in comma-separated (.csv) or tab-separated (.tab) format.

the table	
Toggle between showing/hiding the table header	<p>Use the <b>Show quick details</b> tool () in the top right of the view. An example of the table header appears below boxed in red:</p> <div><div> NC_000004</div><div>SNPs Summary </div></div> <div>0 confirmed 0 rejected 76 putative 0 mixed 15,126 filtered</div>


# Filtering in the Variants view

From the [Variants view](#), you can open a filter dialog to specify which variants are shown in the table. To open the Variant Filter Criteria dialog, use the **Variants > Filter** command, press the **Filter** tool () in the upper-right of the view, or right-click anywhere in the view and choose **Filter**.



The Variant Filter Criteria dialog box is used to specify which variants are shown in the table. It contains several sections for filtering criteria:

- Type**
  - ☒ Substitution
  - ☒ Indel Min size:  bp
  - Genotype:
- Functional impact**
  - ☒ Non-coding Max:  bp from coding feature
  - ☒ Synonymous Splice sites:
  - ☒ Non-synonymous
    - ☒ Substitution ☒ No-start ☒ Inframe Indel
    - ☒ Nonsense ☒ No-stop ☒ Frameshift
- Alignment**
  - P not ref:  %
  - Q call:
  - SNP min:  % SNP max:  %
  - Depth min:  Depth max:
  - ☒ Include homopolymer length discrepancies
- Databases**
  - dbSNP:
  - VCF SNP:
  - COSMIC:
  - GERP Score
- ☐ In targeted regions only

Buttons at the bottom: 

In the Type section:

Choose to display **Substitutions** and/or **Indels** by checking the boxes next to their name(s). If you want to see only variants greater than a certain length threshold, type that threshold into the **Min size** box.

Use the **Genotype** drop-down menu to limit the display to **Annotated SNPs only** or **Novel SNPs only**. To display both types, keep the default setting of **Any**.

In the Functional impact section:

Choose whether to display indels having the following functional impacts:

- **Non-coding** – The indel appears in a non-coding region of the sequence, and will therefore have no functional impact.
- **Synonymous** – The indel does not cause an amino acid change, and will therefore have no functional impact.
- **Non Synonymous** – Amino acid substitution only.
  - **Substitution** – Any change from one nucleotide to another, regardless of effect or lack thereof.
  - **Nonsense** – Amino acid to translational stop.
  - **No Start** – A change that disrupts the start codon.
  - **No Stop** – A change that converts a stop codon to an amino acid, and thereby extends the reading frame.
  - **Inframe Indel** – An insertion or deletion within a coding region whose length is divisible by 3. The type is followed by the word Conservative if the indel occurs between two codons, and Disruptive if it occurs with a codon.
  - **Frameshift** – An indel within a coding region and which is not a multiple of 3, thereby changing the reading frame.

In the Alignment section:

- **P not ref** – The probability that the called base at this position is not the reference base. For coalesced variants, this value is equal to the minimum value of all “child” values. The minimum allowed value is 30%.
- **Q call** – The Phred-like quality score of the called genotype. It is a measure of the probability that the called genotype is correct.

- **SNP min / SNP max** – The minimum and maximum percentage of reads that should contain a SNP at a given position.
- **Depth min / Depth max** – The minimum and maximum depth of reads needed to include the SNP.
- **Include homopolymer length discrepancies** – If you are using Illumina or Sanger data, we recommend leaving this box checked. If using Ion Torrent data, we recommend unchecking the box. If you uncheck the box, SeqMan Ultra will remove all homopolymeric run length variants from the table.

In the Databases section:

- Use the **dbSNP**, **VCF SNP** and **COSMIC** drop-down menus to filter in/out subsets of variants from these databases.
- To filter out variants with a **GERP score** below a specified threshold number, enter that number in the text box.
- To filter out variants from locations other than the targeted regions (i.e., from a *.bed* or manifest file used in the assembly), check the box next to **In targeted regions only**.

If you wish to return to default settings, click **Reset to Default**.

To save your settings, click **Apply** or **OK**, or simply press **Enter**.

# Exporting variant data

---

To export data from the Variants view:

1. Select a contig in the [Explorer panel](#) and open the [Variants view](#).
2. [Filter the table](#) to include the variants of interest.
3. Use the **Change alignment options in this view** tool ( ) to select data columns and place them in the desired order. If exporting for use in ArrayStar, see the section below.
4. Click anywhere in the Variants view and choose **Edit > Select All**, then **Edit > Copy**.
5. Use **Edit > Paste** to paste the data into a suitable spreadsheet program and save the file in any desired format.

Exporting variant data for use in ArrayStar:

If you plan to open the variant table in ArrayStar, you must include the following data columns in Step 3, above:

- **Contig ID**
- **Contig Pos**
- **Ref Pos**
- **Type**
- **Ref Base**
- either **SNP Base** or **Called Base**
- **SNP%**

If you wish to determine coding vs. non-coding regions or synonymous vs. non-synonymous changes in ArrayStar, you must also export the following columns:


- **Genotype**
- **Feature Name**
- **Feature Type**
- either **Codon** or **Protein Change**

# Make a Selection

Depending on the view, you can select one or more sequences, any portion of a sequence, or a specific portion of sequence corresponding to a feature. The active selection is highlighted in light blue. The selection can then be copied to the clipboard if desired.

To select...	Do this...
A contig	In the <a href="#">Explorer panel</a> , click on the name of the contig.
Multiple read sequences	In the <a href="#">Alignment</a> or <a href="#">Strategy</a> view, use <b>Ctrl/Cmd+click</b> or <b>Shift+click</b> to select multiple read sequences.
One entire sequence	In the Alignment or “Strategy view, click on the sequence name.
A portion of a constituent or consensus sequence	<p>In the Alignment or Strategy views:</p> <ul style="list-style-type: none"> <li>• Hold down the left mouse button and drag the mouse along the portion of sequence you wish to highlight. To select multiple non-contiguous portions, use <b>Ctrl/Cmd+drag</b>.</li> <li>• Click on the left-most residue you wish to highlight. Hold down the <b>Shift</b> key and click on the right-most residue.</li> <li>• To select an entire sequence, insert the cursor anywhere along the sequence, then right-click and choose <b>Select All</b>.</li> </ul>
A portion of one sequence	In the lower portion of the Alignment view, drag the mouse across residues in a sequence track while holding down the left mouse button.
A track	Click on a track.
One or more features and corresponding sequence	Click on a <a href="#">feature</a> to select it. To add additional features to the selection, use <b>Ctrl/Cmd+click</b> .

# View Information About a Selection

The Details panel shows details of the current selection or project, and is represented by a blue “information” icon (). If the panel is not already visible, reveal it by using **View > Details** or **View > Reset Panel Layout**; or right-click within a [view](#) and choose **Show Details Panel**.

Current selection	Information shown
The <a href="#">Welcome screen</a> is open	<p><b>Open an assembly</b> launches a file chooser</p> <p><b>Watch a training video</b> opens the <b>Training</b> page on the DNASTAR website.</p> <p>The <b>Recent Documents</b> section provides links for opening a recent document, if any exist.</p>
<b>A single read sequence</b>  <div> <b>Selected Read</b>  <b>NC_002973 selection:</b> ID_1530469  NC_002973: 1 read </div>	Contig name, read ID, and the number of reads selected (i.e. “1”).
<b>Multiple read sequences</b>  <div> <b>Selected Reads</b>  <b>NC_002973 selection:</b>  NC_002973: 2 reads </div>	Contig name and the number of reads selected.
<b>A portion of the consensus sequence</b>  <div> <b>Selected Region</b>  <b>NC_002973 selection:</b> 16526&gt;19276 = 2751 </div>	Contig name and the inclusive coordinates of the beginning and end of the selection
<b>A portion of a read sequence</b>  <div> <b>Selected Region</b>  <b>Sequence:</b> ID_1396536_r1_dup[40]  <b>Sequence selection:</b> 8&lt;23 = 16 </div>	Read ID and the inclusive coordinates of the beginning and end of the selection
<b>One or more tracks</b>	See <a href="#">View information about a selected track</a> and <a href="#">View information about multiple selected tracks</a> .
<b>A single feature</b>	See the section below this table.
<b>Multiple features</b>	The number of features selected and the feature type(s).

<b>2 features</b> <b>Type:</b> CDS, gene	
---	--

If a single feature has been selected in a Feature [track](#), the Details panel contains the following information:

<b>Heading</b> <b>CDS: LMOF2365_0043</b> <b>Type:</b> CDS	Feature name (in blue) and type.
<b>Location</b> <b>▼ Location</b> 44215>45573	Inclusive range of the feature in nucleotides.
<b>Properties</b> <b>▼ Properties</b> <b>product:</b> PTS system, beta-glucoside-specific, IIC component <b>locus_tag:</b> LMOF2365_0043 <b>protein_id:</b> YP_012654.1 <b>db_xref:</b> GI:46906265 <b>db_xref:</b> GeneID:2798815 <b>codon_start:</b> 1 <b>parentID:</b> LMOF2365_0043 <b>transl_table:</b> 11	Properties shown vary by feature and may include the organism, gene name NCBI feature qualifiers, etc.
<b>Notes</b> <b>▼ Notes</b> <b>note:</b> identified by similarity to SP:P46317; match to protein family HMM PF02378; match to protein family HMM TIGR00410	The NCBI feature qualifier “/note”.
<b>Translation</b> <b>▼ Translation</b> MSSERSLSSRFIEGLSIFAQKISSQKHIMAIRDGFAAMIPIT IIAAFFLLVNNVLLQPEGLLKFPNVENYLGVGIOVY SGYLATGGDIRASILQAIIVGTLIYLPFVLVANRAYVLE QKAAGKVEAETVTNGEV	The translated sequence shown in <a href="#">IUPAC amino acid codes</a> .

# View information about a selected track

If a single track has been selected in a view or panel, the [Details panel](#) contains the following information.

<p><b>Heading</b></p> <p><b>Consensus translation</b></p> <p><b>Summary:</b> The Consensus translation track displays the translation of the consensus sequence.</p>	<p>Type of track (in blue).</p> <p><b>Summary</b> – Basic information about what is displayed in the selected track.</p>
<p><b>Actions</b></p> <p>▼ Actions</p> <p><a href="#">Rename...</a></p> <p><a href="#">Rename with Fields...</a></p> <p><a href="#">Show track layout</a></p> <p><a href="#">Show track options</a></p>	<p>Only a subset of links may be available, depending on the selected track.</p> <ul style="list-style-type: none"> <li>• <b>Rename</b> – Shortcut to <b>Edit &gt; Rename</b> (See <a href="#">Work with Data Tracks</a>).</li> <li>• <b>Rename with Fields</b> – Shortcut to <b>Edit &gt; Rename with Fields</b> (See <a href="#">Work with Data Tracks</a>).</li> <li>• <b>Show track layout</b> – Shortcut to <b>View &gt; Tracks &gt; Layout</b>. (See <a href="#">Change track height or spacing between tracks</a>).</li> <li>• <b>Show track options</b> – Shortcut to <b>View &gt; Tracks &gt; Options</b>. (See <a href="#">Edit track options</a>).</li> </ul>

# View information about multiple selected tracks

---

If two or more tracks have been selected in a view or panel, the [Details panel](#) contains the following information.

<div>Heading</div> <div><div>2 tracks</div><div>Summary: Multiple tracks are selected. They can be moved or hidden together.</div></div>	<div>Number of tracks currently selected (in blue).</div> <div>Summary – Description of options available for the multiple track selection.</div>
<div>Actions</div> <div><div>Actions</div><div>Rename with Fields...</div></div>	<div>Rename with Fields – Shortcut to Edit &gt; Rename with Fields (See <a href="#">Work with Data Tracks</a>).</div>

# Search for Sequences Online

---





This application lets you search [National Center for Biotechnology Information](#) (NCBI) databases for matches to text or sequence that you specify.

- [Query part or all of a nucleotide or protein sequence](#) against one of NCBI's sequence databases using the [Basic Alignment Search Tool \(BLAST\)](#).
- [Query text or accession numbers](#) against entries NCBI's [Entrez](#) database.

Once a search has been initiated, you can use the [Jobs panel](#) to monitor the search, stop an in-progress search, or open search results in the [Text](#) and [Table](#) views.

## Tools available in the Search wizards:

Some or all of the following optional tools may be available in the bottom left corner of the Search wizards:

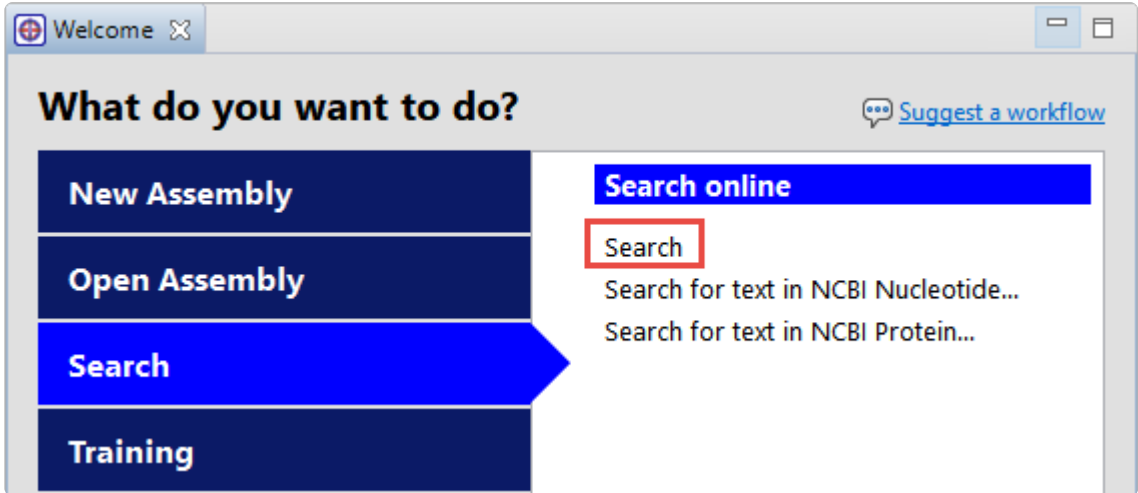

1. To display the help for this topic, press the question mark icon (.
2. To enter an NCBI Access ID, press the **Set Access ID** () tool. This causes the Access ID dialog to pop up. See [Access ID dialog](#) for information.
3. To submit an idea for a new feature, press the **Suggest a feature** () tool.
4. To view information about your service plan, press the **Show Service Plan** () tool. For sequence or text search services, the ensuing dialog describes the state of your service plan and any warning or error text messages. The license control has three states: **Active**, **Expiring**, and **Expired**.

# Search for sequence matches (BLAST)

The BLAST Search wizard is used to [search](#) a query sequence against one of NCBI's databases. The sequence can be part or all of an active sequence, another sequence file on your computer, or typed/pasted-in text. The wizard consists of four consecutive screens, each described in a separate topic:

- [Query screen](#)
- [Search screen](#)
- [Options screen](#)
- [Job screen](#)

To launch the BLAST Search wizard, use one of the following commands or tools. Some commands let you launch the wizard with the workflow pre-selected and the Search list already pre-populated.

Command, shortcut or tool	Description
Press <b>Search</b> on the left of the Welcome screen, then choose <b>Search</b> on the right  	Launches the wizard at the <a href="#">Query screen</a> .
<b>Search &gt; Sequence Search</b>  Use the “Change search strategy” tool (  ) from the <a href="#">Text</a> , <a href="#">Table</a> or <a href="#">Pairwise</a> views	Launches the wizard at the <a href="#">Query screen</a> .
<b>Search &gt; Sequence Search with Selection</b>  Press <b>Ctrl/Cmd+B</b>	Launches the wizard at the <a href="#">Query screen</a> , whose table is pre-populated with the currently-selected portion of sequence.
<b>Search &gt; Sequence Search with All Sequences</b>	Launches the wizard at the <a href="#">Query screen</a> .

	whose table is pre-populated with all sequences in the document.
--	--

Page 111 of 178

**Enter**

**Enter sequences in FASTA format**

```
>Hemoglobin (Alpha Chain)
VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTK
TYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNAL
SALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTISKYR
```

**Range**

Start:  End:

☐ Add another

OK Cancel

In the Enter dialog:

1. Type or paste in the desired sequence following the “dimmed” example in the dialog. A header line is optional.
2. (optional) If you wish to use only a range of this sequence, specify the **Start** and **End** positions.
3. (optional) If you want to add another query sequence after this one, check the **Add another** box. After clicking **OK** to the first sequence, a new Enter dialog will open to add the next query. Otherwise, leave this box unchecked.
4. Press **OK**. If you add additional sequences, each sequence will appear on a separate row in the table.

To remove a file that has already been added:

Select the unwanted file from the list and click **Remove**.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of [this topic](#).

Once you have set up the query sequence(s) as desired, do either of the following:

- To choose the BLAST program to use and database to search, press **Next** to continue to the [Search screen](#).

- To run the search directly, press **Run**.
- To close the Search wizard without starting a search, press **Cancel**.

# Search wizard: Search

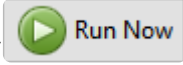
The Search screen is the second of four screens in the [BLAST Search wizard](#). The Search screen is used to choose the BLAST program and database.

- Specify whether you wish to search for **Nucleotides** or **Proteins**.
- Select the desired BLAST program from the **Using** drop-down menu. For information on the different BLAST programs, see the [BLAST Program Selection Guide PDF](#). Available options depend on the combination of the selections you made in **Search with** ([Query screen](#)) and **Search for** (this screen).
  - If you selected **Proteins** on both screens, the **Using** menu contains the unchangeable default of **Similar proteins (blastp)**.
  - If you selected **Nucleotides** on both screens, Choices are **Highly similar sequences (megablast)**, **Similar nucleotides (blastn)**, and **Similar translated nucleotides (tblastx)**.
  - If you selected **Nucleotides** in the Query screen and Proteins in the Search screen, the **Using** menu contains the unchangeable default of **Similar proteins (blastx)**.
  - If you selected **Proteins** in the Query screen and **Nucleotides** in the Search screen, the **Using** menu contains the unchangeable default of **Similar translated nucleotides (tblastn)**.
- Select the desired BLAST database from the **In** drop-down menu. Available choices depend on

whether you selected **Nucleotides** or **Proteins** in this screen.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of [this topic](#).

Once you have made the desired selections, do any of the following:

- To go back to the [Query screen](#), press **Back**.
- To add additional parameters to the search, press **Next** to continue to the [Options screen](#).
- To start the search immediately, press the **Run Now** button () or **Run** button. The [Jobs panel](#) in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings...** button. Use the [Jobs panel](#) to monitor an [online search](#), stop an in-progress search, or open search results in the [Text](#), [Table](#) and [Pairwise](#) views.
- To close the Search wizard without starting a search, press **Cancel**.

# Search wizard: Options

The Options screen is the third of four screens in the [BLAST Search wizard](#), and has two tabs: **Selection** and **Matches**. These tabs allow you to customize miscellaneous parameters relating to the BLAST search.

- Click the **Selection** tab to make it active.

In the **Selection** tab:

- To filter the database you're searching using a text term, enter a text string in the **Entrez** query field.
- To filter low-complexity regions in the query, check the box next to **Ignore low complexity in query sequence**. The **blastn** search program will use **DUST** for filtering; all other programs will use [SEG](#).
- To mask the query while producing seeds used to scan database (but not for extensions), check the box by **Ignore low complexity during database lookup**.
- To mask any letters that were lower-case in the FASTA input, check the box next to **Ignore lower case letters**.


- Click the **Matches** tab to make it active.

In the **Matches** tab:

- Enter a number next to **Maximum sequences**, the maximum number of aligned sequences to display. The actual number of alignments may be greater than this number. The default is **100**.
- Enter a number beside **Expect threshold**, the expected number of chance matches in a random model. The default is **10**.
- Enter a number next to **Maximum regional matches** to limit the number of matches to a query range. The default is **0**.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of [this topic](#).

Once you have made the desired selections, do any of the following:

- To continue filtering, press **Next** to continue to the [Job screen](#).
- To start the search immediately, press the **Run Now** button (  ) or **Run** button. The [Jobs panel](#) in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings...** button. Use the [Jobs panel](#) to monitor an [online search](#), stop an in-progress search, or open search results in the [Text](#),

[Table](#) and [Pairwise](#) views.

- To close the Search wizard without starting a search, press **Cancel**.

# Search wizard: Job

The Job screen is the fourth of four screens in the [BLAST Search wizard](#). This screen allows you to overwrite the default search name with a custom name and initiate the BLAST search.

License Agreement'. At the bottom, there are three icons (a question mark, a speech bubble, and a key) on the left, and three buttons: '< Back', 'Run', and 'Cancel'."/>

The text box contains an automatically-generated name for the search job. If you wish, you can overwrite this default name with a custom name.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of [this topic](#).

After entering a custom job name (or keeping the default name), do either of the following:

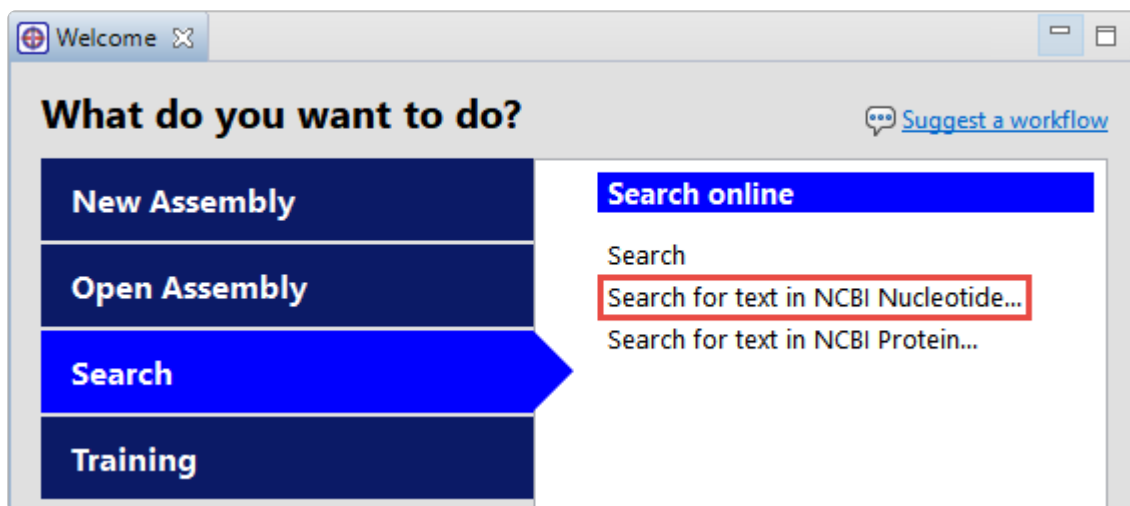
- To start the search, press the **Run** button. The [Jobs panel](#) in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings...** button. Use the [Jobs panel](#) to monitor an [online search](#), stop an in-progress search, or open search results in the [Text](#), [Table](#) and [Pairwise](#) views.
- To close the Search wizard without starting a search, press **Cancel**.

# Search for text or accession number matches (Entrez)

The Entrez Search wizard is used to [search](#) a text query or accession number against NCBI's Entrez databases for nucleotides or proteins.

To launch the Entrez search dialog, do any of the following:

- To initiate a search of the nucleotide database, use **Search > Text Search in > NCBI Nucleotide** or click on the **Welcome Tab**, choose **Search** on the left and then **Search for Text in NCBI Nucleotide** on the right.

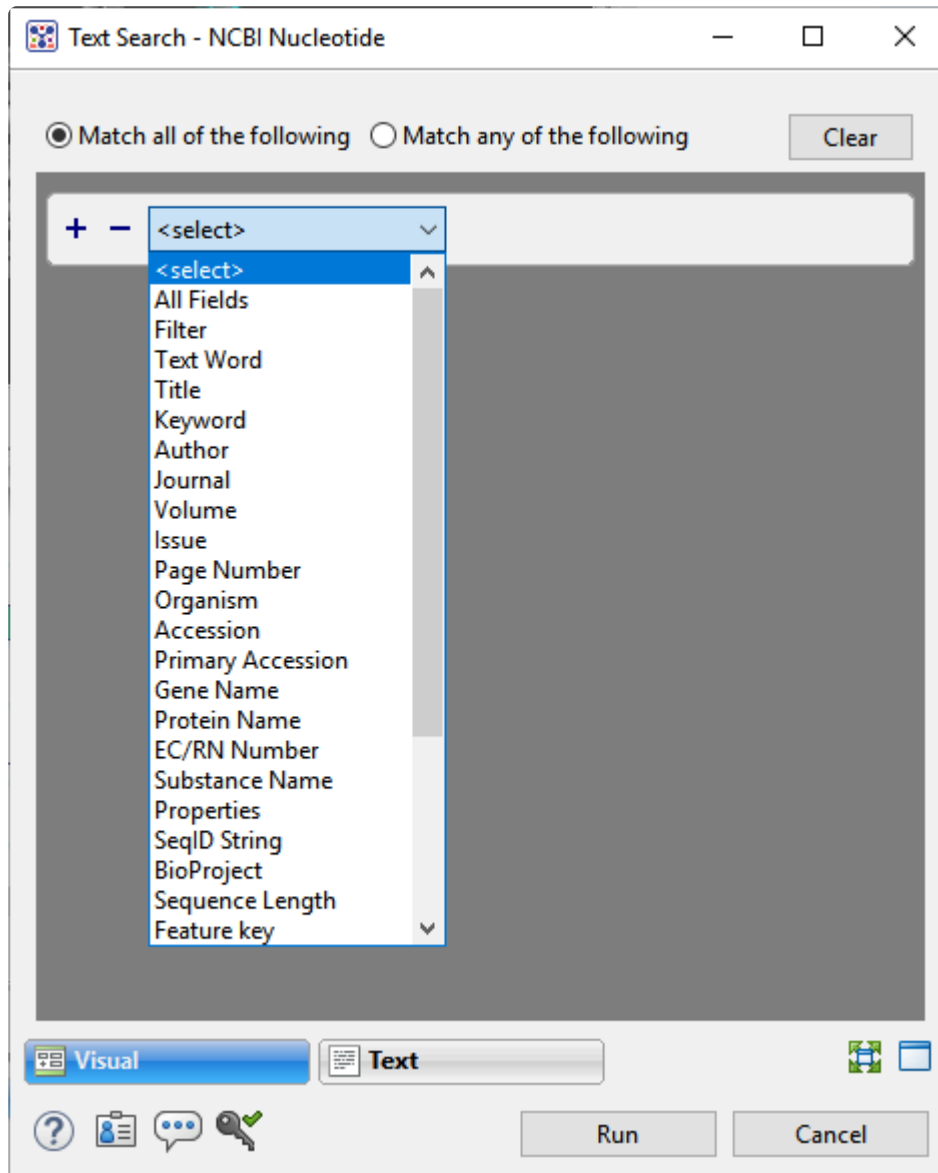


- To initiate a search of the protein database, use **Search > Text Search in > NCBI Protein** or click on the **Welcome Tab**, choose **Search** on the left and then **Search for Text in NCBI Protein** on the right.

The **Visual** view is displayed by default and is described below.

\* **Note:** The Visual view is only for non-nested queries using the same type of operator (AND, OR). To perform a more complex search, press the **Text** tab near the bottom left corner and create the search in the [Text view](#).

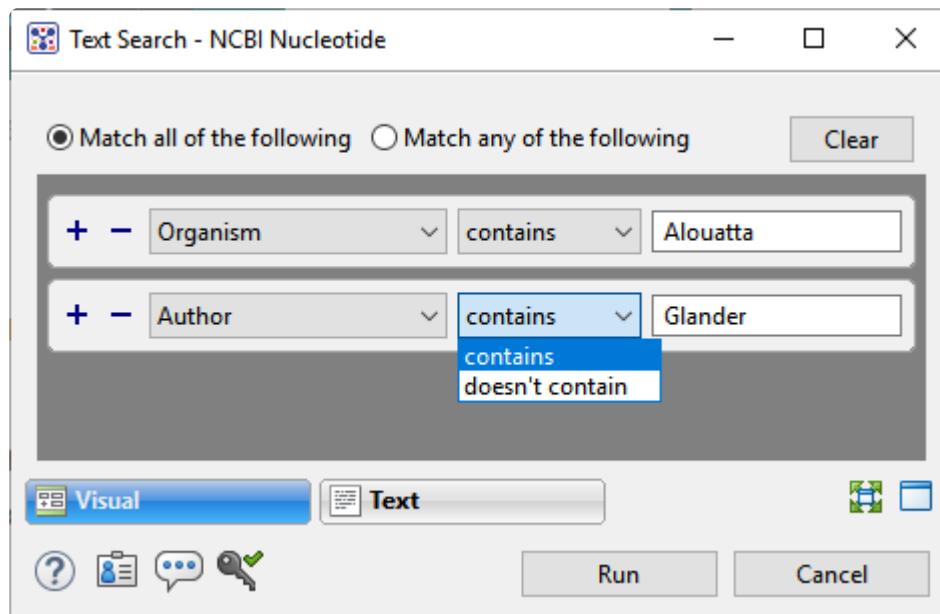
- Use the **<select>** drop-down menu to choose a criterion. The field headings vary depending upon the online database that you are using. Choosing **All fields** will cause the application to search every word of every listing for the query term in the database.



Once you make a selection, a second drop-down menu and an empty text box appear to the right.

- In the second drop-down menu on the row, choose **contains** or **doesn't contain** (see image below).
- In the text box, type the text that should be found in any results (in the case of **contains**) or should not be found in any results (in the case of **doesn't contain**).

2. To add another criterion, press the plus icon (+). To delete a criterion, press the minus icon (-).



3. Once you have set up all desired criteria, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** criteria.
4. To learn about the optional tools in the bottom left corner of the screen, see the bottom of [this topic](#).
5. Click **Run** to send the query to the database. To instead exit the dialog without searching, press **Cancel**.

During the search, a progress dialog is displayed. Once the search is completed, any matches are displayed in the [Jobs panel](#).

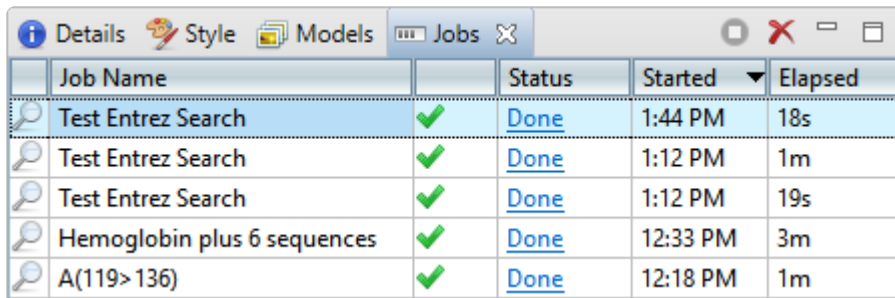
# Jobs panel

After initiating an NCBI [BLAST](#) or [Entrez](#) search, you can monitor its progress and download its eventual output from the Jobs panel. This panel is not visible in a project until you have started running at least one search.

To open the Jobs panel:






Do any of the following:

- Click on the **Jobs** tab.
- Press **Ctrl+Alt+J** (Win) or **Option+Cmd+J** (Mac).
- Use **View > Jobs** or **Search > Show Searches**.





Job Name	Status	Started	Elapsed
Test Entrez Search	Done	1:44 PM	18s
Test Entrez Search	Done	1:12 PM	1m
Test Entrez Search	Done	1:12 PM	19s
Hemoglobin plus 6 sequences	Done	12:33 PM	3m
A(119>136)	Done	12:18 PM	1m
















The Jobs panel consists of a header and a table with the following columns:

Column	Description
<b>Job Name</b>	<a href="#">Text searches</a> are all named with the field and term of the first search criterion followed by “plus # terms,” where # represents the number of additional criteria. <a href="#">Sequence searches</a> use the name specified in the <a href="#">Job screen</a> of the Search wizard. The same <b>Job Name</b> can be used for more than one prediction.
<b>Status</b>	<p>A word and icon indicating the status of the search.</p> <ul style="list-style-type: none"> <li>• <b>Started – Run or Run Job</b> were pressed in the Search wizard</li> <li>• <b>Queued</b> –  – Submitted and placed into the queue</li> <li>• <b>Running</b> –  – Running</li> <li>• <b>Done</b> –  – Job complete; results available</li> <li>• <b>Failed</b> –  – Job failed</li> <li>• <b>Stopped</b> –  – Job stopped by user</li> </ul>
<b>Started</b>	The date and time the job began.
<b>Elapsed</b>	The total elapsed time between starting and completing the job.

Tasks that pertain to the Jobs panel:

Task	How To
To open the results of one or more jobs	Click on the <b>Done</b> ( <a href="#">Done</a> ) or <b>Expiring</b> links for that job. See below this table for more information on job expiration. You can also open a saved .search file using <b>Search &gt; Open Search File</b> .
To stop one or more jobs that are in progress	Select one or more rows with a Status of <b>Started</b> , <b>Queued</b> or <b>Running</b> . Then press the <b>Stop</b> tool (  ) in the header above the table. Once stopped, a job cannot be restarted from the Jobs panel. You must instead <a href="#">initiate a new search</a> .
To delete a finished job	Select one or more rows with a Status of <b>Done</b> , <b>Failed</b> or <b>Stopped</b> . Then press the <b>Delete</b> tool (  ) in the header above the table.
To open/reopen the progress dialog for a job	Select one or more rows with a Status of <b>Started</b> , <b>Queued</b> or <b>Running</b> . Then right-click and choose <b>Open Progress dialog</b> .

Sequence (BLAST) and text (Entrez) search jobs are intended to be ephemeral and expire after 7 days. The Jobs panel notifies you of upcoming expiration dates.

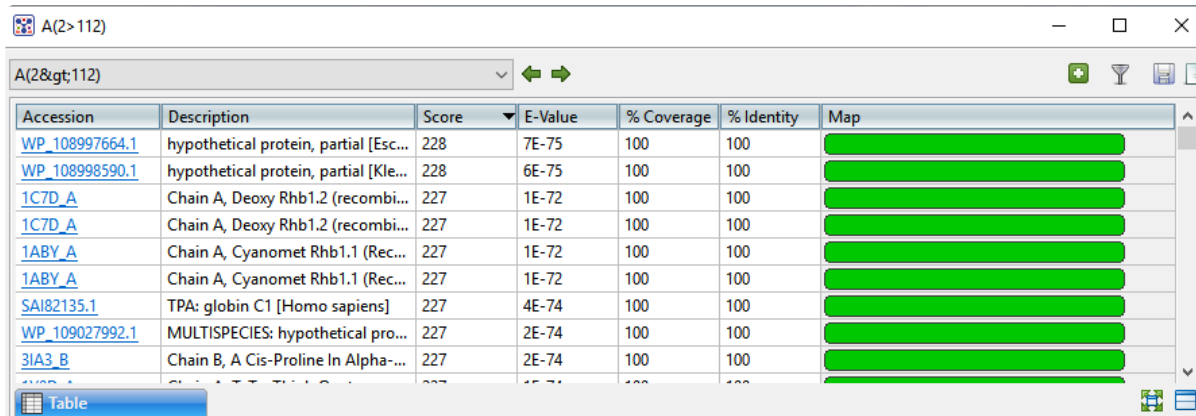
 Details  Style  Models  Jobs 		
Job Name	Status	
 Hotspot for 7TIM (2>248, 2>248)		<a href="#">Done</a>
 GEKZA8J08JS1CF plus 5 sequences		<a href="#">Expiring at 6/6/19 5:33 PM</a>
 Variant for 4HHB (1 position)		<a href="#">Done</a>
 Hemoglobin plus 6 sequences		<a href="#">Expiring at 6/6/19 12:33 PM</a>
 A(119>136)		<a href="#">Expiring at 6/6/19 12:18 PM</a>

All jobs are recorded in the location specified in the [Temporary File preferences](#) dialog. If you are a Macintosh user and elect to keep the default save location, the operating system may remove jobs after a reboot. To avoid this, we recommend choosing a different location in which to save temporary files.

# Table view

After the [Jobs panel](#) shows the **Status** of a [BLAST search](#) or [Entrez search](#) is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the [Text](#), Table, and [Pairwise](#) views. All three views are contained in a large pop-up window that is separate from the main SeqMan Ultra window.

If no significant matches were identified, the Table view will be empty. An example of a populated Table view is shown below:







Accession	Description	Score	E-Value	% Coverage	% Identity	Map
<a href="#">WP_108997664.1</a>	hypothetical protein, partial [Esc...	228	7E-75	100	100	
<a href="#">WP_108998590.1</a>	hypothetical protein, partial [Kle...	228	6E-75	100	100	
<a href="#">1C7D_A</a>	Chain A, Deoxy Rhb1.2 (recombi...	227	1E-72	100	100	
<a href="#">1C7D_A</a>	Chain A, Deoxy Rhb1.2 (recombi...	227	1E-72	100	100	
<a href="#">1ABY_A</a>	Chain A, Cyanomet Rhb1.1 (Rec...	227	1E-72	100	100	
<a href="#">1ABY_A</a>	Chain A, Cyanomet Rhb1.1 (Rec...	227	1E-72	100	100	
<a href="#">SAI82135.1</a>	TPA: globin C1 [Homo sapiens]	227	4E-74	100	100	
<a href="#">WP_109027992.1</a>	MULTISPECIES: hypothetical pro...	227	2E-74	100	100	
<a href="#">3IA3_B</a>	Chain B, A Cis-Proline In Alpha...	227	2E-74	100	100	

The right side of the header contains a [toolbar that is shared between all three views](#).

Columns in the view’s table are described below:

Column	Description
<b>Accession</b>	The NCBI “Accession” number for the match. Press the link to open the corresponding page on the NCBI website.
<b>Description</b>	The NCBI “Definition” field text.
<b>Score</b>	The similarity score for the match. (See the <a href="#">NCBI definition for raw score</a> ). In general, a higher score denotes a better match.
<b>E-Value</b>	The chance that the match is expected to occur by chance. (See the <a href="#">NCBI definition for E value</a> ). In general, a lower score denotes a better match.
<b>%Coverage</b>	The percentage of the query sequence that aligns to a match in the BLAST database. Example: Query has 596 bases, of which 565 align to a match in the NCBI database. <b>%Coverage</b> = 565/596 = 95%.
<b>%Identity</b>	The percentage of bases that match when the query and result sequence are aligned. ( <a href="#">NCBI definition</a> ). Example: Query has 565 bases that align to the match and 3 bases that don’t match. <b>%Identity</b> = 562/565 = 99.5%.
<b>Map</b>	Graphic showing areas of agreement between the query and the match.

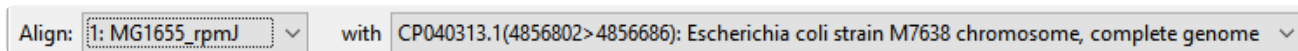
Tasks pertaining to the Table view:

Task	How To...
<b>If a BLAST or Entrez search was performed</b>	
<b>Launch the NCBI page for matching database entry</b>	Click the hyperlink in the table's Accession column.
<b>Copy all information from a search result row</b>	Right-click on the row and choose <b>Copy</b> .
<b>Save the sequence match in .gbk format</b>	Right-click on the row and choose <b>Download sequences</b> . In the ensuing Save As dialog, choose a location in which to save the sequence. By default, the sequence is named using its Accession number.
<b>Add sequences or matches</b>	Use the <b>Add sequences or matches</b> tool (  ). Sequences are added back into the host application, not the job's document.
<b>Filter results from a sequence or text search</b>	Use the <b>Filter results</b> tool (  ). Learn more about <a href="#">sequence search</a> or <a href="#">text search</a> filtering.
<b>Export data from the view</b>	Use the <b>Export data</b> tool (  ), described in the topic <a href="#">Export data</a> .
<b>Only if a BLAST search was performed</b>	
<b>To save results</b>	Click the <b>Save as</b> tool (  ). A save dialog opens and allows you to save the search results as a .search format file.
<b>To display a pairwise alignment using a particular search result</b>	If both the Table and <a href="#">Pairwise</a> views are open, click on the desired row in the Table view to see the corresponding pairwise alignment in the Pairwise view. As an alternative, open the Pairwise view manually and choose the desired target from the <b>with</b> drop-down menu.


# Pairwise view

After the [Jobs panel](#) shows the **Status** of a [BLAST search](#) or [Entrez search](#) is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the Text, [Table](#), and [Pairwise](#) views. All three views are contained in a large pop-up window that is separate from the main SeqMan Ultra window. The Pairwise view consists of three areas: toolbar, summary, and sequences.

## Toolbar area:



This area consists of two drop-down menus and one tool. The **Align** drop-down menu is used to specify the target sequence, which should be the longest sequence of the two. The \*with\* drop-down menu is used to specify the query sequence. Specifying a different sequence in either menu will initiate a new pairwise alignment.

The Export image\* tool () located on the right of the header, lets you export the summary and sequences areas.

## Summary area:

DNA alignment [Matrix: "NUC44" Gap penalty: 10 Gap extension penalty: 1]									
	MG1655_rpmJ	CP040313.1				Gap	Gap		
	1>117	1>117	%Identity	%Gaps	Identical	Count	Length	Score	Length
Alignment	1>117	1>117	100.0%	0.0%	117	0	0	585	117

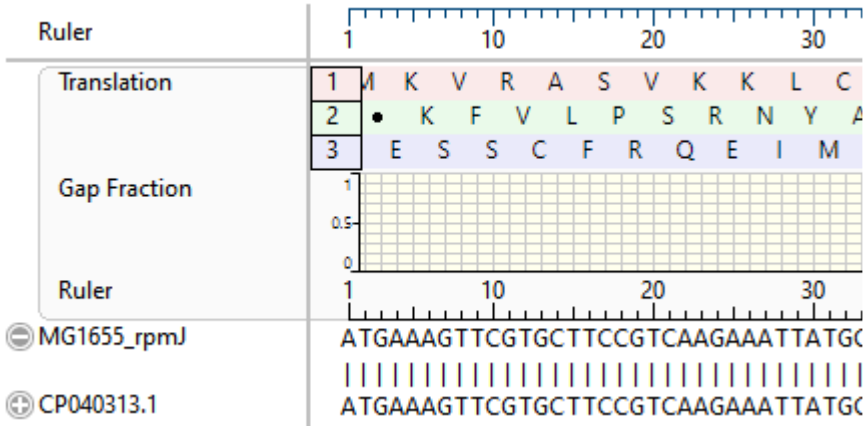
The summary area describes the alignment engine and settings that were used to create the pairwise alignment.

Value	Description
<b>% Identity</b>	Number of matching residues divided by the sum of all residues and gap characters. The quotient is multiplied by 100 to give the similarity as a percent.
<b>% Gaps</b>	The gap length divided by the consensus length. The quotient is multiplied by 100 to give the similarity as a percent.
<b>Identical</b>	Number of residues that are identical in both sequences.
<b>Gap Count</b>	Number of gaps (of any length) in either sequence of an alignment.
<b>Gap</b>	Number of gap characters in an alignment.

Length	
Score	The cumulative score for an alignment derived by adding the matrix-derived match scores and subtracting gap penalties.
Length	Number of residues in the consensus, including gaps that match in both sequences.

Sequences area:

The sequences area consists of nucleotide or amino acid codes for each sequence on the right, and their labels on the left. The order is always “target sequence” first, and “query sequence” second. A variety of optional [data tracks](#) can be displayed in this section.



# Text view

After the [Jobs panel](#) shows the **Status** of a [BLAST search](#) or [Entrez search](#) is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the Text, [Table](#), and [Pairwise](#) views. All three views are contained in a large pop-up window that is separate from the main SeqMan Ultra window.

An example Text view is shown below:

```
<p><!--
QBlastInfoBegin
  Status=READY
QBlastInfoEnd
--><p>
<PRE>
BLASTN 2.10.0+
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and
Webb Miller (2000), "A greedy algorithm for aligning DNA
sequences", J Comput Biol 2000; 7(1-2):203-14.

Reference for database indexing: Aleksandr Morgulis, George
Coulouris, Yan Raytselis, Thomas L. Madden, Richa Agarwala,
Alejandro A. Schaffer (2008), "Database Indexing for
Production MegaBLAST Searches", Bioinformatics 24:1757-1764.

RID: 1HSZN6Z7016

Database: Nucleotide collection (nt)
      55,411,115 sequences; 253,561,826,700 total letters
Query= MG1655_rpmJ
Length=117

Sequences producing significant alignments:
```

	Score (Bits)	E Value	Max Ident
CP046271.1 Enterobacter hormaechei strain ECL69214 chromosome,...	217	9e-53	100%
CP033387.1 Salmonella enterica subsp. enterica strain CFSA1007...	217	9e-53	100%
CP033384.1 Salmonella enterica subsp. enterica strain CFSA300 ...	217	9e-53	100%
CP033360.1 Salmonella enterica subsp. enterica strain EQAS2016...	217	9e-53	100%
CP023165.1 Escherichia coli O22:H8 strain RM10809-3 chromosome...	217	9e-53	100%






In order, the view shows the following:

- The version of BLAST used in the search (e.g., “BLASTP 2.9.0+”).
- References for the BLAST search engine.
- A description of the BLAST database used in the search, including the number of sequences.


- Information about the Query sequence, including its length and the Request ID (RID) number.
- A table of sequences producing significant alignments.
- Pairwise alignments of each significant match with summary information on the match.

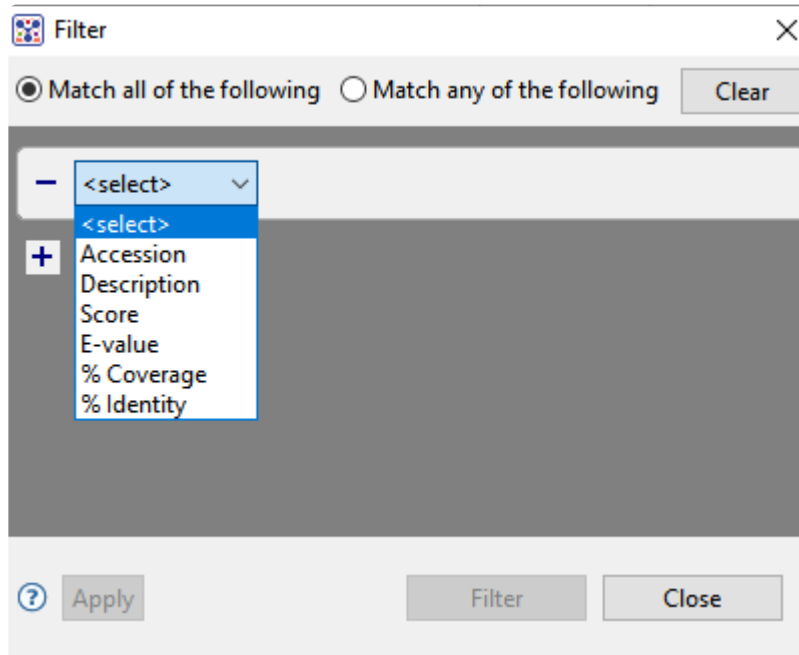
# Shared tools in the Table, Text and Pairwise views

After successfully searching for sequences online, the [Table](#), [Text](#) and [Pairwise](#) views appear. All three views share a common toolbar. The table below shows tasks that can be performed using the shared tools.

Task	How To...
To filter results from a sequence search	Use the <b>Filter results</b> tool (  ) . Filtering for this situation is described <a href="#">here</a> .
To save the search job to disk	Use the <b>Save as</b> tool (  ) .
To export data or sequences from the view	<p>Use the <b>Export data</b> tool () . Select from the following options:</p> <ul style="list-style-type: none"> <li>• <b>Matches without Gaps</b> – To export an ungapped FASTA (<i>.fasta</i>) file consisting of matching regions of displayed matches, but not the query sequence itself.</li> <li>• <b>Aligned Matches</b> – To export a gapped FASTA (<i>.fasta</i>) file with displayed matches aligned against the query.</li> <li>• <b>Table</b> – To save <a href="#">Table</a> view contents in either Comma- or tab-separated value (<i>.csv</i> or <i>.tab</i>) formats.</li> <li>• <b>Text</b> – To save the <a href="#">Text</a> view contents in Text (<i>.txt</i>) format.</li> </ul>
To download a selected match sequence in GenBank ( <i>.gbk</i> ) format	Click on the <b>Export data</b> tool (  ) and choose <b>Download Sequences</b> .
To launch the <a href="#">Access ID dialog</a>	Click on the <b>Export data</b> tool (  ) and choose <b>Set Access ID</b> .

# Filter sequence search results

When you press the **Filter results** tool () in the header above the [Table](#) or [Pairwise](#) views, there are two options: **Filter** and **Clear Filter**. Clicking on the **Filter** option opens the following dialog:



- Use the **<select>** drop-down menu to choose a filter type, then enter the desired options.

Filter type	Description	Options
<b>Accession</b>	The Accession number.	
<b>Description</b>	The NCBI "Definition" field text.	
<b>Score</b>	The similarity score for the match. (See the <a href="#">NCBI definition for raw score</a> ). In general, a higher score denotes a better match.	
<b>E-value</b>	The chance that the match is expected to occur coincidentally. (See the <a href="#">NCBI definition for E value</a> ). In general, a lower score denotes a better match.	
<b>% Coverage</b>	The percentage of the query sequence that aligns to a match in the BLAST database.  Example: Query has 596 bases, of which 565 align to a match in the NCBI database. <b>%Coverage</b> = 565/596 = 95%.	
<b>% Identity</b>	The percentage of bases that match when the query and result sequence are aligned. ( <a href="#">NCBI definition</a> ).  Example: Query has 565 bases that align to the match and 3 bases that don't match. <b>%Identity</b> = 562/565 = 99.5%.	

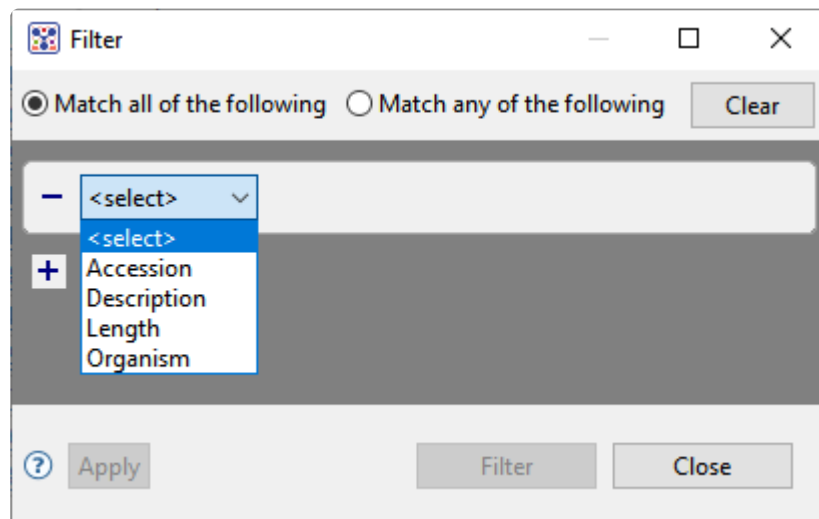
- To add another filter, press the plus icon (+). To delete a filter row, press the minus icon (-).
- Once you have set up all desired queries, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.

Once you are finished setting up filters, do any of the following:

- To apply the filters, press **Apply** or **Filter**.
- To clear all filters and start over, press **Clear**.
- To exit the dialog without filtering, press **Close**.

# Filter text search results

When you press the **Filter results** tool () in the header above the [Table](#) or [Pairwise](#) views, there are two options: **Filter** and **Clear Filter**. Clicking on the **Filter** option opens the following dialog:



- Use the **<select>** drop-down menu to choose a filter type, then enter the desired options.

Filter type	Description	Options
<b>Accession</b>	The Accession number.	
<b>Description</b>	The NCBI “Definition” field text.	
<b>Length</b>	The length of the match sequence in bases.	
<b>Organism</b>	The species name.	

- To add another filter, press the plus icon (+). To delete a filter row, press the minus icon (−).
- Once you have set up all desired queries, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.


Once you are finished setting up filters, do any of the following:

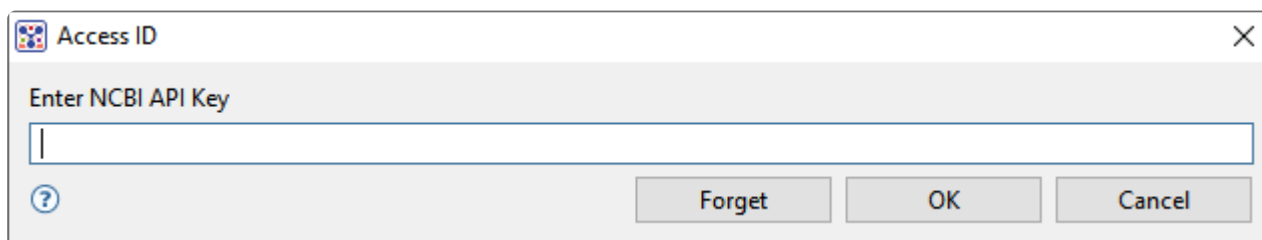
- To apply the filters, press **Apply** or **Filter**.
- To clear all filters and start over, press **Clear**.
- To exit the dialog without filtering, press **Close**.

# Access ID dialog

As of December 2018, NCBI requires an API key for faster access to NCBI services, or to run more searches concurrently. To get a key, start by [setting up an NCBI account](#). After registering your account, click on your user name and request your key using **API Key Management**. Once you have your key, you can enter it in Lasergene applications to enable NCBI's premium search access.

To add your key to a Lasergene application:

1. Begin an [Entrez database search](#) using **Search > Text Search in > NCBI Nucleotide** or **Search > Text Search in > NCBI Protein**.
2. Press the **Set Access ID** () tool in the bottom left corner of the dialog; or, from the [Table view](#), press the **Export data** tool and choose **Set Access ID**. In either case, the Access ID pop up appears.



The image shows a dialog box titled "Access ID" with a close button (X) in the top right corner. Inside the dialog, there is a label "Enter NCBI API Key" above a text input field. Below the input field, there is a help icon (question mark in a circle) on the left and three buttons: "Forget", "OK", and "Cancel" on the right.

- If an NCBI API key is not already present in the text box, enter your key and press **OK**.
- If an NCBI API key is present but you no longer use it, press **Forget** to remove it from the application's memory. Then enter the desired key and press **OK**.
- To exit without saving your changes, press **Cancel**.

# Customize Appearance and Behavior

---

Customize the appearance of the project and its data views:

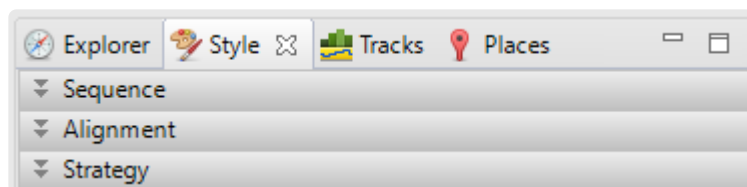
- [Customize the appearance of views](#)
- [Customize tables in the views](#)
- [Customize the interface layout](#)

Customize the behavior of SeqMan Ultra:

- [Specify the consensus calling method](#) parameters}
- [Specify the naming convention for paired reads](#)
- [Specify where to store temporary files](#)
- [Opt in or out of receiving update notifications](#)
- [Opt in or out of usage logging](#)

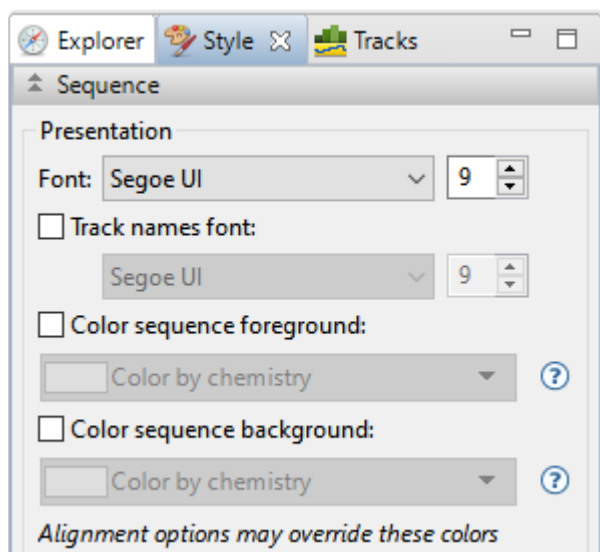
# Customize the appearance of views



The appearance of the [Alignment](#) and [Strategy](#) views is controlled via the Style panel, which is composed of three expandable sections.



## Sequence section:

To access options affecting the sequence font and background/foreground color, click on the Style panel expand bar entitled **Sequence**, or choose **View > Style > Sequence**.



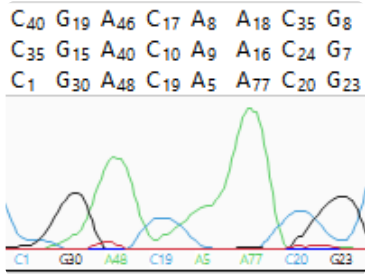
Task	How to...
<b>Select font options for sequence names</b>	Select a font from the <b>Font</b> drop-down menu. Recently selected fonts (if any) appear at the top of the menu, and other available fonts appear below the row of dashes. Enter the preferred font size by typing a number between 4-72 in the box at right, or by using the corresponding up and down arrows (  ).
<b>Select font options for tracks</b>	Check the box next to <b>Track names font</b> , then use the corresponding drop-down menu to change the font from the current value. Enter the preferred font size by typing a number between 4-72 in the box at right, or by using the corresponding up and down arrows (  ).
<b>Change the sequence</b>	To change the letter color in the Alignment view, check the <b>Color sequence foreground</b> box. To change the background color (boxes around each letter), check the <b>Color sequence</b>

<b>color scheme in the Alignment view</b>	<b>background</b> box. Then make selections from the corresponding drop-down menus. Color schemes are described in <a href="#">Available color schemes</a> . To specify which parts of the sequence should be colored, see <b>Alignment coloring</b> in the Alignment section below.
---	--

### Alignment section:

To access options affecting the Alignment view, click on the Style panel expand bar entitled **Alignment**, or choose **View > Style > Alignment**.

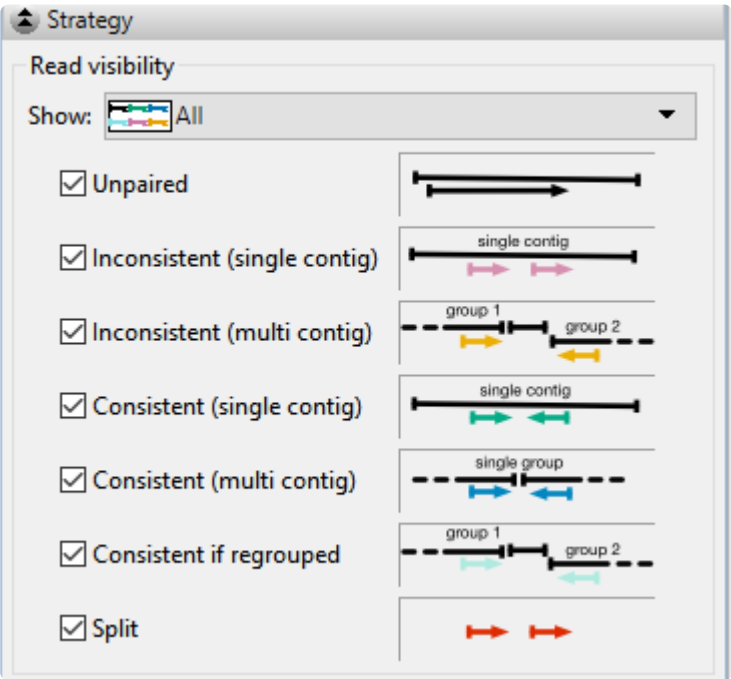
Task	How to...
<b>Specify the type(s) of editing allowed in an .sqd assembly</b>	<p>Make a selection from the <b>Editing</b> drop-down menu. Choices are:</p> <ul style="list-style-type: none"> <li>• <b>All editing allowed</b> – Allows you to type in new nucleotides or gaps into the consensus sequence, or to delete the same. To add gaps, use dashes (-).</li> <li>• <b>Only gap editing allowed</b> – Allows you to add gaps using dashes (-) or to remove gap-indicating dashes. Does not allow nucleotide characters to be added or deleted.</li> <li>• <b>No editing allowed</b> – Disallows adding or removing nucleotides or gaps in the Alignment view. If you choose this option and attempt to edit the consensus, you will receive a warning message asking if you would like to allow editing. If you respond with <b>Yes</b>, the <b>Editing</b> menu will be switched to <b>All editing allowed</b>.</li> </ul>

	<ul style="list-style-type: none"> <li>• <b>No editing allowed permanently</b> – This option is similar to <b>No editing allowed</b>. However, the only way to override the option is to manually make a different selection from the <b>Editing</b> menu.</li> </ul>
<b>Limit color-coding to the targeted coverage region only</b>	<p>Check the <b>Color only in targeted coverage regions</b> box to cause coloring options selected elsewhere in this section to only be applied to nucleotides in the targeted region of the assembly.</p>
<b>Display quality scores</b>	<p>To display quality scores as subscripts next to each nucleotide, check the <b>Quality</b> box and then make a selection from the drop-down menu to the right:</p> <ul style="list-style-type: none"> <li>• <b>Show scores</b> – Causes quality scores to be displayed as a subscript next to the base. When bases with very high quality scores (&gt;99% of the 95th percentile) are displayed, two asterisks (**) will be shown in place of the usual numerical subscripts. Example: G<sub>52</sub> A<sub>84</sub> G<sup>**</sup> C<sub>85</sub>. The quality score of a peak is calculated based on the shape and height of each peak and is adjusted relative to the maximum height of any peak in the entire sequence. Taller, sharper peaks receive the highest quality scores. The heights of any underlying peaks are subtracted from the highest peak's score. Illumina data will typically have higher scores than Sanger data. For <i>.fas</i> files, SeqMan NGen automatically checks for <i>.qual</i> files with the same name in the same folder. If found, the quality values are displayed. Note that gaps are not assigned quality scores.</li> </ul>  <ul style="list-style-type: none"> <li>• <b>Show averaged scores</b> – Displays quality scores averaged over a defined window. This mode is useful for examining trimming. Note that gaps are not assigned quality scores.</li> </ul>
<b>Specify where background and foreground coloring should be applied</b>	<p>To show where background and/or foreground coloring should be applied, make a selection from the <b>Alignment coloring</b> drop-down menu. This option is only active if at least one of the <b>Color sequence foreground</b> or <b>Color sequence background</b> boxes are checked in the <b>Sequence</b> section described above. Options include: <b>No special coloring</b>, <b>Color only differences from consensus</b>, <b>Color only matches to consensus</b>, <b>Show only differences from consensus</b>, and <b>Color matches and differences</b>. The last option will apply the foreground and/or background colors to all nucleotides.</p>
<b>Display negated</b>	<p>Check the corresponding font color box (left) or highlight box (right) to enable color choosing. Then click on the "A" and/or Color "highlighter" icons to open their color choosers.</p>

weights or trimmed sequence with a custom font color or highlighter

Strategy section:

To access options affecting the visibility of reads in the Strategy view, click on the Style panel expand bar entitled **Strategy**, or choose **View > Style > Strategy**.



Task	How to...
Show or hide specific types of paired reads in the Strategy view	Choose an option from the <b>Show</b> drop-down menu or manually check or uncheck the individual boxes: <b>Unpaired</b> , <b>Inconsistent (single contig)</b> , <b>Inconsistent (multi contig)</b> , <b>Consistent (single contig)</b> , <b>Consistent (multi contig)</b> , <b>Consistent if regrouped</b> , <b>Split</b> . If your group of checked/unchecked boxes does not match any of the presets, the <b>Show</b> menu will automatically change to <b>Custom</b> .

# Available color schemes

The [Sequence section](#) allows you to select color schemes for sequences in the [Alignment view](#). Each available color scheme is described below:

**Color by Chemistry** - This is the default color scheme, and matches the colors used in DNASTAR's Protean 3D and MegAlign Pro applications. Amino acids are colored according to their side chain chemistry:

Nucleotide
A
C
G
T/U

Side Chain Chemistry - Amino Acids
Aromatic - Phenylalanine (F), Tryptophan (W), Tyrosine (Y)
Acidic (negatively charged) - Aspartate (D), Glutamate (E)
Basic (positively charged) - Arginine (R), Histidine (H), Lysine (K)
Nonpolar (aliphatic) - Alanine (A), Glycine (G), Isoleucine (I), Leucine (L), Methionine (M), Proline (P), Valine (V)
Polar (neutral) - Cysteine (C), Asparagine (N), Glutamine (Q), Serine (S), Threonine (T)

**Shapely** - This color scheme matches the RasMol amino acid and RasMol nucleotide [color schemes](#), which are, in turn, based on Robert Fletterick's "Shapely models."

Nucleotide
A
C
G
T
U

Amino Acids
Alanine (A)
Arginine (R), Lysine (K)
Asparagine (N), Glutamine (Q)
Aspartate (D), Glutamate (E)
Cysteine (C), Methionine (M)
Glycine (G)
Histidine (H)
Isoleucine (I), Leucine (L), Valine (V)
Phenylalanine (F), Tyrosine (Y)
Proline (P)
Serine (S), Threonine (T)
Tryptophan (W)

**Taylor** - This color scheme is taken from [Taylor](#) and is also used in JalView ([Waterhouse et al., 2009](#)).

Nucleotides
A
C
G
T/U

Amino Acids
Alanine (A)
Arginine (R)
Asparagine (N)
Aspartate (D)
Cysteine (C)
Glutamate (E)
Glutamine (Q)
Glycine (G)
Histidine (H)
Isoleucine (I), Leucine (L)
Lysine (K)
Methionine (M)
Phenylalanine (F)
Proline (P)
Serine (S)
Threonine (T)
Tryptophan (W)
Tyrosine (Y)
Valine (V)

**Zappo** - This scheme colors residues according to their physico-chemical properties, and is also used in [JalView](#) ([Waterhouse et al., 2009](#)):

Nucleotide
A
C
G
T
U

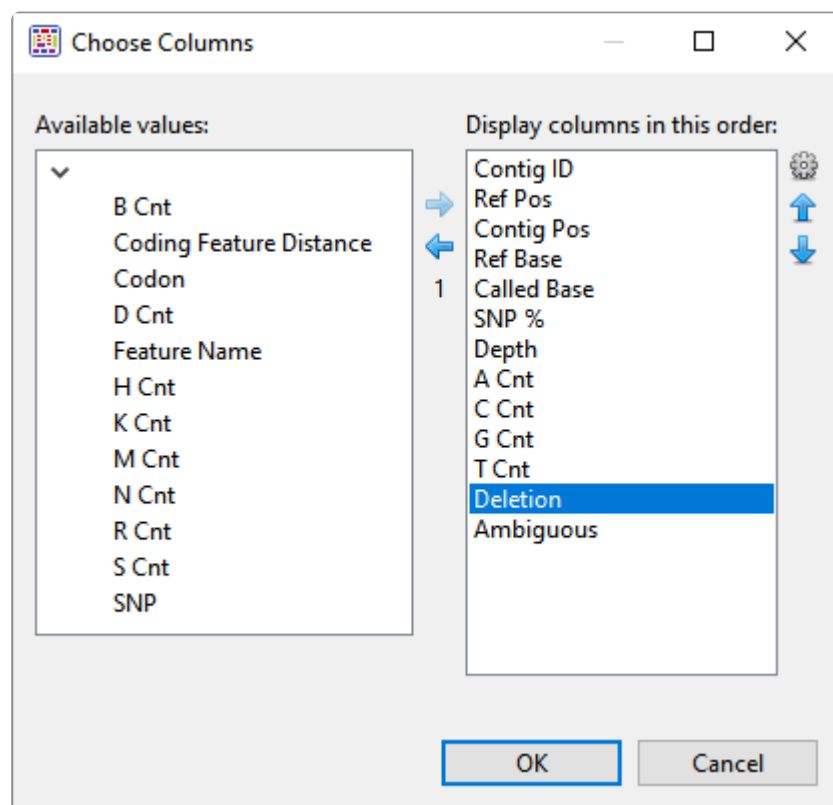
Physico-Chemical Property - Amino Acids
Aliphatic/hydrophobic - Alanine (A), Isoleucine (I), Leucine (L), Methionine (M), Valine (V)
Aromatic - Phenylalanine (F), Tryptophan (W), Tyrosine (Y)
Conformationally special - Glycine (G), Proline (P)
Cysteine (C)
Hydrophilic - Asparagine (N), Glutamine (Q), Serine (S), Threonine (T)
Negatively Charged - Aspartate (D), Glutamate (E)
Positively Charged - Arginine (R), Histidine (H), Lysine (K)

**Solid** – These scheme colors residues with the color shown. After selecting **Solid**, you can specify the color by clicking on the colored box to the right of the menu.



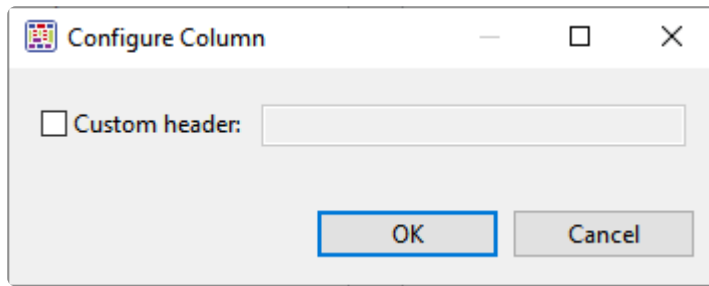
# Customize tables in the views

The [Contig Report](#), [Features](#), and [Variants](#) views allow you to add data columns to a table, rename the columns, change the order of the columns, or remove existing columns. This is done by clicking the **Change options in this view** (⚙️) tool in the top left of the header.




Available columns are on the left, while currently-applied columns are on the right. To see definitions for each available column, see [Descriptions of data columns](#).

- To add a column to the display, select its name on the left and press the right arrow key to move it to the right.
- To remove a column from the display, select its name on the right and press the left arrow key to move it to the left.
- To change the order of displayed columns, select the column name you wish to move on the right, then use the up/down arrows to place it in the desired order.
- To create a custom header for a displayed column, select its name on the right, then choose the **Configure column** tool (⚙️). In the popup dialog, type in the desired name and press **OK**.



# Descriptions of table data columns

When you click the **Change options in this view**  tool from within the [Contig Report](#) and [Features](#) views you have the option of selecting which columns should appear in the view's table. Only a subset of column types is available for each view and may further depend on the type of assembly that is open. Descriptions of each column appear below in alphabetical order.

Column Name	Description
'N' Cnt	The residue count, i.e., the number of a particular residue called in the aligned column. A dash (-) represents the reference base.
Ambiguous	A single base call could not be made.
Amino Acid Change	<p>This column is only available if <b>Show Codon Bases &amp; Distance to feature</b> is unchecked. The column shows the change(s) in the amino acid sequence, using the nomenclature established by the <a href="#">Human Genome Variation Society</a> and <a href="#">The Sequence Ontology Project</a>. This includes:</p> <ul style="list-style-type: none"> <li>Conservative in-frame insertions. Example: p.K2_M3insQSK denotes that the sequence GlnSerLys (QSK) was inserted between amino acids Lysine-2 (K) and Methionine-3 (M).</li> <li>Disruptive in-frame insertions. Example: p.C28delinsWV denotes a 3 bp insertion in the codon for Cysteine-28, generating codons for Tryptophan (W) and Valine (V).</li> <li>Conservative in-frame deletions. Example: p.(C28_M30del) a deletion of three amino acids, from Cysteine-28 to Methionine-30.</li> <li>Disruptive in-frame deletions. Example: p.(C28_M30delinsL) denotes a 9 bp deletion including 2 bp from the codon for Cysteine-28 and 1 bp from the codon for Methionine-30 resulting in replacement of C28 to M30 with leucine (L).</li> </ul>
Called Base	The dominant variant in the aligned column. In the case of a heterozygote call, both bases at the position are shown, separated by a vertical bar. For multi-base insertions, the inserted string is shown. For multi-base deletions, the deleted bases are represented with dashes (-).
Coding Feature Distance	<p>Shows whether variants are within or near a named feature, and the distance from that feature. For <i>.assembly</i> files and certain <i>.sqd</i> files (e.g., from <i>de novo</i> or special templated workflows), the following color scheme may be used:</p> <ul style="list-style-type: none"> <li>Gray + feature name + – Variant is within the named feature.</li> <li>Pink + arrow + feature name – Distance from the variant to the closest upstream coding feature.</li> </ul>

	<ul style="list-style-type: none"> <li>Orange + arrow + feature name – Distance from the variant to the closest downstream coding feature.</li> </ul> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <div style="background-color: #f0f0f0; padding: 2px; text-align: center;">Coding Feature Distance</div> <div style="padding: 5px;"> <p>CT &lt;- 146</p> <p>CT &lt;- 104</p> <p>CT &lt;- 103</p> <p>CT &lt;- 94</p> <p>CT &lt;- 90</p> <p>CT &lt;- 66</p> <p style="color: magenta;">139 -&gt; CT</p> <p>CT &lt;within&gt;</p> <p>CT &lt;within&gt;</p> <p>CT &lt;within&gt;</p> </div> </div>
<b>Codon</b>	When a translated feature is present on the reference sequence at the position of a variant, a codon change is displayed. The codon and amino acid translation is shown for the reference sequence and compared to the codon and amino acid translation for the selected variant. The position number of the amino acid change is also displayed. If more than one translated feature is present at the variant position, SeqMan Ultra will use the first feature based on the current sorting in the <a href="#">Features view</a> .
<b>Cons Pos</b>	Consensus position that includes gaps.
<b>Contig ID</b>	The name of the reference sequence or chromosome.
<b>Contig Pos</b>	Consensus position that includes gaps
<b>COSMIC</b>	The Catalogue of Somatic Mutations in Cancer ( <a href="#">COSMIC</a> ) ID for positions with known variants. Double-clicking on the entry opens the corresponding page at COSMIC. For human assemblies only.
<b>Count</b>	(isoforms only) The total number of aligned reads assigned to an isoform feature, after adjustments for repeat distribution.
<b>Coverage Depth</b>	The average depth of coverage across the feature, calculated as the sum over all segments of the total number of bases covering each segment divided by the length of that segment. Note that the coverage for a gene may differ substantially from that of its corresponding CDS (e.g. exome or transcriptome sequencing). Values range from zero to the maximum coverage in the project and values are shown to the second decimal point to account for areas of very low, but non-zero coverage.
<b>Coverage Strand Depth</b>	(stranded RNA-seq only) The total number of sequenced bases within the feature and on the correct strand, divided by the length of the feature.
<b>dbSNP ID</b>	The dbSNP rs ID, if available, for positions with known variants. Double-clicking on the entry opens the corresponding page at <a href="#">dbSNP</a> .
<b>Deletion</b>	The number of deleted bases in the aligned column.
<b>Depth</b>	The number of reads overlapping the aligned column. Since this calculation disregards bases below the quality threshold, the <a href="#">Alignment view</a> may show a greater number of sequences than the <b>Depth</b> shown in tabular views. The default quality threshold for assembly in SeqMan NGen is 5. The threshold can

	be changed either pre-assembly, in SeqMan NGen, or post-assembly, in SeqMan Ultra.
<b>DNA Change</b>	<p>Change(s) in the DNA sequence affecting either CDS features or splice sites are indicated using the nomenclature established by the <a href="#">Human Genome Variation Society</a>. A “c.” prefix, followed by coordinates taken from the ORF, denotes a change in a CDS feature. For example:</p> <ul style="list-style-type: none"> <li>• Substitutions. Example: c.76A&gt;C denotes that at nucleotide 76 an A is changed to a C.</li> <li>• Insertions within coding regions. Example: c.76_77insT denotes that a T is inserted between nucleotides 76 and 77.</li> <li>• Deletions within coding regions. Example: c.76_78delACT denotes an ACT deletion from nucleotides 76 to 78.</li> </ul> <p>A “g.” prefix followed by genomic coordinates denotes a change in the intronic region of a splice site.</p> <p><b>Note:</b> When a multibase variant affects both the intron and exon portions of a splice site, it is represented under two separate entries: one with g. coordinates and the other with c. coordinates.</p>
<b>Deletion</b>	
<b>Depth</b>	
<b>Feature Name</b>	<p>If a variant is located within an annotated feature in the reference sequence, the feature type and name are displayed. A single nucleotide change may sometimes be reported as affecting multiple overlapping features. These can include different overlapping genes on the same or opposite strands, as well as alternatively spliced messages from the same gene. In this case, SeqMan NGen produces multiple VCF Variant table entries at the same position, one for each reported feature. A bracketed number follows the <b>Feature Name</b> to indicate which isoform from the Feature view table was used (e.g., TP53 [2]).</p> <p><b>Note:</b> If a non-gene feature (“mRNA”, “CDS”, etc.) exists in the template file, but has no corresponding “gene” feature, SeqMan NGen adds the “gene” feature automatically during assembly. The locations of any automatically added “gene” annotations are indicated by asterisks (*) in this column.</p>
<b>Feature Type</b>	<p>For variants within a gene feature, the feature type is shown in the following order of precedence: <b>CDS, mRNA, Gene</b>. If <b>Show Codon Bases Distance to feature</b> is selected, this column also contains a feature designation if the variant is within 150 bases of the nearest exon. Therefore, it is possible for a variant that is in a gene to also be listed as a <b>CDS, mRNA</b>, etc. When <b>Show Codon Bases</b> is checked, the <b>Feature Type</b> column will also show the distance to the nearby exon and an arrow indicating the direction of the feature. Feature types for different variant locations are shown below:</p>

	<ul style="list-style-type: none"> <li>• <b>gene</b> – Within a gene feature, but not included in an mRNA or CDS feature for that gene. (Variants within the intron portion of a splice site are indicated as CDS features.)</li> <li>• <b>CDS</b> – Within an exon or splice site.</li> <li>• <b>mRNA</b> – In the 5' and 3' untranslated portions of an mRNA.</li> </ul>
<b>GERP</b>	<p>The <a href="#">Genomic Evolutionary Rate Profiling</a> (GERP) score representing the calculated evolutionary constraint at that position. GERP data is automatically delivered when you use DNASTAR's human template package prior to performing a templated assembly in SeqMan NGen. To limit the size of the data file required, only positions with scores of 1.0 or greater are displayed.</p> <p>GERP is a tool that provides a score for each position in the human genome that estimates whether that position is under purifying selection or not (Davydov et al. 2010). GERP uses alignments between the human genome and 33 other mammalian genomes to quantitate the position-specific constraint in terms of rejected substitutions, defined as the difference between the neutral rate of substitution and the observed rate, estimated by maximum likelihood. Substitutions in sites under selection are assumed to be more deleterious than those not under selection. Scores range from negative values to ~6. Positions with scores below or near zero are not under selection. Conversely, the more positive the score, the more constrained the position. GERP information can be useful in evaluating the impact of non-synonymous variants in coding regions and the impact of changes in or near promoter elements, among others.</p>
<b>Genotype</b>	<p>When the "Diploid" SNP detection method is used in a SeqMan NGen assembly, there are four possibilities: 1) homozygous variant (both alleles have the same base and it is different from the reference), 2) reference (both alleles have the same base and it is the same as the reference), 3) heterozygous reference (two different alleles are called, one with the same base as the reference, the other with a variant base), and 4) heterozygous not reference (two different alleles, neither of which match the reference base). It is quite rare for the reference case to occur in the table. This only happens in cases where there is sufficient evidence of the possibility of a variant to pass the filtering threshold, but where the evidence is still quite weak. These cases are usually eliminated by even modest filtering. When the Haploid SNP detection method is used, only variant and reference are possible.</p> <p><b>Note:</b> In this column, if one or more of the adjacent variants is called as a heterozygote, the coalesced variant is also called a heterozygote. Therefore, for a coalesced variant to be called homozygous, all positions must be called homozygous.</p>
<b>Group</b>	<p>Specifies whether the variant was called from the NGS or the Sanger data. The NGS group is named using the Read technology selected in the SeqMan NGen wizard. If you are investigating a putative variant that was identified in the initial run of the NGS data, you can find that position here and open it in the <a href="#">Alignment view</a> to check whether there is a Sanger-called variant at the same position.</p>
<b>Homopolymer</b>	<p>Indicates whether the variant occurs within a homopolymeric run, which is defined as two or more identical bases in a row. When using Pacific Biosciences (PacBio) or Ion Torrent data, SeqMan Pro and SeqMan Ultra may not list all</p>

	<p>homopolymeric indels.</p> <p><b>Note:</b> When possible, insertions or deletions are placed at the 5' end (top strand) of the run during alignment.</p>
<b>Impact</b>	<p>The impact of the variant or indel on the protein, displayed as:</p> <ul style="list-style-type: none"> <li>• <b>Synonymous</b> – No amino acid changes.</li> <li>• <b>Non Synonymous</b> – Amino acid substitution only.</li> <li>• <b>Nonsense</b> – Amino acid to translational stop.</li> <li>• <b>Frameshift</b> – An indel within a coding region and which is not a multiple of 3, thereby changing the reading frame.</li> <li>• <b>No Start</b> – A change that disrupts the start codon.</li> <li>• <b>No Stop</b> – A change that converts a stop codon to an amino acid, and thereby extends the reading frame.</li> <li>• <b>Inframe Insertion</b> – An insertion within a coding region whose length is divisible by 3. The type is followed by the word <b>Conservative</b> if the insertion occurs between two codons, and <b>Disruptive</b> if it occurs with a codon.</li> <li>• <b>Inframe Deletion</b> – A deletion within a coding region whose length is divisible by 3. The type is followed by the word <b>Conservative</b> if the insertion occurs between two codons, and <b>Disruptive</b> if it occurs within two codons.</li> </ul> <p>If sorting by the <b>Impact</b> column, the column is ordered by severity. For example, a <b>Frameshift</b> is more severe than a <b>Nonsense</b> change.</p>
<b>Location</b>	The range of sequence associated with the feature, including gaps.
<b>MID</b>	Displays variants separately for each MID sample. If the same variant (same base change, same position) occurs in more than one sample, there will be an entry for each sample. Similarly, if the same position is affected, but the base change is different, there will be separate entries and columns will correspond to that sample only, and not all the reads at that position.
<b>Name</b>	The name of the feature. The /dnas_title qualifier is used for the feature name. If no /dnas_title is available, SeqMan Ultra will use the value of the first qualifier listed for the name.
<b>P Not Ref</b>	The probability that the called base at this position is not the reference base. For coalesced variants, this value is equal to the minimum value of all “child” values. The minimum allowed value is 30%.

<b>PDB ID</b>	<a href="#">Worldwide Protein Data Bank</a> (PDB) ID number.
<b>Q Call</b>	The Phred-like quality score of the called genotype. It is a measure of the probability that the called genotype is correct.
<b>Quality</b>	
<b>Raw_count</b>	(isoforms only) The number of reads that were initially and uniquely assigned to an isoform, fragment or peak. This value is not normalized and does not account for repeat reads.
<b>Raw_repeat_count</b>	(isoforms only) The total fraction of repeat reads that map to the peak. For example, if one read maps equally well to two peaks, each peak would have a Raw_repeat_count of 0.5. If multiple repeat reads map to a single peak, the proportions are summed to get the Raw_repeat_count. This column is only available if repeat handling was specified in SeqMan NGen.
<b>Ref</b>	
<b>Ref Base</b>	The reference sequence base in this position. For multi-base deletions, the reference sequence of the string is shown, beginning with the base at the Ref Pos coordinate. If there is no reference sequence present, the Ref Base column displays the most frequently occurring non-ambiguous base at this position. If no such base exists, the consensus base at this position is shown.
<b>Ref ID</b>	Reference sequence or chromosome.
<b>Ref Pos</b>	Reference position that does not include gaps. Coordinates matching entries in the VCF Variant table are shown. For deletions, <b>Ref Pos</b> is the genomic coordinate of the first deleted base. For insertions, <b>Ref Pos</b> is the genomic coordinate of the base preceding the insertion.
<b>Region Capture</b>	Indicates whether the variant occurs within a region specified in the <i>.bed</i> or manifest file used. Values are Yes and No.
<b>Repeat_distrib_count</b>	(isoforms only) The proportional number of repeated reads assigned to this exon, gene or isoform.
<b>Repeat_distrib_percent</b>	(isoforms only) A rough estimate of the percentage of the total repeated reads which could have been assigned and that were assigned. In other words, the percentage of Repeat_distrib_count in Raw_repeat_count.
<b>Replicate set total count</b>	(RNA-seq with no normalization specified only) The total number of aligned reads averaged over the set of replicate samples. Only available if replicates and replicate sets were specified in SeqMan NGen.
<b>Replicate set total RPK</b>	(RNA-seq with RPK normalization specified only) The RPK value averaged over the set of replicate samples. Only available if replicates and replicate sets were specified in SeqMan NGen.
<b>Replicate set total RPKM</b>	(RNA-seq with RPKM normalization specified only) The RPKM value averaged over the set of replicate samples. Only available if replicates and replicate sets were specified in SeqMan NGen.
<b>Replicate set total RPM</b>	(RNA-seq with RPM normalization specified only) The RPM value averaged over the set of replicate samples. Only available if replicates and replicate sets were specified in SeqMan NGen.
<b>Sequence</b>	The name of the sequence that contains the feature.
<b>Sequence Pos</b>	












<p><b>Skew</b></p>	<p>When a maximum “strand bias” (also called “skew”) is set for a templated SeqMan NGen assembly (i.e., by setting the scripting parameter <code>snp_maxStrandBias</code> to ‘true’), SeqMan NGen calculates the strand bias for each variant. The results can be viewed in the Skew column of this table.</p> <p>The strand bias for a variant is a bias in the variant appearing on one strand instead of the other. It is measured relative to the strand bias in the assembly at the location of the variant. For example, in a column with 60 forward reads and 40 backward reads, 6 variant bases on the forward strands and 4 on the reverse would be unbiased and have a skew of zero.</p> <p>SeqMan NGen calculates strand bias using the following formula:</p> $\text{Strand bias} = (\text{SNP}\%f - \text{SNP}\%r) / (\text{SNP}\%)$ <p>... where:</p> <ul style="list-style-type: none"> <li>• SNP%f= Strand-specific SNP percentage for the forward strand</li> <li>• SNP%r= Strand-specific SNP percentage for the reverse strand</li> <li>• SNP%= Overall SNP percentage</li> </ul> <p>Below are interpretations of several <b>Skew</b> values:</p> <ul style="list-style-type: none"> <li>• 0 = unbiased</li> <li>• 2 = all SNPs on one strand, where strands are equally abundant</li> <li>• &gt; 2 = SNPs are present in large numbers on a strand that is, itself, rare.</li> </ul> <p>The maximum theoretical strand bias is equal to the depth of coverage. In practice, however, numbers over four are seldom seen, as they require such low variant percentages that they are unlikely to be called as variants.</p> <p>When a variant occurs only on strands in one direction, no strand bias can be calculated. In this case, the variant is not filtered. This situation can be prevented by setting a Minimum strand coverage value in SeqMan NGen prior to assembly.</p>
<p><b>SNP</b></p>	<p>Shows the symbol for whether this is a confirmed (checkmark), putative (question mark), rejected (‘x’) or mixed (dot) variant.</p>
<p><b>SNP %</b></p>	<p>The percentage of the single most prevalent non-reference base in the aligned</p>

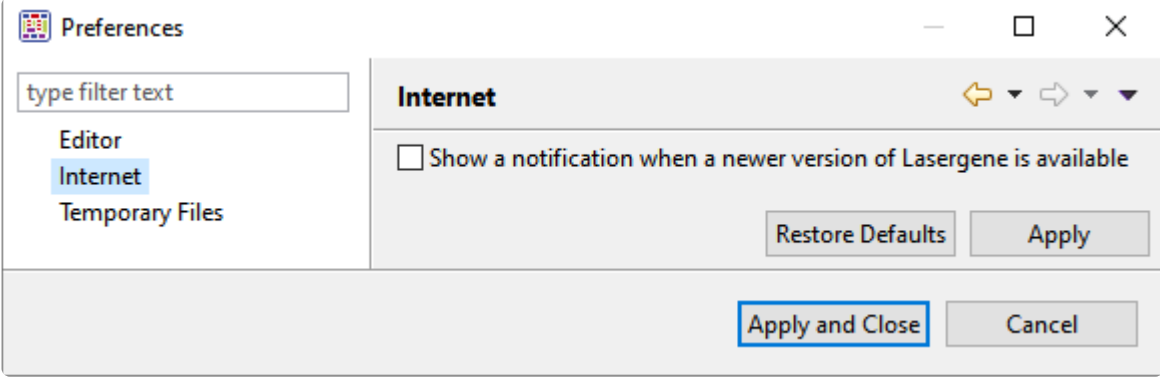
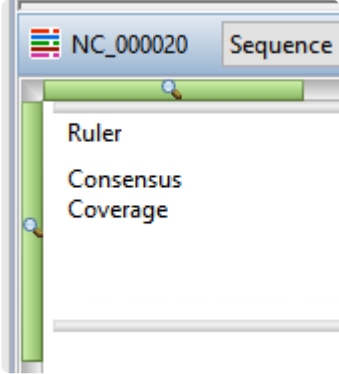




	column. As a very general rule, significant variants tend to occur at 25% and higher.
<b>Splice</b>	Variant is in or near an exon splice site. Splice site variations are changes to the 5' ("donor") or 3' ("acceptor") consensus splice site sequences. The DNA sequence for the donor is 5'-AGGTRAGT-3' and for the acceptor is 5'-YYYYYYYYCAGGT-3'. Note that the AG dinucleotide on the 5' end of the donor and the GT dinucleotide on the 3' end of the acceptor are within the exon. Therefore, changes at these positions can also cause changes in the amino acid sequence of the resulting proteins. Changes in the intron portion of the splice site are marked as "Splice" in the column while those in the exon portion of the site are labeled "Splice in CDS." Only the position where the change occurs is considered, not the identity of the base.
<b>Total_count</b>	(RNA-seq with no normalization specified only) The sum of the counts for all of the isoform features within a gene.
<b>Target Length</b>	The target length of the fragment, equal to the end position minus the start position.
<b>Total raw count</b>	(genes only) The total number of reads which were initially, uniquely assigned to any isoform within the gene.
<b>Total_repeat_distrib_count</b>	(genes only) The total count of reads which were proportionally distributed to any isoform within the gene.
<b>Total RPK</b>	(RNA-seq with RPK normalization specified only) RPK (reads assigned per kilobase of target) = The signal values for each experiment divided by the total bases of target sequence divided by one thousand.
<b>Total RPKM</b>	(RNA-seq with RPKM normalization specified only) Reads assigned per kilobase of target per million mapped reads (RPKM) = The signal values for each experiment divided by the total bases of target sequence divided by one thousand; the resulting number is then divided by the total number of mapped reads divided by one million.
<b>Total RPM</b>	(RNA-seq with RPM normalization specified only) Reads assigned per million mapped reads (RPM) = The signal values for each experiment divided by the total number of mapped reads divided by one million.
<b>Trace%</b>	
<b>Transcript ID</b>	Transcript ID number from ENSEMBL.
<b>Type</b>	Specifies the variation type as SNP, Del (deletion) or Ins (insertion). For .assembly files and certain .sqd files (e.g., from <i>de novo</i> or special templated workflows), pink typeface may be used to indicate non-synonymous variants.
<b>User ID</b>	Positions corresponding to a custom VCF Variant Table are labeled with the ID from that set.
<b>Variant Count</b>	The number of called variants in the feature. Note that a gene feature in an exome project can have a very low <b>Coverage Depth</b> value and still have one or many variants.
<b>Visible</b>	A checkmark in this column indicates that the feature is currently displayed in the <a href="#">Alignment view</a> (for features on constituent sequences) or the <a href="#">Strategy view</a> (for features on the consensus sequence). The absence of a checkmark indicates that this feature is hidden.

**Note:** The **Visible** column does not apply to variant features, which are available only in .sqd projects. To control the display of variants, select **Variant > Show/Hide Variants**. For more information, [see Work with Variants](#).

# Customize the interface layout

The following table shows methods for organizing, resizing, and changing the layout of items in the SeqMan Ultra window.

Task	How to...
Show a view, report or table	Use <b>View &gt; (Name of view, report or table)</b> . In some cases, you may need to specify the contig name using a subcommand. For instance, to see the Alignment view for Contig x, you would choose <b>View &gt; Alignment &gt; Contig x &gt; Show</b> .
Scroll vertically	Use the vertical scrollbar or the mouse wheel.
Scroll horizontally	Use the horizontal scrollbar, or hold down the <b>Alt</b> key (Win) or <b>Option</b> key (Mac) and drag the mouse.
To show the active view in full-screen mode	Use the <b>View &gt; Full Screen</b> command or the <b>Full screen</b> tool  at the bottom right corner of the view.
To exit full screen mode	Press the <b>Esc</b> key or use the <b>Exit full screen</b> tool  at the bottom of the full-screen view.
To configure the number and location of visible views.	<p>Use the <b>View &gt; Change layout</b> command or the <b>Change layout</b> tool (image varies) in the bottom right corner of the view. The following options are available:</p> <ul style="list-style-type: none"> <li>• <b>One</b> –  <b>One</b> – To show only the current view.</li> <li>• <b>Two Horizontal</b> (  <b>Two Horizontal</b> ) – To show the current view on top and one additional view below it.</li> <li>• <b>Two Vertical</b> (  <b>Two Vertical</b> ) – To show the current view on the left, and one additional view to the right.</li> <li>• <b>Three Horizontal</b> (  <b>Three Horizontal</b> ) – To show the current view on top and two additional views below it.</li> <li>• <b>Three – Left</b> (  <b>Three - Left</b> ) – To show the current view on the left and two additional views stacked one on top of the other on the right.</li> <li>• <b>Three – Right</b> (  <b>Three - Right</b> ) – To show the current view on the right and two additional views stacked one on top of the other on the left.</li> <li>• <b>Three – Top</b> (  <b>Three - Top</b> ) – To show the current view on the top and two additional views side-by-side on the bottom.</li> <li>• <b>Three – Bottom</b> (  <b>Three - Bottom</b> ) – To show the current view on the bottom and two additional views side-by-side on the top.</li> </ul>
Detach a specific display	Right click on its tab and choose <b>Detach &gt; (View Name)</b> or use the <b>Detach display</b> tool (  ) in the upper right corner of some views. Detaching a view turns it into a free-floating window. To reattach the view, see two rows down on this table.
Detach all	Choose <b>Edit &gt; Preferences</b> , select <b>Editor</b> on the left, add a checkmark next to <b>Open</b>

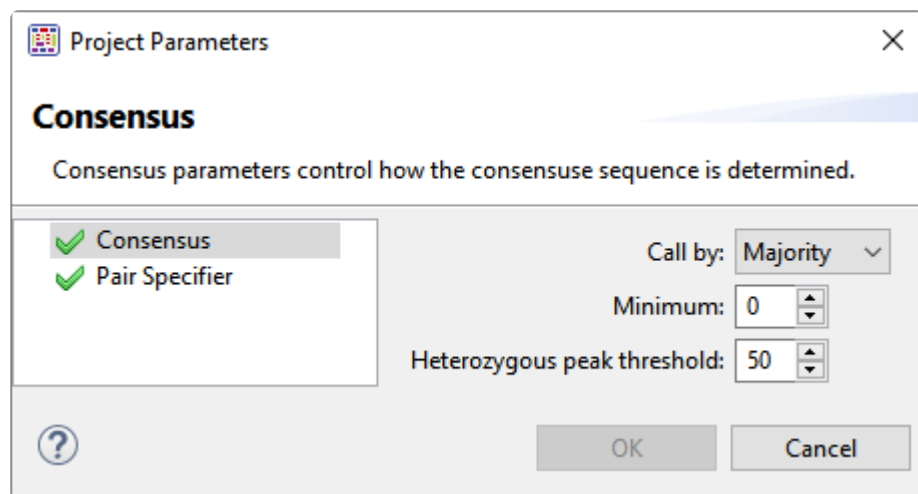
<p>future Alignment views by default</p>	<p>alignments in a detached window and press <b>Apply and Close</b>.</p> 
<p>Reattach all detached displays to their original locations</p>	<p>Use <b>View &gt; Reattach All Views</b>.</p>
<p>Change the relative sizes of panes</p>	<p>Drag the light gray divider bar between them to the left, right, up or down.</p>
<p>Zoom in or out on a view</p>	<p>Some SeqMan Ultra views have independent vertical and horizontal zoom in/out sliders. Grab any green slider with your mouse and drag left/right or up/down to zoom in or out in the view.</p> 
<p>Show or open a hidden item</p>	<p>Use the associated <b>View</b> menu command, click on a closed expand bar, or click on a plus sign.</p>
<p>Hide or close a visible item</p>	<p>Use the <b>Close</b> icon (  ) or <b>Minimize</b> icon (  ), click on an open expand bar, or click on a minus sign. To close one or more tabs, right-click on a project, view or panel tab and choose <b>Close</b>, <b>Close Others</b>, <b>Close Tabs to Right</b>, or <b>Close All</b>. Some of these options are only available for certain tabs.</p>
<p>Increase or decrease the size of an item</p>	<p>Drag the bottom or left sides of the item or use the <b>Maximize</b> icon (  ) or <b>Restore</b> icon (  ). To view more of an open expand bar, close neighboring bars and/or increase the size of the whole panel.</p>
<p>Move an item</p>	<p>Using the mouse, grab a project, view or panel tab, then drag and drop it in the desired location.</p>
<p>Move a view or panel up/</p>	<p>Right-click on the view or panel tab and choose <b>Move Up (Down)</b>.</p>

down	
Return to the default SeqMan Ultra display	Choose <b>View &gt; Reset Panel Layout</b> .
Return to a view or bring a floating view to the top	Use commands in the <b>View</b> menu.

# Consensus parameters

SeqMan NGen calculates the consensus sequence for a *de novo* assembly. After opening the project in SeqMan Ultra, you can elect to change the consensus calling parameters so that SeqMan Ultra can recompute the consensus. The Consensus dialog lets you choose the method for calling the consensus sequence.

To access the dialog, use **Project > Project Parameters**. If the Consensus section is not already active, click on the word **Consensus** on the left.



In the **Call by** drop-down menu, choose between **Majority** and **Trace evidence**. In the **Minimum** box, enter an integer as described below.

- **Trace Evidence** directly reflects the quality of peaks in the trace data and is the most accurate method for generating consensus sequences for contigs. It is strongly recommended you use **Trace Evidence** for consensus calling whenever trace data are available. Sharp, well-defined peaks are assigned higher scores than less well-defined peaks.
- **Majority** should be chosen when trace data are not available. If you select **Majority**, use the **Minimum** box to enter the minimum percentage of identical residues at each position required to identify the consensus base at this position. Ambiguous bases in the consensus denote a column of bases whose constituents do not meet this threshold. Ambiguous codes are assigned in these cases.

In the **Minimum majority percentage** box, enter the minimum percentage of a called base in a column to consider it the majority. The default is 75.

Enter a value for the **Heterozygous peak threshold**, used for called the secondary peak at heterozygous positions. With trace data, diploid heterozygous positions show two trace peaks at the same position. However, the secondary peak must be at least 50% the “size” of the primary peak to be considered a heterozygous variant. Trace data also has some background noise, so adjusting the **Heterozygous peak**

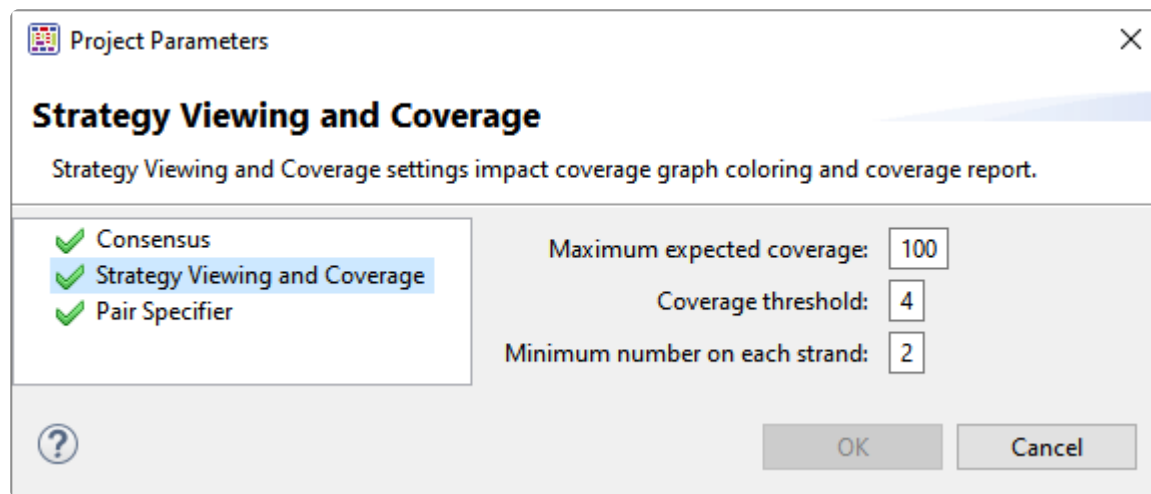
**threshold** value is often necessary to get accurate variant calling in trace data.

Click **OK**.

# Specify coverage parameters

Parameters in the Strategy Viewing and Coverage dialog impact coverage graph coloring in the [Strategy view](#) and tabular coverage data in the [Contig Coverage](#) report.

To access the dialog, use **Project > Project Parameters**. Click on the phrase **Strategy Viewing and Coverage** on the left.



- **Maximum expected coverage** – The maximum coverage expected at any position in the assembly. Positions/ranges above this value will be indicated as “Coverage above maximum” in the [Contig Coverage report](#). In the Strategy view, areas meeting or exceeding this maximum value will be indicated by thick, red areas in the Coverage histogram. The **Maximum expected coverage** value must be greater than the **Coverage threshold** value. The default is 100.
- **Coverage threshold** – The minimum coverage required to be considered sufficient. Positions/range below this value will be indicated as “Below coverage threshold” in the Contig Coverage report. For random sequencing projects, coverage of five to tenfold is commonly desired. However, if you are using the Trace Evidence method to [call the consensus](#), you may well achieve the accuracy you seek with coverage of only threefold. The default is 4.
- **Minimum number on each strand** – The minimal number of reads on each strand required to be considered sufficient. Positions/ranges where either strand is below this value will be indicated as “Only one strand” in the Contig Coverage report. The default value is 2.

# Specify the naming convention for paired reads

The Pair Specifier dialog lets you choose the method for recognizing paired reads.

To access the dialog, use **Project > Project Parameters**. If the Pair Specifier section is not already active, click on **Pair Specifier** on the left.

**Project Parameters**

**Pair Specifier**

Pair specifiers allow changing the insert size when dealing with paired end data.

☒ Consensus  
☒ **Pair Specifier**

**Pair Pattern:**

Forward:  Reverse:

**Paired-End Insert:**

Minimum:  Maximum:

- Choose from typical pair naming conventions in the **Pair Pattern** drop-down menu.

```
sample_f.abi <> sample_r.abi
sample100.f_abc.abi <> sample100.r_abc.abi
sample_n100.abi <> sample_f100.abi
sampleD1234.abi <> sampleE1234.abi
sample#001/1 <> sample#001/2
ID_1234_f <> ID_1234_r
sample_1 <> sample_1
Custom
```

If you choose **Custom**, you will need to enter the custom naming convention in the **Forward** and **Reverse** boxes.

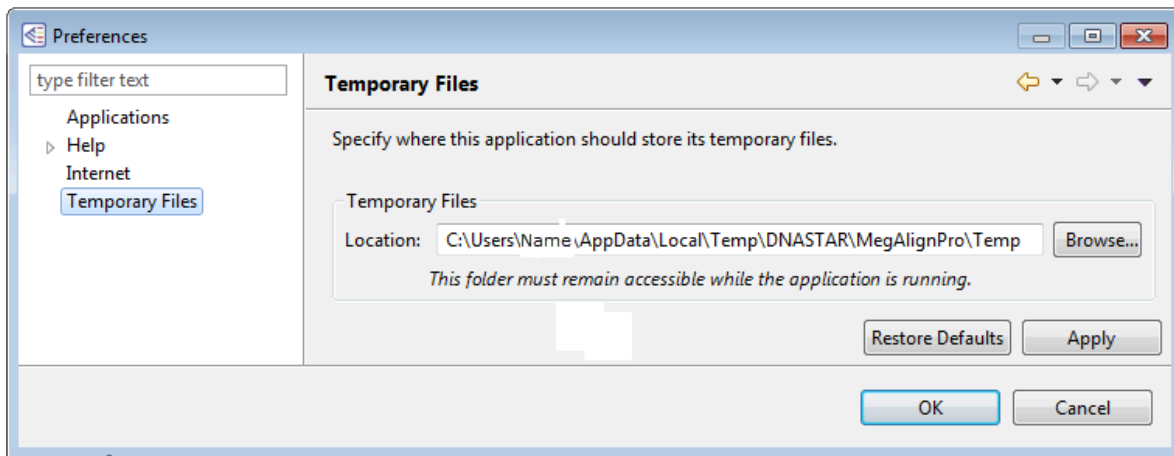
- Under **Paired-End Insert**, enter integers for the **Minimum** and **Maximum** expected distance (in bp) between the forward and reverse paired reads.

Click **OK**.

# Specify where to store temporary files

SeqMan Ultra creates and uses temporary files while a project is open. The need for temporary file space grows with the number and length of the sequences and amount of track data in open projects. If you expect the temporary files will be too large to fit in the default location, you may specify a different location for these files using the Temporary Files Preferences dialog.

To access the dialog, use **Edit > Preferences** (Win) or **SeqMan Ultra > Preferences** (Mac), then click on **Temporary Files** on the left.



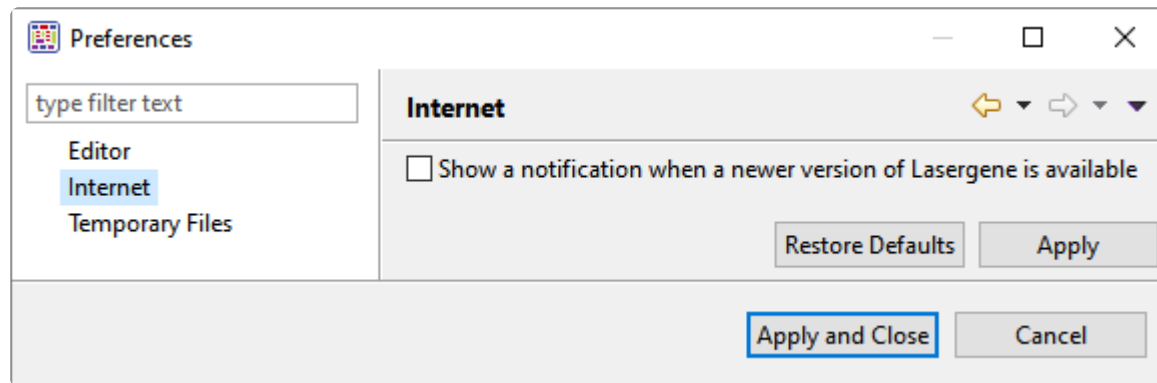
- To change the temporary file location, close all SeqMan Ultra projects and then return to this dialog. Select the **Browse** button, navigate to the new location and click **OK**. As noted in the dialog, the location you choose must remain accessible while SeqMan Ultra is running. The new location will be used for all projects that are subsequently created or opened.
- To return to the default (temporary) storage location, click **Restore Defaults**.

Click **Apply** and **Close**.

# Opt in or out of showing update notifications

The Internet Preferences dialog lets you choose whether you wish to be notified about Lasergene updates.

To access the dialog, use **Edit > Preferences** (Win) or **SeqMan Ultra > Preferences** (Mac). If the Internet section is not already active, click on the word "Internet" on the left.



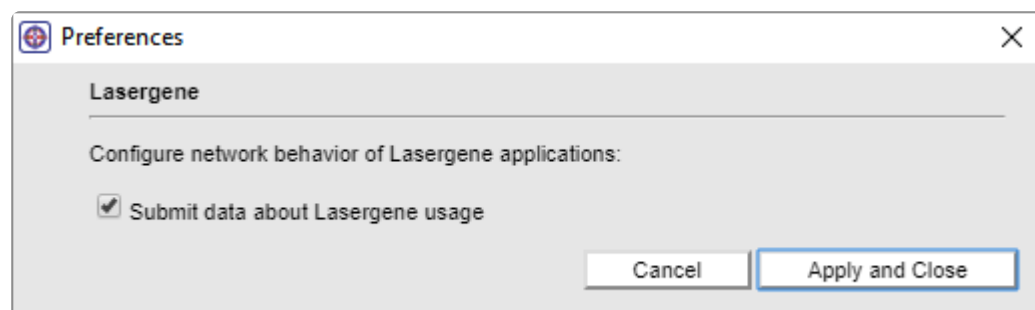
- Check the box if you would like SeqMan Ultra to display a notification when a newer version of Lasergene is available. Uncheck the box if you do not wish to receive this notification.

Click **Apply and Close**.

# Opt in or out of usage logging

Usage logging allows DNASTAR to learn which applications and workflows are of most interest to our users. By default, [usage logging](#) is enabled in Lasergene version 11 and later.

To change the usage logging setting, launch the DNASTAR Navigator and choose **Window > Preferences**.







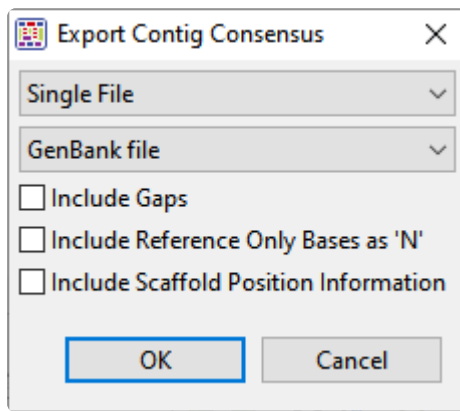
- Check or uncheck the box next to **Submit data about Lasergene usage**.

Click **Apply and Close**.

# Export Data

The following table shows how to export one or more consensus sequences or to export tabular data from a view in SeqMan Ultra.

Item to Export	How To
Any version of the <a href="#">Project Report</a> , <a href="#">Contig Report</a> or <a href="#">Variants</a> views	With the view open, click on the <b>Export data</b> tool (  ) at the right of the header.
<a href="#">Features</a> view	With the Features view open, click on the <b>Export data</b> tool (  ) at the right of the header or choose the <b>File &gt; Export Data &gt; Features &gt; (Contig Name)</b> command. The tabular data can be saved in either comma- or tab-separated value (.csv or .tab) formats.
Data or sequences from the <a href="#">Text</a> , <a href="#">Table</a> or <a href="#">Pairwise</a> views	<p>Click on the <b>Export data</b> tool (  ) at the right of the Table view header. Select from the following:</p> <ul style="list-style-type: none"> <li>• <b>Matches without Gaps</b> – To export an ungapped FASTA (.fasta) file consisting of matching regions of displayed matches, but not the query sequence itself.</li> <li>• <b>Aligned Matches</b> – To export a gapped FASTA (.fasta) file with displayed matches aligned against the query.</li> <li>• <b>Table</b> – To save <a href="#">Table</a> view contents in either Comma- or tab-separated value (.csv or .tab) formats.</li> <li>• <b>Text</b> – To save the <a href="#">Text</a> view contents in Text (.txt) format.</li> </ul>
Consensus sequences	<p>Select one or more contigs in the <a href="#">Explorer panel</a>, then choose <b>File &gt; Export Data &gt; Consensus &gt; (Contig Name)</b>, <b>File &gt; Export Data &gt; Consensus &gt; Selected Contigs</b> or <b>Contig &gt; Export Consensus of (Contig Name)</b>. Alternatively, from the <a href="#">Alignment view</a>, use the <b>Export data tool</b> (  ).</p> <p>In each case, the following dialog opens:</p>



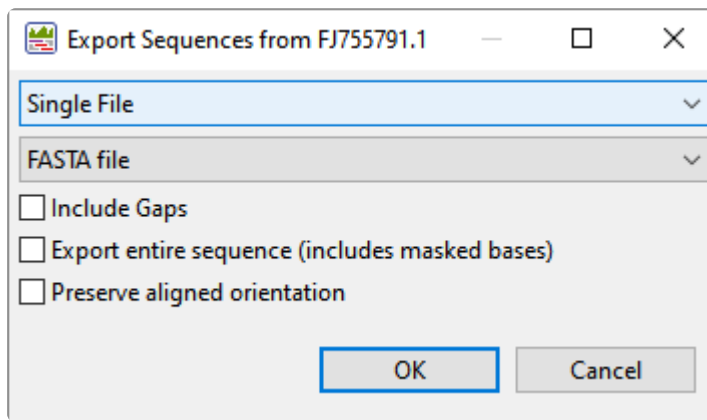
- In the upper drop-down menu, choose **Single File** to save the reads in a single file with the name of your choice. Or choose **Multiple Files** to save reads in separate files.
- Use the lower drop-down menu to choose the save format.
- If you want SeqMan Ultra to remove consensus gaps when the contig(s) are saved, keep **Include gaps** unchecked (recommended). Check the box only if you want the gaps in the contig(s) to be retained.
- Check **Include Reference Only Bases as 'N'** to export all bases that were only found in the reference sequence (not any of the constituent sequences) as Ns. An example in *.fasta* format is shown below.

[illegible]

- Check **Include Scaffold Position Information** to include the position number in the header of the exported data file. If you check this box, the file header will including the text: */dnas\_scaffold\_POS=x* where 'x' is the position of the contig in the scaffold.

When you have finished choosing the desired settings, click **OK**.

<b>Reads from a contig</b>	Select one or more contigs in the <a href="#">Explorer panel</a> and choose <b>Contig &gt; Export Sequences of (Contig Name)</b> .
----------------------------	--




- In the upper drop-down menu, choose **Single File** to save all selected read sequences in a single file with the name of your choice. Each contig in the file will begin with a note listing the contig name, scaffold ID and scaffold position. Or choose **Multiple Files** to save each selected contig as a separate file that is automatically assigned the same name as the contig.
- Use the lower drop-down menu to choose the save format.
- If you want SeqMan Ultra to include gaps when the contigs are saved, check the box next to **Include Gaps**.
- To include masked bases in the exported file, check the box next to **Export entire sequence (includes masked bases)**.
- To keep the orientation that was used when the read was aligned, check the box next to **Preserve aligned orientation**.

When you have finished choosing the desired settings, click **OK**.

# Save, Close and Exit

To save, close, or exit:

The following table describes how to save files, close projects and exit from SeqMan Ultra:

Task	How to...
To save all of the open projects simultaneously	Choose <b>File &gt; Save All</b> or press <b>Ctrl/Cmd+Shift+S</b> .
To save the current project	Choose <b>File &gt; Save</b> , select the <b>Save</b> tool (  ) , or use <b>Ctrl/Cmd+S</b> . If you want to preserve the earlier version of a project, or store a copy of the project with a different name or location, use <b>File &gt; Save As</b> instead. By default, SeqMan Ultra saves to the desktop unless directed to save elsewhere.
To close an active project	Choose <b>File &gt; Close</b> or <b>Ctrl/Cmd+W</b> . If multiple projects are open, first select the tab corresponding to the project you wish to close.
To exit from SeqMan Ultra	On Windows, use <b>File &gt; Exit</b> . On Macintosh, use <b>SeqMan Ultra &gt; Quit SeqMan Ultra</b> . The shortcut <b>Cmd+Q</b> works on both platforms.

Closing or quitting when there are unsaved documents:

If you select **File > Close**, or if you quit SeqMan Ultra when an unsaved project is open, SeqMan Ultra will open the Save Resource dialog. Click **Yes** to save changes, **No** to close/quit without saving changes, or **Cancel** to return to SeqMan Ultra without closing/quitting.

# Appendix

---

The Appendix contains the following topics:

- [Troubleshoot the error](#)
- [IUPAC codes](#)
- [Supported file types](#)
- [Installed Lasergene file locations](#)
- [Research references](#)

# Troubleshoot the error “A different version is running”

---

If you receive the error “A different version of Lasergene is running” when you try to launch a Lasergene application, but another version of Lasergene actually isn’t running, there may be a lingering .state file from another version that needs to be deleted.

The solution is to locate and delete files named *STARDM\*.state*, ...where ‘\*’ represents the version number.

These files may exist in the following directories:

## Windows 7 and Windows 8:

- *C:\Users\username\AppData\Local\DNASTAR\DataManager*
- *C:\ProgramData\DNASTAR\DataManager*

## Macintosh:

- *Hard Drive:Library:Preferences:DNASTAR:DataManager*
- *Hard Drive:Users:username:Library:Preferences:DNASTAR:DataManager*

If you find any of the .state files listed above, delete them. Once the files are deleted, the error should no longer appear.

# IUPAC codes

---

## IUPAC codes for amino acids:

1-Letter Code	3-Letter Code	Amino Acid
A	Ala	Alanine
B	Asx	Aspartic Acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
J	Xle	Leucine or Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
O	Pyl	Pyrrolysine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
U	Sec	Selenocysteine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Unspecified or unknown
Y	Tyr	Tyrosine
Z	Glx	Glutamic Acid or Glutamine

## IUPAC codes for nucleotides:

Symbol	Description	Bases Represented			
A	adenosine/adenine	A			
C	cytidine/cytosine		C		
G	guanosine/guanine			G	
T	thymidine/thymine				T
U	uridine/uracil				U
W	weak	A			T
S	strong		C	G	
M	amino	A	C		
K	keto			G	T
R	purine	A		G	
Y	pyrimidine		C		T
B	not A		C	G	T
D	not C	A		G	T
H	not G	A	C		T
V	not T	A	C	G	
N or -	any base (not a gap)	A	C	G	T

✿ The non-standard nucleotide codes “I” (Inosine) or “X” (xanthine) are not supported.

# Supported file types

---

For a list of file formats supported by this DNASTAR product, please see the [File Formats](#) page of our website.

# Installed Lasergene file locations

The following file names use 'x' to represent the version number.

File Category	Application	Path
Application <sup>2</sup>	ArrayStar	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x\ ArrayStar
	SeqNinja (command line)	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x\ SeqNinjaCL
	All others	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x Macintosh: /Applications/ DNASTAR/Lasergene x
Data Manager <sup>2</sup> ( <i>DMx</i> , <i>DMx.exe</i> )	SeqBuilder Pro, Protean 3D, GeneQuest, MegAlign Pro	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x Macintosh: /Applications/ DNASTAR/Lasergene x
Data Manager State File	SeqBuilder Pro, Protean 3D, GeneQuest, MegAlign Pro	Windows: C:\Program Data\ DNASTAR\DataManager Macintosh: ~/Library/ Application Support/ DNASTAR/DataManager, ~/Library/Preferences/ DNASTAR/DataManager
License Manager <sup>2</sup>	All	Windows: C:\Program Files (x86)\DNASTAR\License Manager Macintosh: /Applications/ DNASTAR/License Manager
Server License File ( <i>Iservrc</i> ), Server License Manager <sup>2</sup> , <i>Server Executables</i> ( <i>_Iserv</i> , <i>Iservnt.exe</i> )	All	Windows: C:\Program Files (x86)\DNASTAR- LicenseServer\Server Macintosh: ~/Library/ DNASTAR-LicenseServer
Standalone & Trial Licenses ( <i>*.license</i> ), License Server Client License ( <i>*.lshost</i> ), Key Server Client License ( <i>*.keyhost</i> )	All	Windows: C:\Program Data\ DNASTAR\Licenses Macintosh: ~/Library/ Application Support/ DNASTAR/Licenses

Preferences	Protean 3D, Navigator, SeqNinja (DNA*), GenVision Pro	Windows: C:\Users\<user>\DNASTAR  Macintosh: ~/Library/DNASTAR
	ArrayStar	Windows: C:\Users\<User>\AppData\Roaming\DNASTAR\ArrayStar
	All others	Windows: C:\Users\<user>\AppData\Local\DNASTAR\  Macintosh: ~/Library/Preferences



AppData is a hidden folder in Windows. To unhide the folder, go to **Organize > Folder and Search Options > View > Show Hidden files and folders**.

# Research references

---

Allex, C.F., Baldwin, S.F., Shavlik, J.W., and Blattner, F.R. (1996). Improving the quality of automatic DNA sequence assembly using fluorescent trace-data classifications. *Proceedings, Fourth International Conference on Intelligent Systems for Molecular Biology*, 3-14. St. Louis, MO. Menlo Park, CA. [ISMB-96 Proceedings](#), AAAI Press.

Allex, C.F., Baldwin, S.F., Shavlik, J.W., and Blattner, F.R. (1997). Increasing consensus accuracy in DNA fragment assemblies by incorporating fluorescent trace representations. *Proceedings, Fifth International Conference on Intelligent Systems for Molecular Biology*, 3-14. Halkidiki, Greece. Menlo Park, CA. [ISMB-97 Proceedings](#), AAAI Press.

Allex CF (1999). Computational Methods for Fast and Accurate DNA Fragment Assembly ([Doctoral thesis](#)). Department of Computer Sciences, University of Wisconsin-Madison.

Allex, C.F., Shavlik, J.W., and Blattner, F.R. (1999). Neural network input representations that produce accurate consensus sequences from DNA fragment assemblies. [Bioinformatics 15\(9\):723-728](#).

Altshuler *et al.* (2000). A SNP map of the human genome generated by reduced representation shotgun sequencing. [Nature 407, 513-516](#).

Anson E and Myers E (1997). "ReAligner: A program for refining DNA sequence multi-alignments." *Journal of Computational Biology* 4(3):369-383.

Blattner FR *et al.* (1997). The Complete Genome Sequence of Escherichia coli K-12. *Science* 05 Sep 1997: Vol. 277, Issue 5331, pp. 1453-1462. DOI: 10.1126/ science.277.5331.1453. [View on PubMed](#)

Davydov *et al.* (2010). "Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++." *PLoS Comput. Biol.* 6: e1001025.

Ewing B, Hillier L, Wendl MC and Green P (1998). "Base-calling of automated sequencer traces using Phred. I. Accuracy assessment." *Genome Research*, 8:175-185.

Ewing B and Green P (1998). "Base-calling of automated sequencer traces using Phred. II. Error probabilities." *Genome Research*, 8:186-194.

Martinez HM (1983). "An efficient method for finding repeats in molecular sequences." *Nucleic Acids Res.* 11: 4629-4634.

Needleman SB and Wunsch CD (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins." *J. Mol. Biol.* 48: 443-453.

Taylor WR (1997). "Residual colours: a proposal for aminochromography." *Protein Engineering* vol.10 no.7

pp.743 – 746. (See the [PDF file](#).) [“Taylor” color scheme]

Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ (2009). “JalView version 2: a multiple sequence alignment and analysis workbench.” *Bioinformatics* 25:1189 – 1191. (See on [PMC](#).) [Discusses methodology later used to create the “Taylor” and “Zappo” color schemes]