User Guide to



SeqMan Ultra™

DNASTAR, Inc. 2022

Table of Contents

SeqMan Ultra Overview **6**

SeqMan Ultra Tutorials **7**

Whole genome reference-guided workflow 8

Whole genome de novo workflow with mate pair data 18

Analysis of a whole genome de novo assembly 23

De novo assembly using Sanger data 27

RNA-Seq de novo transcriptome workflow 32

Part A: Setting up the transcriptome assembly in SeqMan NGen 33

Part B: Viewing annotated transcripts in SeqMan Ultra 36

Gap closure workflow 39

Reference-guided alignment of Sanger/ABI data 42

Getting Started **43**

SeqMan Ultra Interface **44**

Welcome screen 45

Views 47

Unassembled view 49

Project Report view 55

Project Overview 56

Project Details 58

Project Statistics 61

Trimming 63

Alignment View 65

Alignment view toolbar 68

View multi-sample assemblies in the Alignment view 71

Default colors in the Alignment view 73

Strategy View 75

Strategy view toolbar 78

Transcriptome view 79

Contig Report View 82

Contig Information 83

Contig Coverage 85

Contig Structural Variation 87

Features view 90

Variants view 93

Filtering in the Variants view 98

Table view 101

Text view 103

Pairwise view 105

Panels ... 107
Explorer panel ... 108
Contigs and Scaffolds section ... 109
Sequences section ... 111
Style panel ... 112
Sequence section ... 114
Alignment section ... 115
Available color schemes ... 119
Strategy section ... 123
Tracks panel ... 124
Layout section ... 128
Options section ... 129
Details panel ... 130
View information about a selected feature ... 132
View information about a selected track ... 133
View information about multiple selected tracks ... 134
Jobs panel ... 135

Assemble Sequences ... 137
Assemble small-to-medium sets of Sanger data ... 138
Quality score calculations ... 142
End-trimming based on averaged quality scores ... 146
Assemble multi-sample Sanger data ... 147
Assemble NGS data or large sets of Sanger data ... 148

Open an Existing Assembly ... 149
Navigate between open assemblies ... 151

Contigs ... 152
Organize contigs and scaffolds ... 153
Edit a contig ... 157
Close a gap between contigs ... 159
Display contig information in a graphical or tabular view ... 162
Rename a contig ... 164
Reassemble contigs ... 165

Features ... 166

Tracks ... 167
Ruler track ... 170
Contig track ... 171
Coverage track ... 172
Pair consistency track ... 173
Features track ... 174
Split reads track ... 175

Translation track 176
Trace track 178
Majority track 179
Sample-alignment track 180
Rename a track manually 181
Rename tracks automatically using specified data fields 182

Customize the Look and Layout 184
Customize the appearance and layout of views 185
Customize tables in the views 187
Descriptions of table data columns 189
Customize the layout of other window components 199
Save the layout and apply it to other projects 201

Search Within the Project 202
Search for and select a read sequence by name 203
Search for elements within a contig 204
Search for a numerical position on a contig 206

Search for Sequences Online 207
Search for sequence matches (BLAST) 208
Search wizard: Query 210
Search wizard: Search 213
Search wizard: Options 215
Search wizard: Job 218
Search for text or accession number matches (Entrez) 219
Filter sequence search results 222
Filter text search results 224
Access ID dialog 225

Make a Selection 226

Specify Parameters and Preferences 227
Consensus parameters 228
Consensus calling algorithms 230
Strategy Viewing and Coverage parameters 231
Pair Specifier parameters 232
Assembly parameters 233

Export 235
Export an image of a view 236
Export data to a file 238
Export data from the Variants view 241
Export data from the Text, Table or Pairwise views 242

Save, Close and Exit 243

Appendix **244**
Troubleshoot the error “A different version is running” 245
IUPAC codes 246
Supported file types 248
Installed Lasergene file locations 249
Research references 251

SeqMan Ultra Overview

If you are working with Sanger data, SeqMan Ultra serves as an application for performing and analyzing Sanger assemblies. If you are working with NGS data, SeqMan Ultra is a workflow launcher for the SeqMan NGen assembler, and the place to perform downstream analysis after assembly.

SeqMan Ultra can be used for downstream analysis of 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**.

Copyright DNASTAR, Inc. 2022

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	Whole genome reference-guided workflow
2	Whole genome *de novo* workflow with mate pair data
3	Analysis of a whole genome *de novo* assembly
4	*De novo* assembly using Sanger data
5	RNA-Seq *de novo* transcriptome workflow
6	Gap closure workflow

SeqMan Ultra also plays an important role in the tutorial Phage display assembly, translation and prediction, located in the Protean 3D User Guide. This integrated workflow starts in SeqMan Ultra, continues through SeqNinja and MegAlign Pro, and ends in Protean 3D.

Copyright DNASTAR, Inc. 2022

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**.

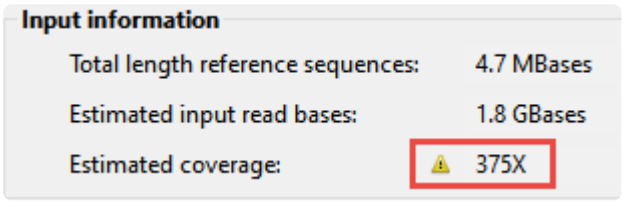
5. In the Input Sequences screen, change the **Experiment setup** to **Single sample**. Press the **Add** button to add the paired reads *SRR1284938_1.fastq* and *SRR1284938_2.fastq*. Then click **Next**.

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.”

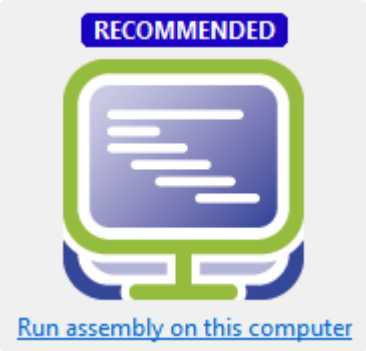


Copyright DNASTAR, Inc. 2022

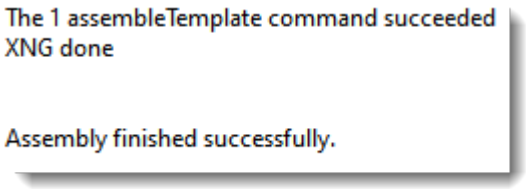
A coverage of 50-100X is ideal, and additional coverage only serves to increase the assembly time. There are several wizard parameters you can change to reduce coverage; here, you'll reduce the **Maximum total reads**.

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.

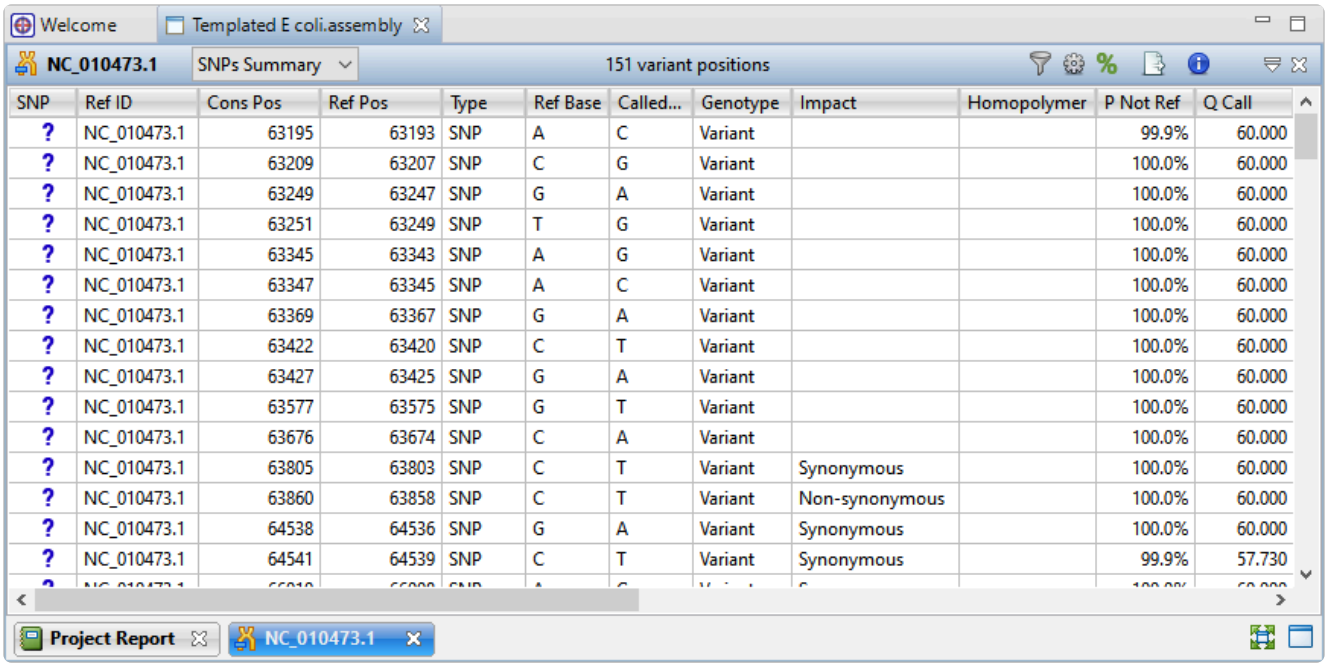
14. Close the SeqMan NGen project by clicking the **Finish** button and confirming **Yes**.

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.

2. 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.

Copyright DNASTAR, Inc. 2022

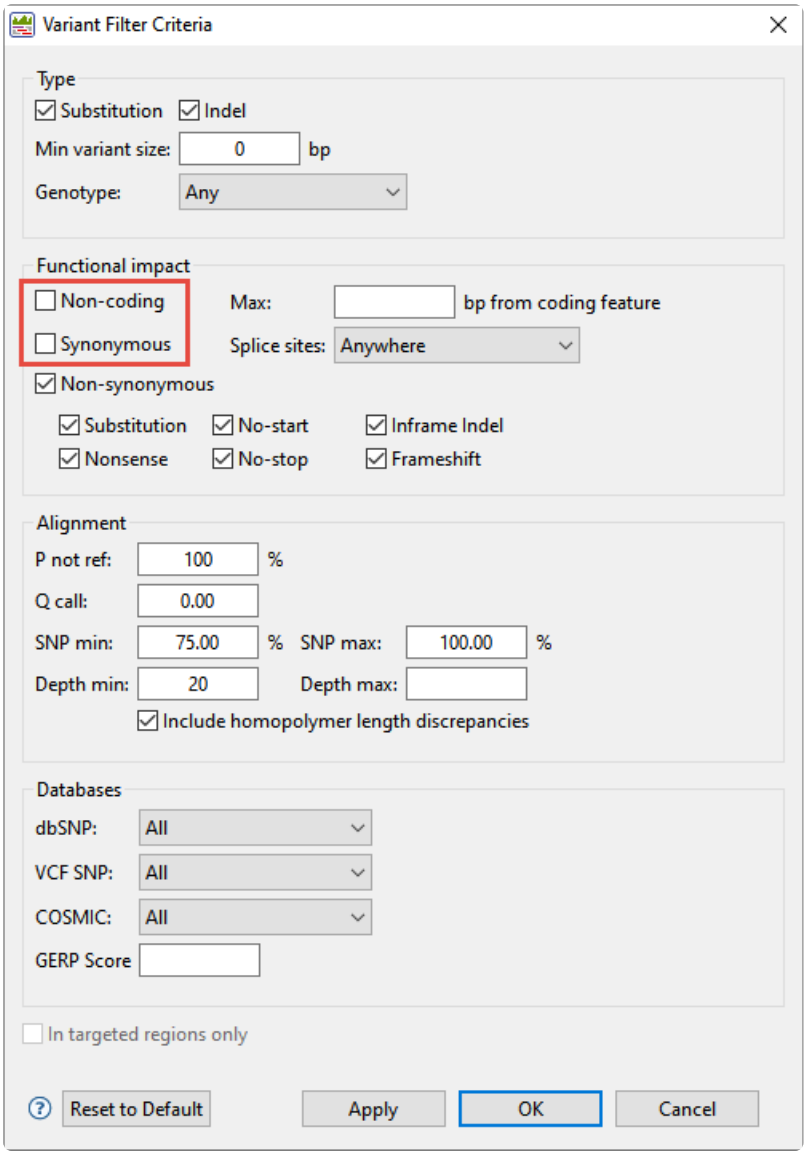


SNP	Ref ID	Cons Pos	Ref Pos	Type	Ref Base	Called...	Genotype	Impact	Homopolymer	P Not Ref	Q Call
?	NC_010473.1	63195	63193	SNP	A	C	Variant			99.9%	60.000
?	NC_010473.1	63209	63207	SNP	C	G	Variant			100.0%	60.000
?	NC_010473.1	63249	63247	SNP	G	A	Variant			100.0%	60.000
?	NC_010473.1	63251	63249	SNP	T	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.

3. 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**.

Copyright DNASTAR, Inc. 2022

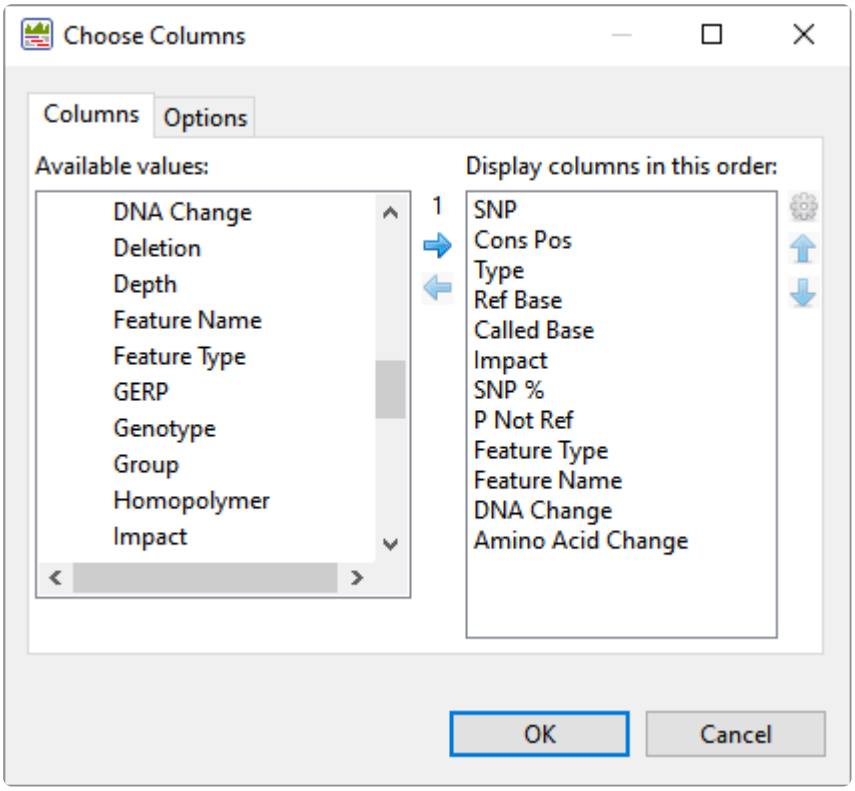


The remaining 22 variants all have a non-synonymous impact or consist of an in-frame deletion.

4. 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.

5. 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**.

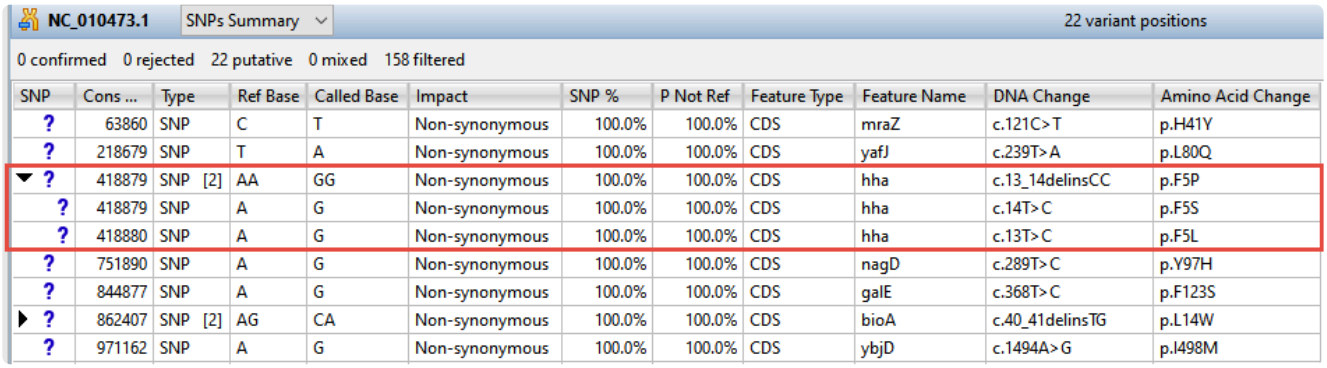
Copyright DNASTAR, Inc. 2022



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.

6. 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.



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

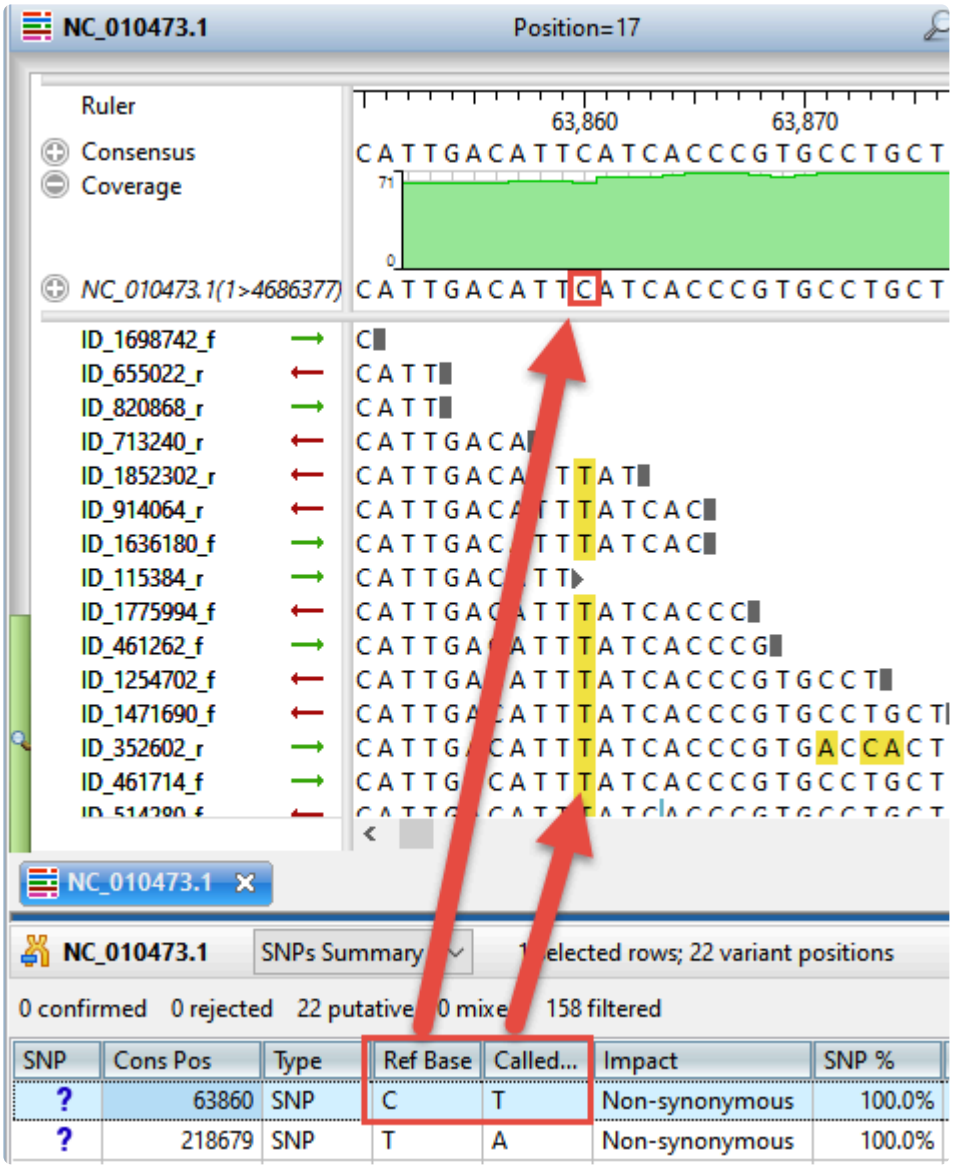
7. Look at the **Impact** column for the variant in the first row (Cons Pos = 63860), and note that a non-synonymous mutation is predicted.

Copyright DNASTAR, Inc. 2022

8. 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.

9. 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.



   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.

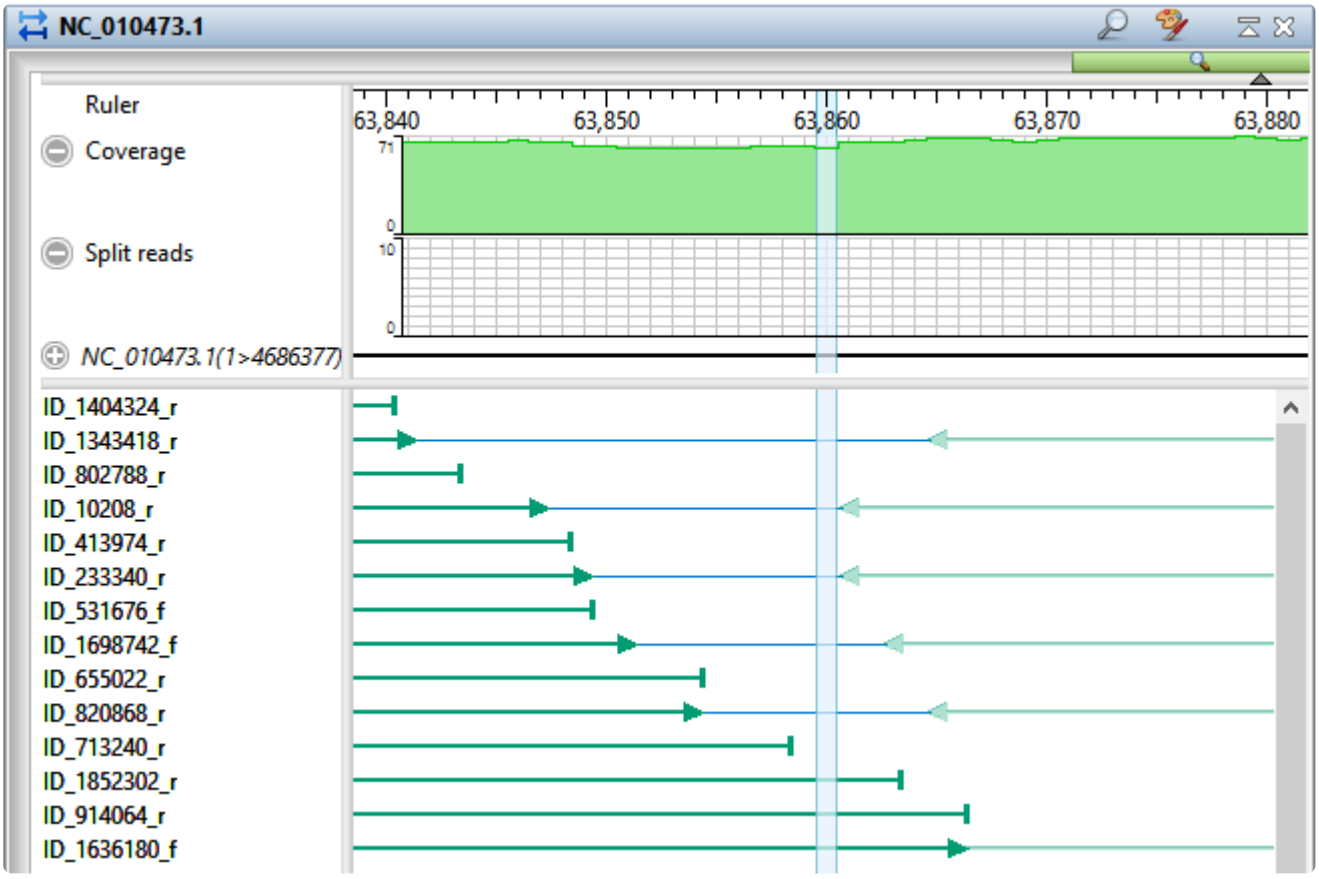
10. To view the variant in the Strategy view, select **View > Strategy > Show Strategy View**. Because two views are already open, the Strategy view appears in a pop-up window. The view graphically

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.

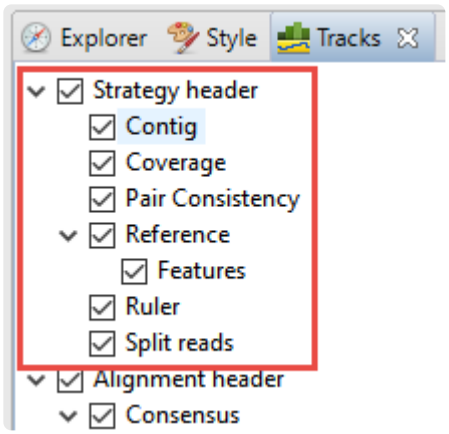
12. Click on the top row of the Variants view table to again move the cursor to the variant of interest.

    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.



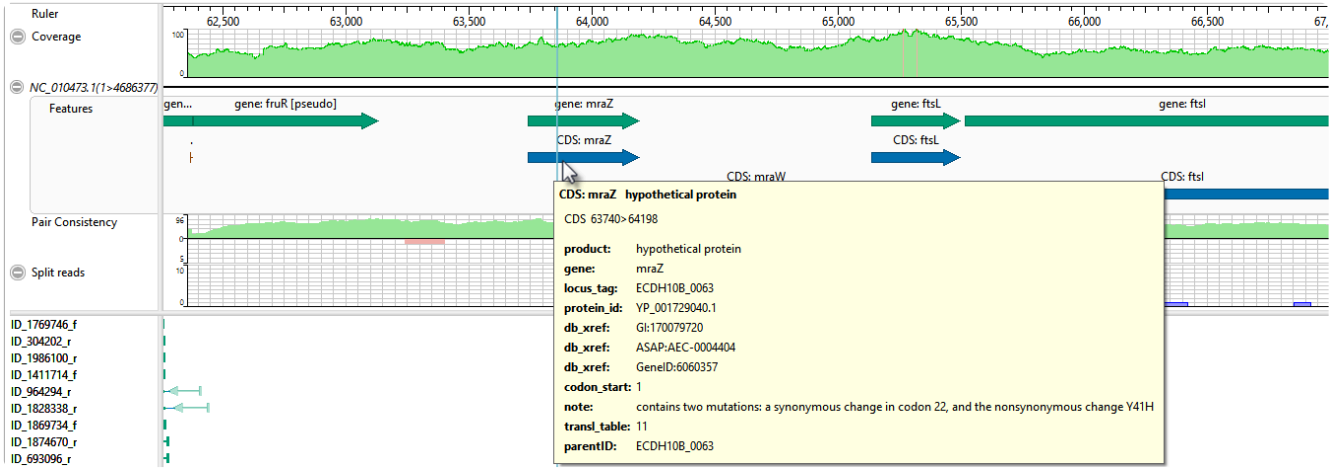
13. To add data tracks to enhance the Strategy view, click on the Tracks tab and check all boxes in the "Strategy header" section.

Copyright DNASTAR, Inc. 2022



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.

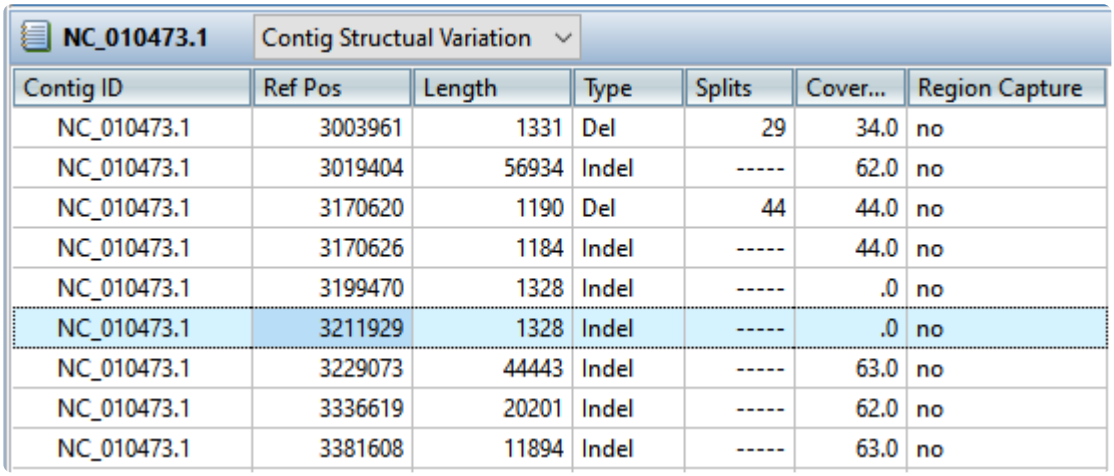
17. Confirm the variant that you just investigated by clicking on the **SNP** (leftmost) column to change the question mark to a checkmark. (If you had found supporting evidence was lacking, you could have

Copyright DNASTAR, Inc. 2022

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.

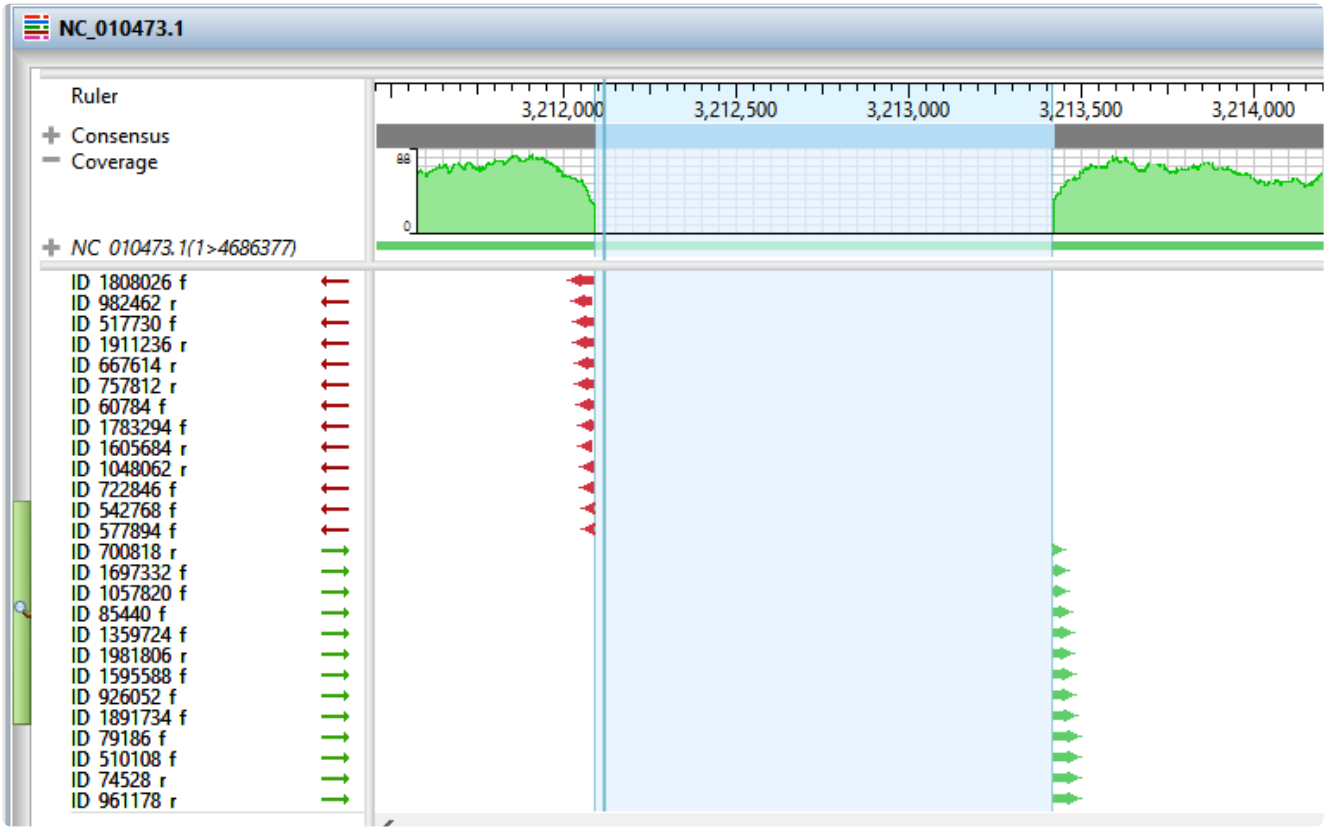
18. To see the longer indels, choose **Contig > NC_010473.1 Structural Variation**. Scroll down to the putative Indel at 3211929.



NC_010473.1 | Contig Structual 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.

Copyright DNASTAR, Inc. 2022

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**.

Copyright DNASTAR, Inc. 2022

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**.

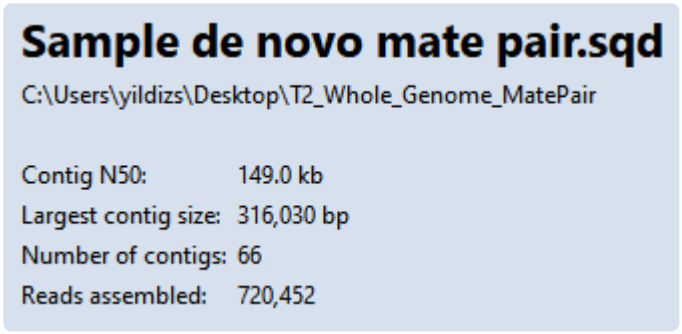
Correcting misjoined contigs and realigning data in SeqMan Ultra:

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*.

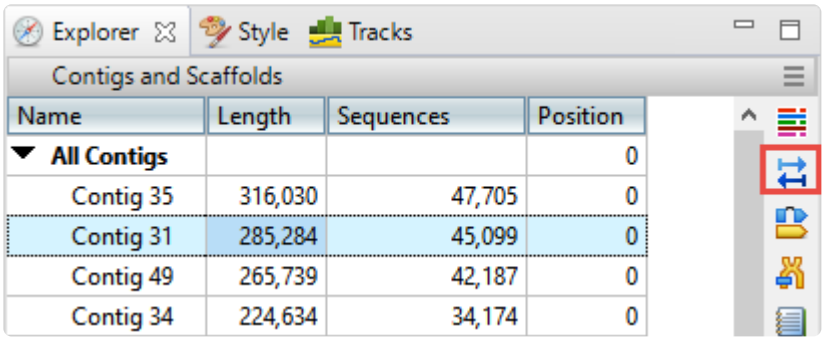
3. Observe the information in the Project Overview on your screen. Briefly note the high **Contig N50** value of 149 kb. This means that half of the genome sequence is located in contigs equal to or longer than 149 kb.



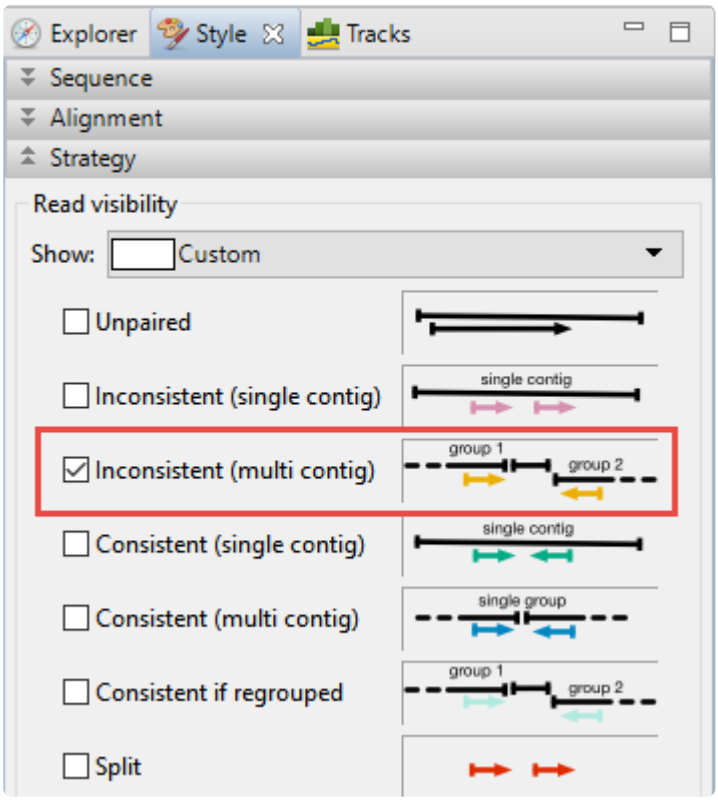
   Using a high stringency in SeqMan NGen would likely have resulted in about ten times as many contigs. However, a tradeoff is that some contigs might have been joined when they should not have been. You will address that issue in the next few steps. In "real life," you would check all of the contigs for 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.

Copyright DNASTAR, Inc. 2022

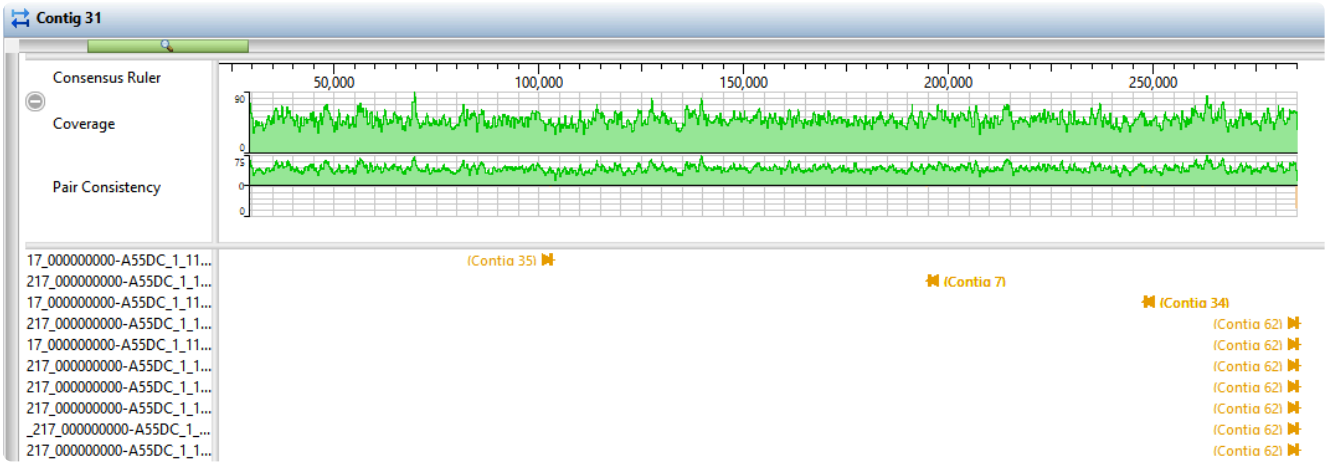


5. You will be looking for locations along the contig that are flagged for inconsistent pairs. To easily see inconsistent pairs, click the **Style** tab on the right. In the **Strategy** section, uncheck all boxes except **Inconsistent (multi-contig)**.



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.

Copyright DNASTAR, Inc. 2022



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 uppermost 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.

Copyright DNASTAR, Inc. 2022

14. 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.

15. 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.

16. Choose **Project > Project Overview** and observe how the steps above have improved the assembly, even though you only split one false join.



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.

Copyright DNASTAR, Inc. 2022

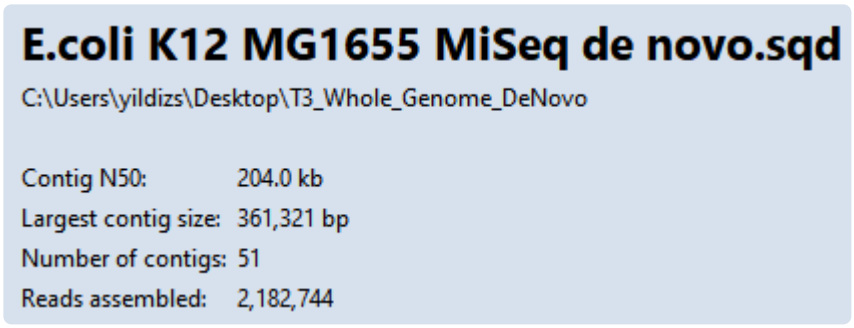
Analysis of a whole genome de novo assembly

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.



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 assembly the reads. In the upper section of the report, observe that there are an average of 42798 sequences per contig and 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.

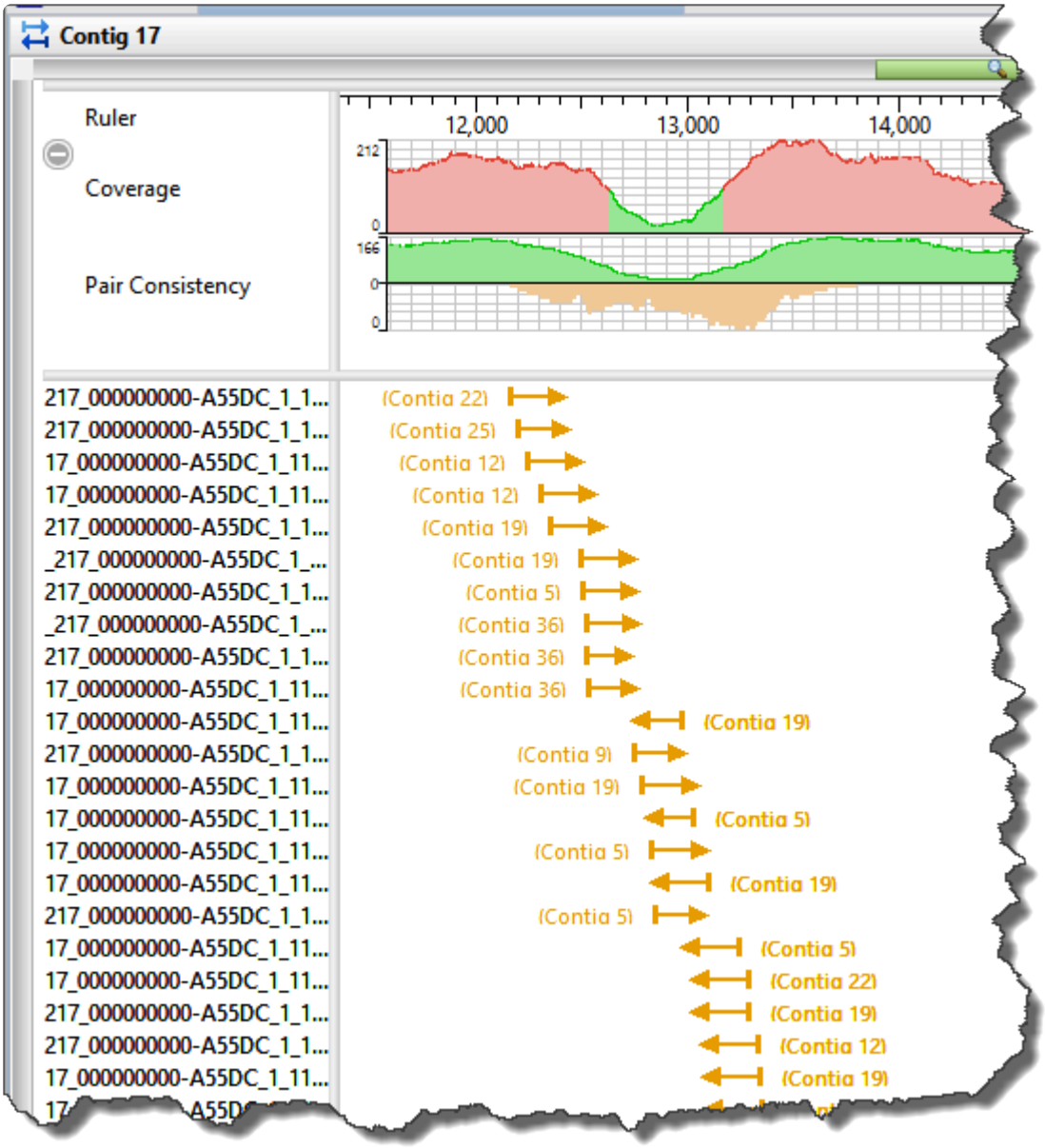
4. In the Explorer panel on the right, select **Contig 17** and press the **Show strategy view** tool ( ) on the right.

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

Copyright DNASTAR, Inc. 2022

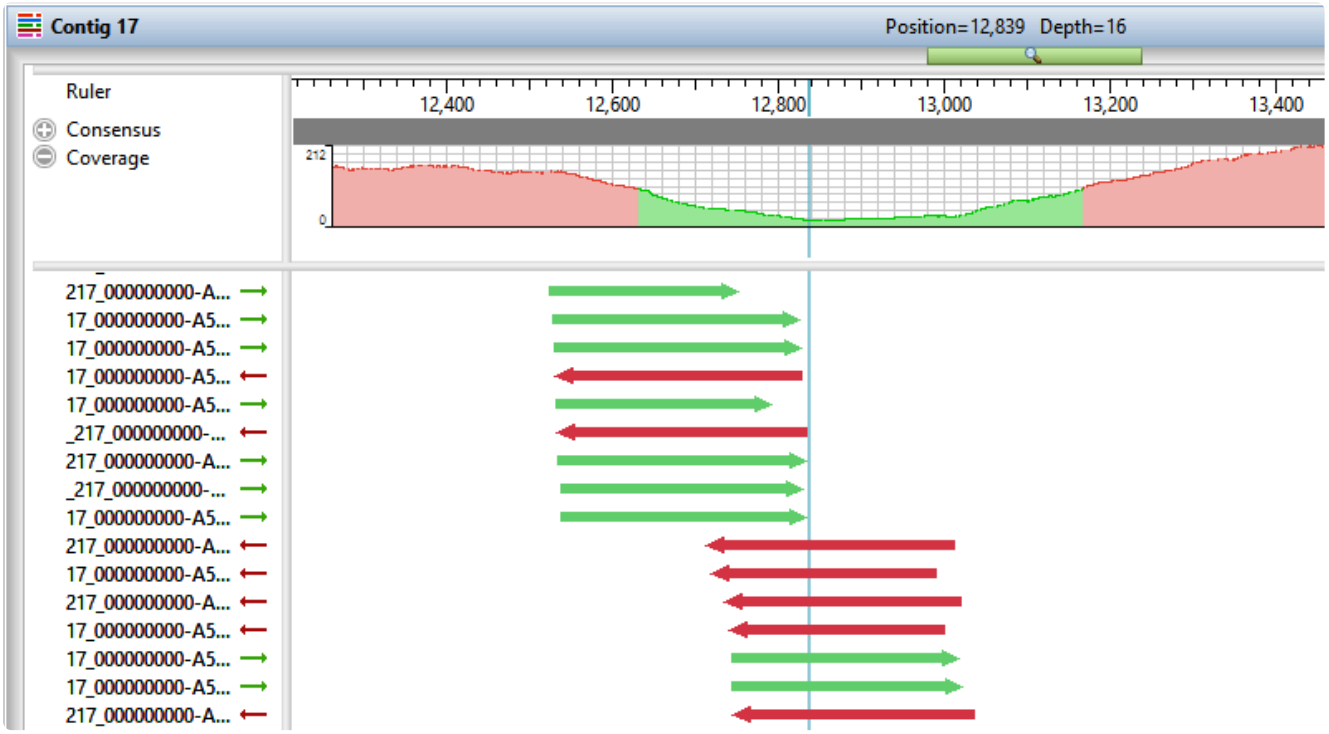
indicate the contig where their “mate” resides.

6. 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.



7. 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.

Copyright DNASTAR, Inc. 2022



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.

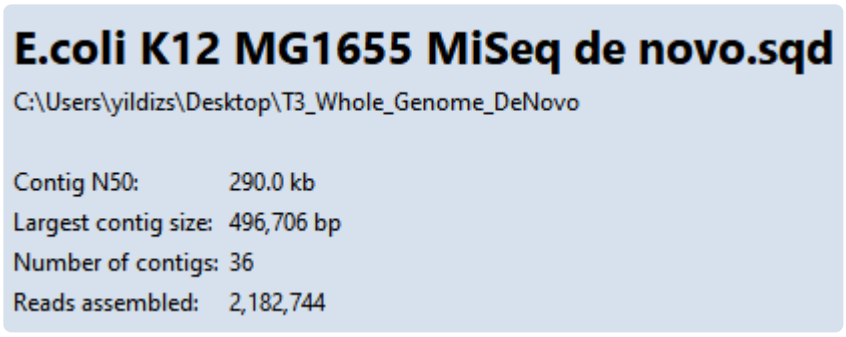
    Fourteen scaffolds are created, numbered from 100-230. The reason that microbial genomes do not assemble into a single contig (when using short reads) is a high prevalence of repetitive elements, like transposons that occur throughout the genome.

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 cach 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**.

Copyright DNASTAR, Inc. 2022



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.

Copyright DNASTAR, Inc. 2022

De novo assembly using Sanger data

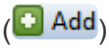
In this tutorial, you will *de novo* assemble fourteen short trace sequences from PE Applied Biosystems, Inc. and then analyze the resulting contig in SeqMan Ultra. Scroll to the bottom of this topic for a short video showing a slightly different version of this tutorial using the same data set.

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.

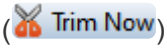
2. Launch SeqMan Ultra. On the left of the window, choose the blue bar named **New Assembly**.

3. Under **Molecular biology**, click on **New Sanger/ABI assembly**. The “Unassembled” window opens.

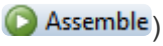
4. Press the **Add** button near the top left (

), 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**.

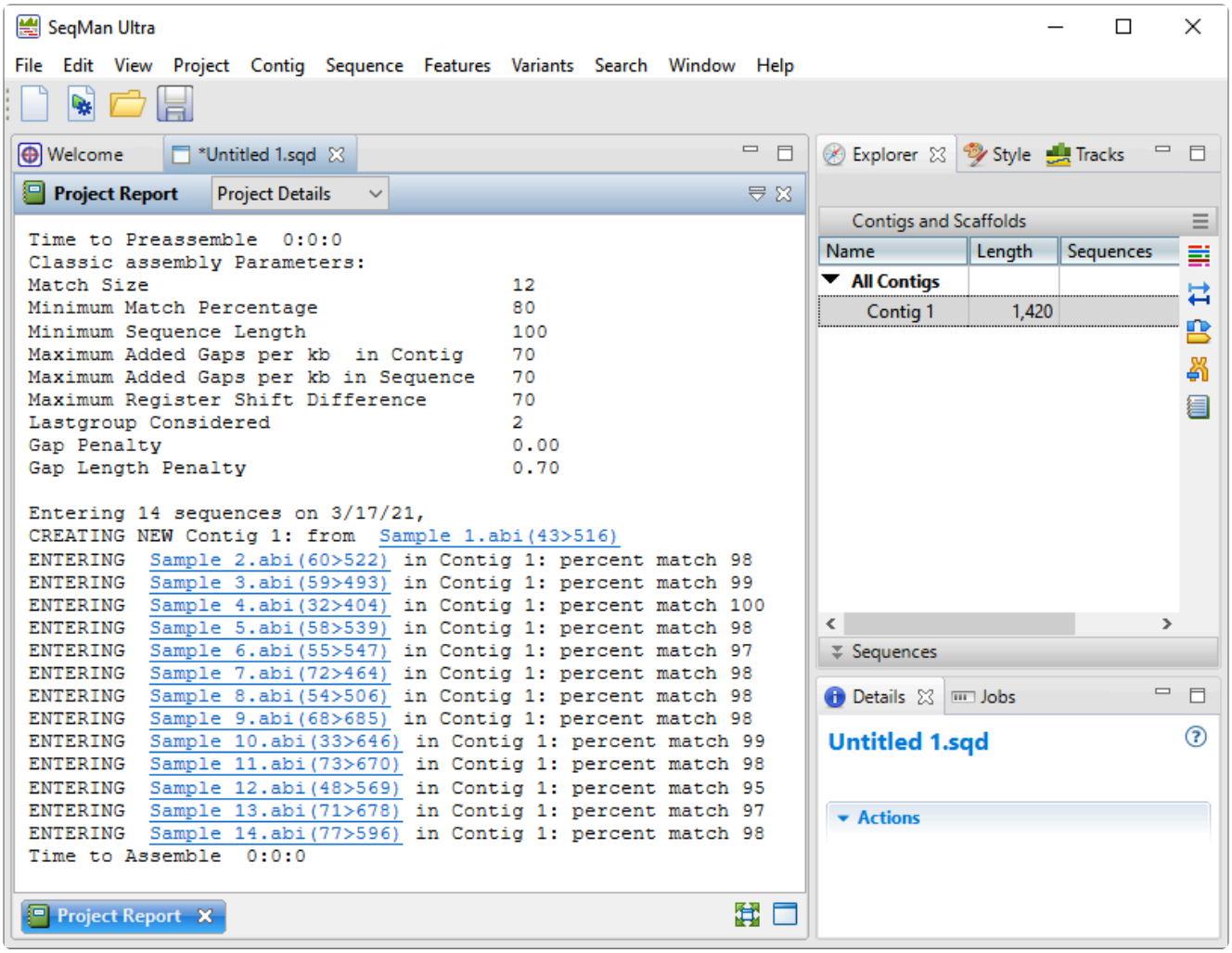
6. Press the **Trim Now** button (

). In the **Vector** column in the table, observe that the vector *InvJanus* was discovered and removed.

7. Press the **Assemble** button (

) to begin the assembly. The assembly should take a few seconds, at the most.

Copyright DNASTAR, Inc. 2022

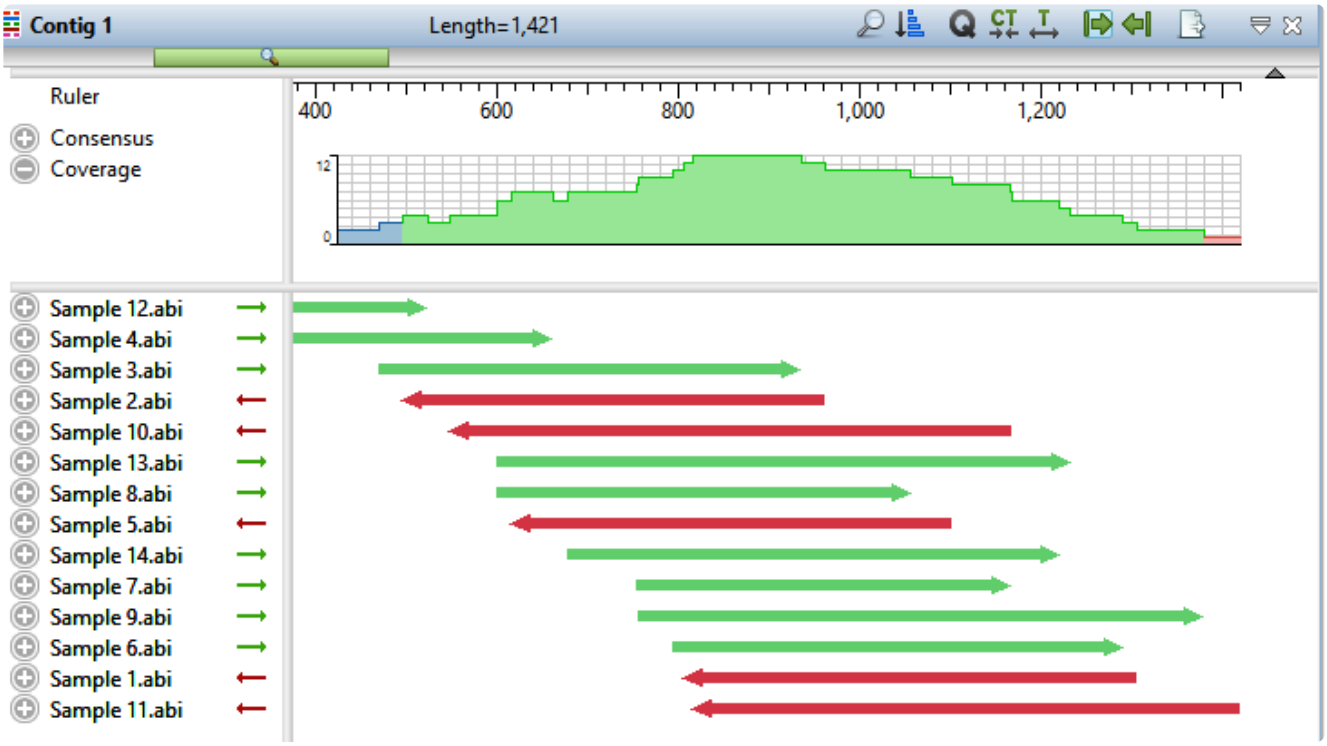


Examining the alignment:

1. In the Explorer tab on the right, note that the assembly resulted in one contig.

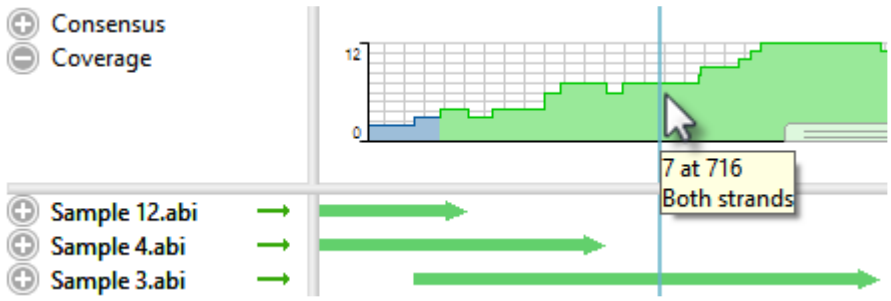
2. Double-click on **Contig 1** to open it in the Alignment view. Move the green horizontal zoom slider (near the top of the view) to the left until you can see the entire contig.

Copyright DNASTAR, Inc. 2022



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.

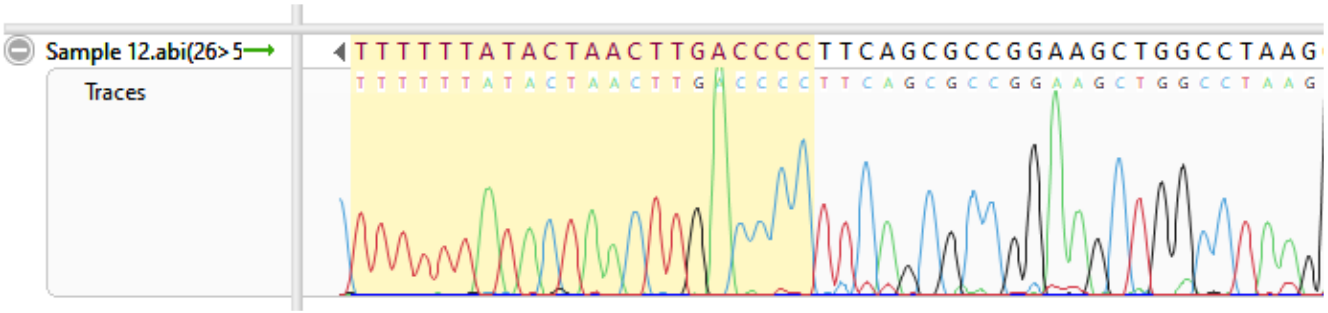
6. To view the trace data, right-click on any sample name on the left and choose **Expand All**.

7. Scroll all the way to the left so that only Sample 12.abi is visible.

Copyright DNASTAR, Inc. 2022

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.

Copyright DNASTAR, Inc. 2022

This marks the end of this tutorial.

Copyright DNASTAR, Inc. 2022

RNA-Seq de novo transcriptome workflow

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 names and annotates 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.

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 and then examine the results. This workflow uses an abbreviated yeast data set with about 1 million reads per file.

The workflow is in two parts, which should be done in order:

- Part A: Setting up the transcriptome assembly in SeqMan NGen
- Part B: Viewing annotated transcripts in SeqMan Ultra

Copyright DNASTAR, Inc. 2022

Part A: Setting up the transcriptome assembly in SeqMan NGen

In part A 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*.

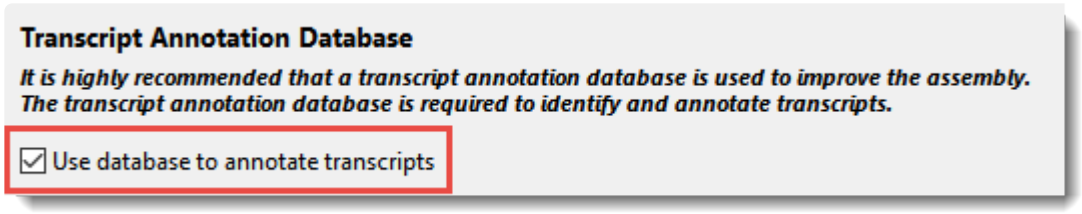
2. Launch SeqMan NGen and choose **RNA-Seq/Transcriptomics** on the left. On the right, click on **De novo transcriptome**.

3. 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**.

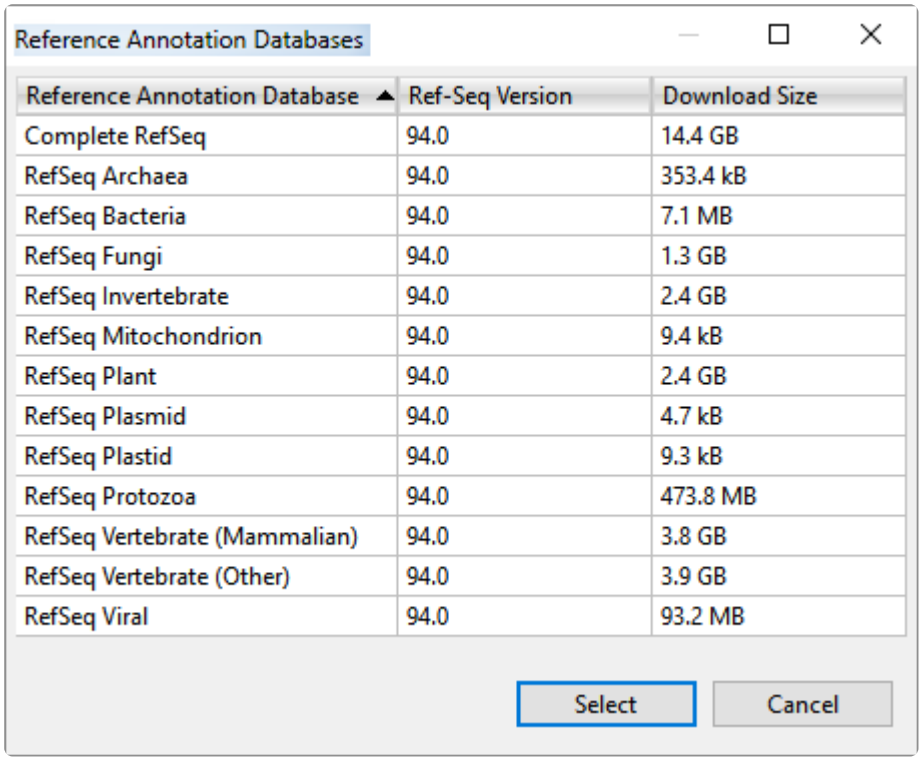
4. In the Input Sequences screen, press **Add** and add the *Yeast_RNASeq_1Mreads_1.fastq* and *Yeast_RNASeq_1Mreads_2.fastq* files. Click **Next**.

5. In the Transcript Annotation Database screen:

    a. Check the box next to **Use database to annotate transcripts**. **This step is very important!**



    b. Click the **Download Database** button to open the following popup.



c. Choose **RefSeq Fungi** and press **Select**.

d. Click **Next**.

6. In the Assembly Options screen, click **Next**.

7. 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**.

8. In the Run Assembly Project screen, note that:

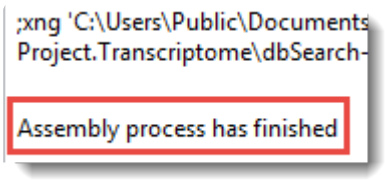
   - The estimated disk requirement of 2.1 TB is based on the total length of the fungal Transcript Annotation Database, which is 4.2 GB: larger than a human genome. That estimate is based on reference guided genome assemblies that have fixed 50X coverage, not reference guided transcriptome assemblies, 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.

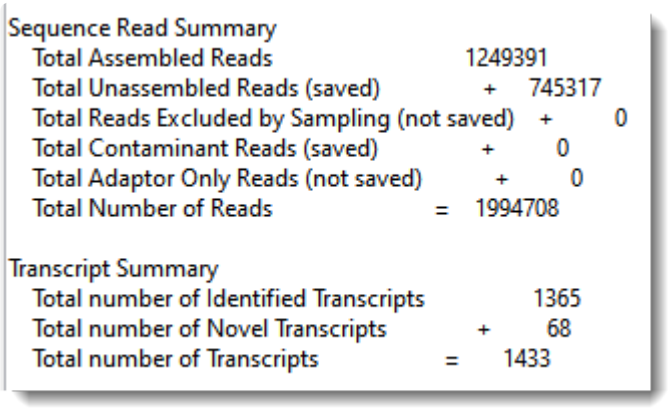
9. Wait until being informed that assembly has finished, then click **Next**.

Copyright DNASTAR, Inc. 2022



During assembly, 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.”

10. In the lower half of the Assembly Summary screen, do a quick check of the results. Your numbers will likely differ from those in the image below, but you should have at least 1300 **Identified Transcripts** and over 40 **Novel Transcripts**.



> **Note:** If **Total number of Identified Transcripts** is **0**, this indicates you forgot to check the box in step 5a, above. In this case, please set up and run the assembly again.

11. In the upper part of the same screen, click the button **View assembled transcripts** to open transcriptome results in SeqMan Ultra.

12. (optional), If desired, you may now return to SeqMan NGen and close that application by clicking **Finish** and then **Yes**.

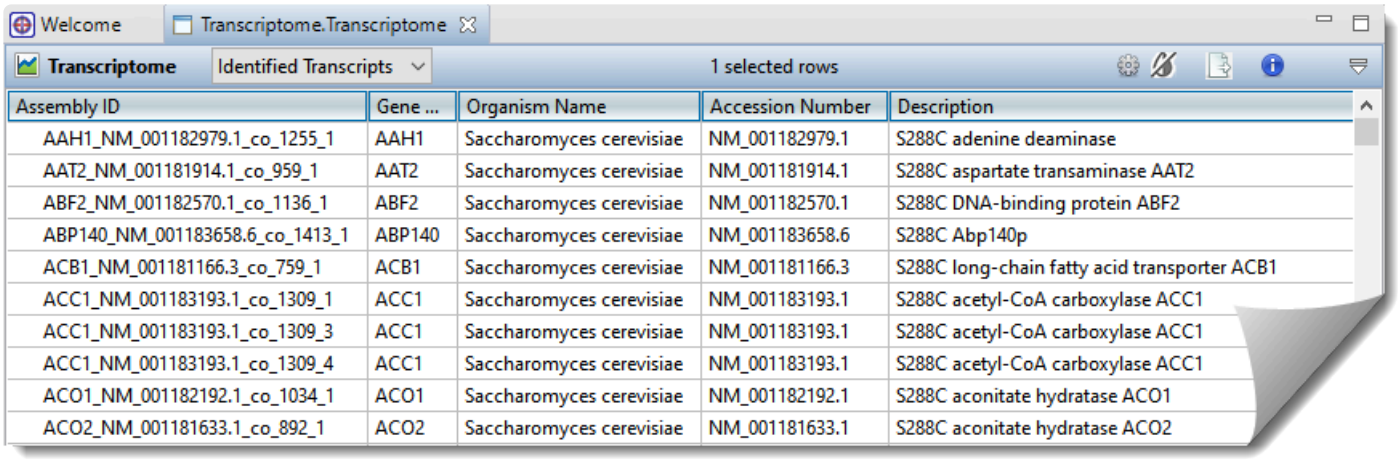
Proceed to Part B: Viewing annotated transcripts in SeqMan Ultra.

Copyright DNASTAR, Inc. 2022

Part B: Viewing annotated transcripts in SeqMan Ultra

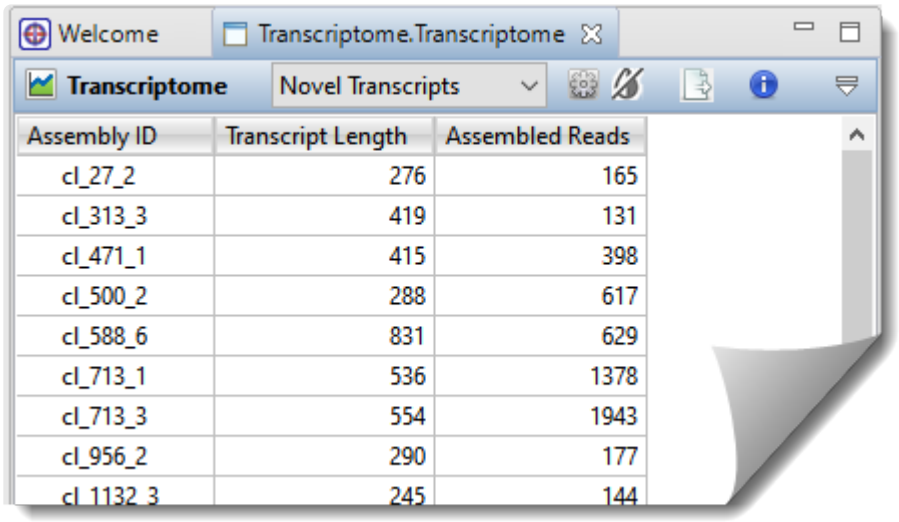
In Part A, you assembled transcriptome data in SeqMan NGen, which automatically annotated the transcripts during assembly. In this part of the tutorial, you will view the transcripts that were assembled and annotated using SeqMan NGen .

When you clicked the blue button in SeqMan NGen, the results opened automatically in SeqMan Ultra's Transcriptome view. The **Identified Transcripts** table is shown initially, and will contain over 1300 entries.



Assembly ID	Gene ...	Organism Name	Accession Number	Description
AAH1_NM_001182979.1_co_1255_1	AAH1	Saccharomyces cerevisiae	NM_001182979.1	S288C adenine deaminase
AAT2_NM_001181914.1_co_959_1	AAT2	Saccharomyces cerevisiae	NM_001181914.1	S288C aspartate transaminase AAT2
ABF2_NM_001182570.1_co_1136_1	ABF2	Saccharomyces cerevisiae	NM_001182570.1	S288C DNA-binding protein ABF2
ABP140_NM_001183658.6_co_1413_1	ABP140	Saccharomyces cerevisiae	NM_001183658.6	S288C Abp140p
ACB1_NM_001181166.3_co_759_1	ACB1	Saccharomyces cerevisiae	NM_001181166.3	S288C long-chain fatty acid transporter ACB1
ACC1_NM_001183193.1_co_1309_1	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288C acetyl-CoA carboxylase ACC1
ACC1_NM_001183193.1_co_1309_3	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288C acetyl-CoA carboxylase ACC1
ACC1_NM_001183193.1_co_1309_4	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288C acetyl-CoA carboxylase ACC1
ACO1_NM_001182192.1_co_1034_1	ACO1	Saccharomyces cerevisiae	NM_001182192.1	S288C aconitate hydratase ACO1
ACO2_NM_001181633.1_co_892_1	ACO2	Saccharomyces cerevisiae	NM_001181633.1	S288C aconitate hydratase ACO2

Now use the drop-down menu in the view header to choose **Novel Transcripts**. You should see >40 **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.

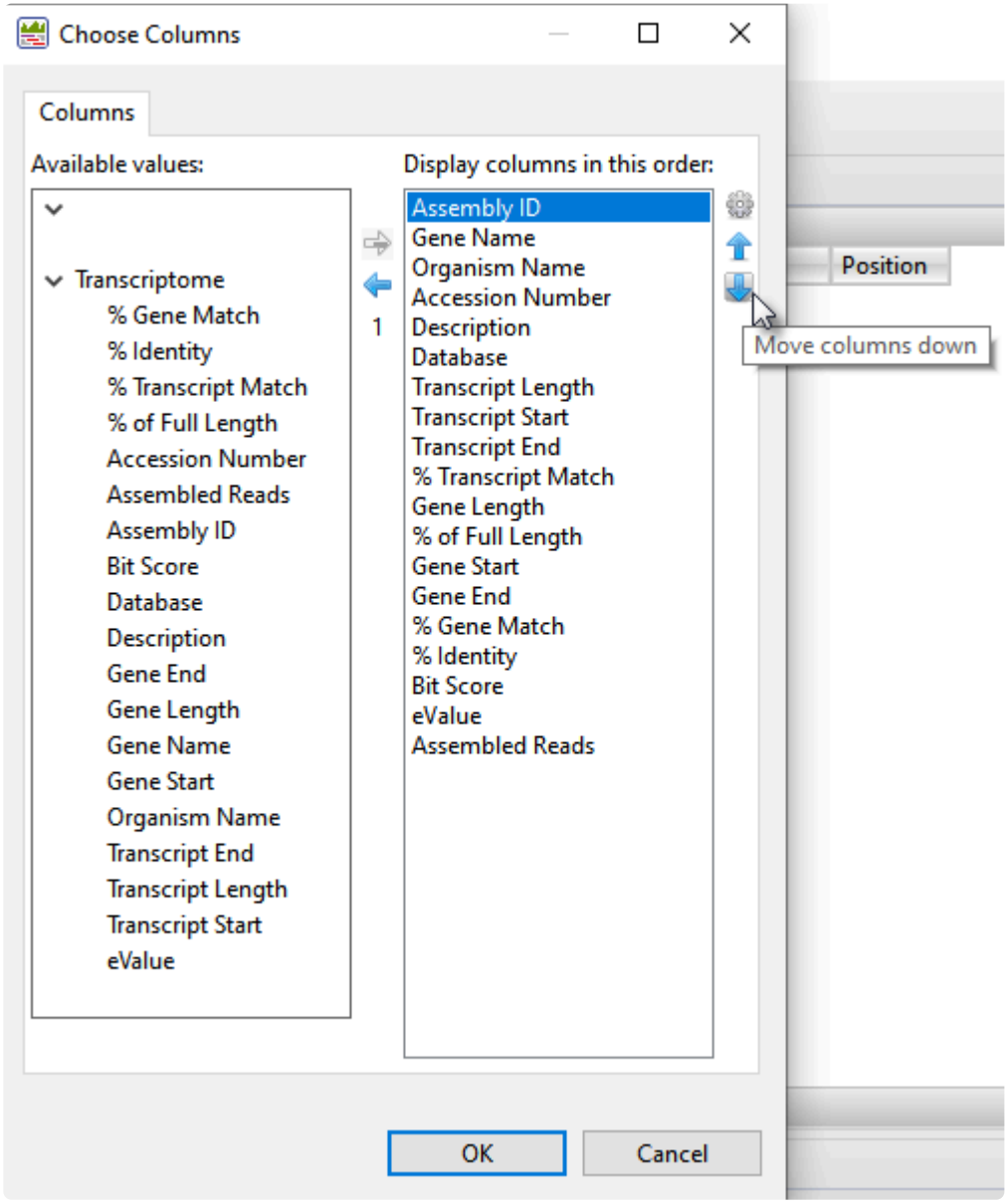


Assembly ID	Transcript Length	Assembled Reads
cl_27_2	276	165
cl_313_3	419	131
cl_471_1	415	398
cl_500_2	288	617
cl_588_6	831	629
cl_713_1	536	1378
cl_713_3	554	1943
cl_956_2	290	177
cl_1132_3	245	144

From within either Transcriptome view table, experiment with the following:

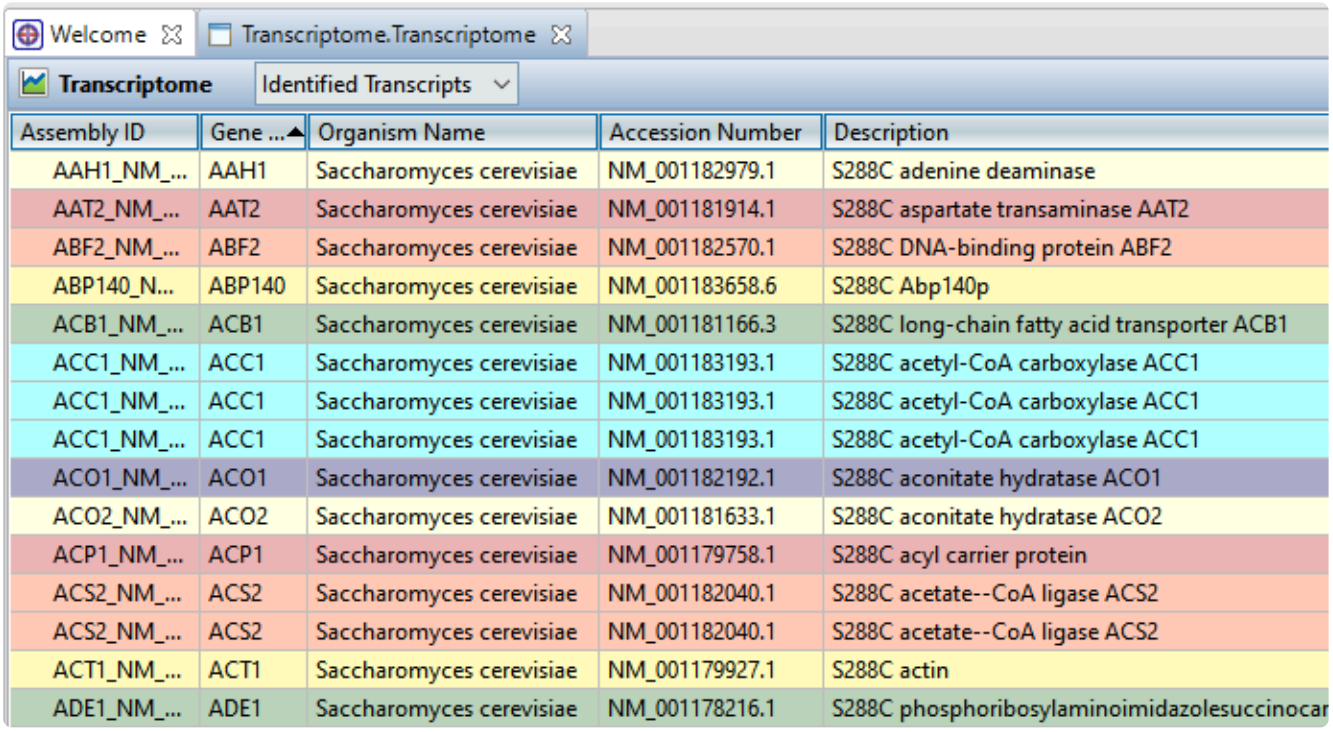
Copyright DNASTAR, Inc. 2022

- Both Transcriptome tables are customizable and can display a wide variety of sortable columns To show/hide columns or change their order, click the **Choose or rearrange columns** tool ( ) in the view header. Use the up/down/left/right arrows to choose which columns to display and their order. Descriptions of each column type can be seen in the table at the bottom of [this SeqMan Ultra User Guide topic]



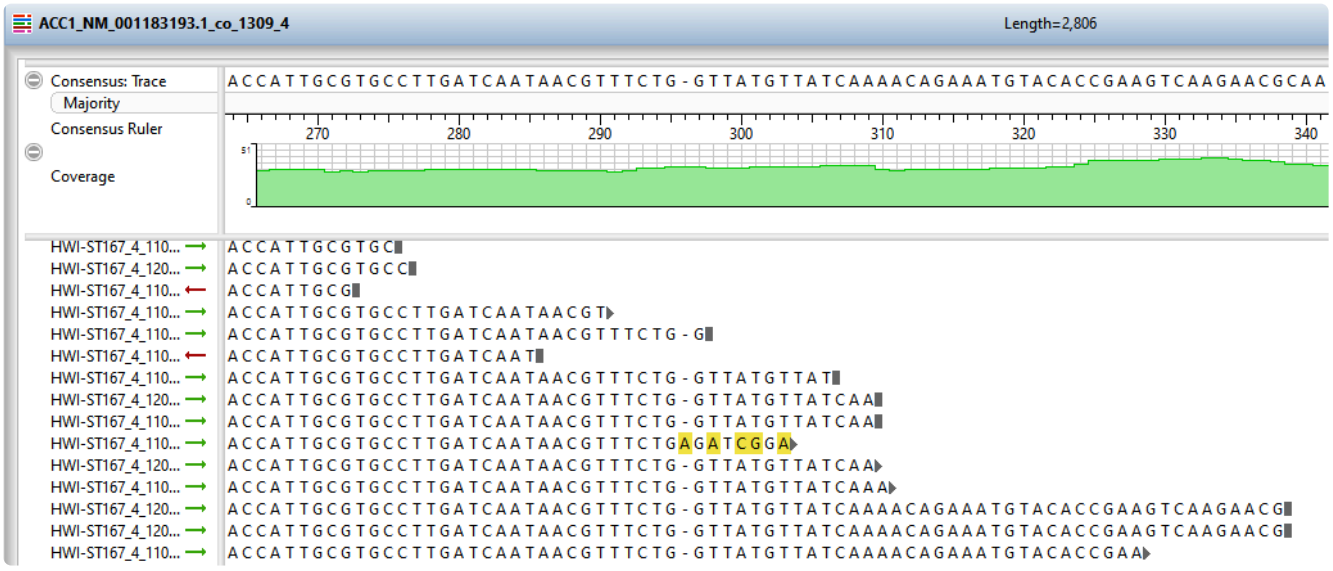
- To sort data in alphanumeric order, click on the header of the column you want to sort. Note that the resulting groups are also shown in different colors to help visually differentiate between them.

Copyright DNASTAR, Inc. 2022



Assembly ID	Gene ...	Organism Name	Accession Number	Description
AAH1_NM_...	AAH1	Saccharomyces cerevisiae	NM_001182979.1	S288C adenine deaminase
AAT2_NM_...	AAT2	Saccharomyces cerevisiae	NM_001181914.1	S288C aspartate transaminase AAT2
ABF2_NM_...	ABF2	Saccharomyces cerevisiae	NM_001182570.1	S288C DNA-binding protein ABF2
ABP140_N...	ABP140	Saccharomyces cerevisiae	NM_001183658.6	S288C Abp140p
ACB1_NM_...	ACB1	Saccharomyces cerevisiae	NM_001181166.3	S288C long-chain fatty acid transporter ACB1
ACC1_NM_...	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288C acetyl-CoA carboxylase ACC1
ACC1_NM_...	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288C acetyl-CoA carboxylase ACC1
ACC1_NM_...	ACC1	Saccharomyces cerevisiae	NM_001183193.1	S288C acetyl-CoA carboxylase ACC1
ACO1_NM_...	ACO1	Saccharomyces cerevisiae	NM_001182192.1	S288C aconitate hydratase ACO1
ACO2_NM_...	ACO2	Saccharomyces cerevisiae	NM_001181633.1	S288C aconitate hydratase ACO2
ACP1_NM_...	ACP1	Saccharomyces cerevisiae	NM_001179758.1	S288C acyl carrier protein
ACS2_NM_...	ACS2	Saccharomyces cerevisiae	NM_001182040.1	S288C acetate--CoA ligase ACS2
ACS2_NM_...	ACS2	Saccharomyces cerevisiae	NM_001182040.1	S288C acetate--CoA ligase ACS2
ACT1_NM_...	ACT1	Saccharomyces cerevisiae	NM_001179927.1	S288C actin
ADE1_NM_...	ADE1	Saccharomyces cerevisiae	NM_001178216.1	S288C phosphoribosylaminoimidazolesuccinocar

- To open an individual contig in SeqMan Ultra 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 Ultra are available.



This marks the end of this tutorial.

Copyright DNASTAR, Inc. 2022

Gap closure workflow

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

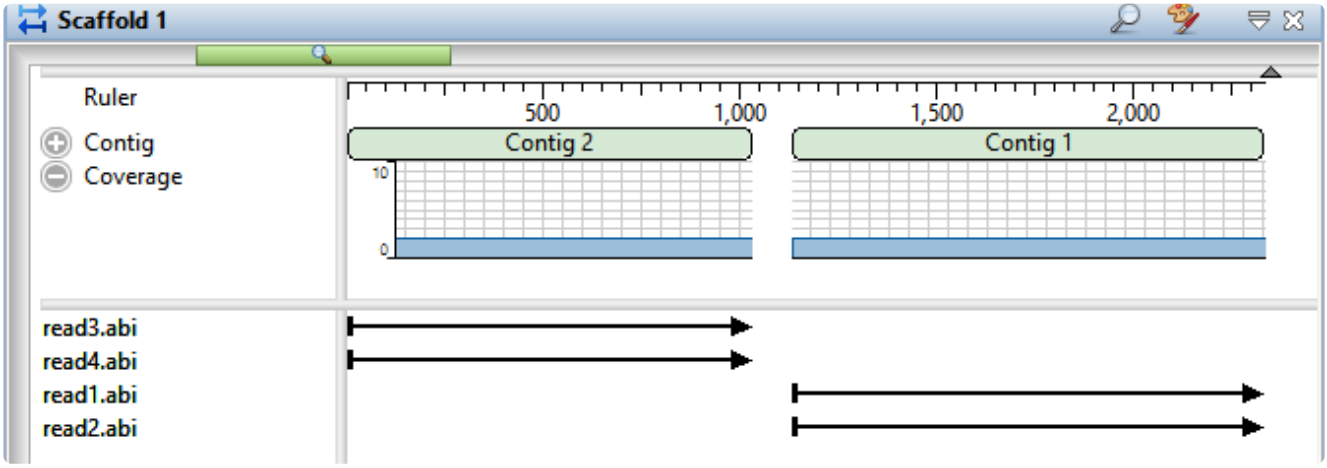
1. Download *T6_Gap_Closure.zip* (45 kb) and extract it to any convenient location (i.e., your desktop). The data set consists of one very small SeqMan Ultra project and one *.abi* file.

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**.

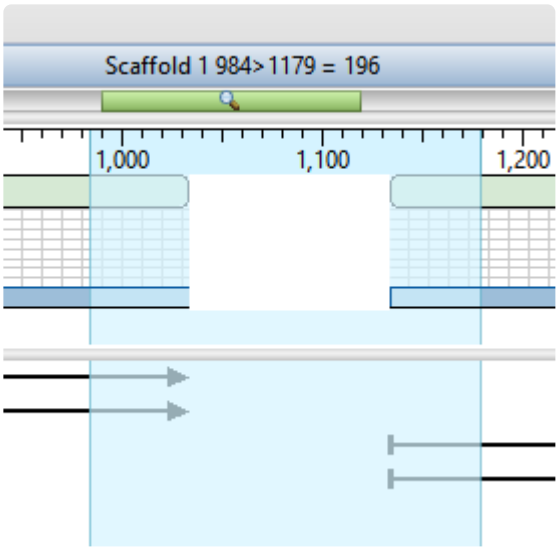
4. Select the scaffold (“Scaffold 1”) and press the **Show Strategy view** tool ( ) to the right of the selection.

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



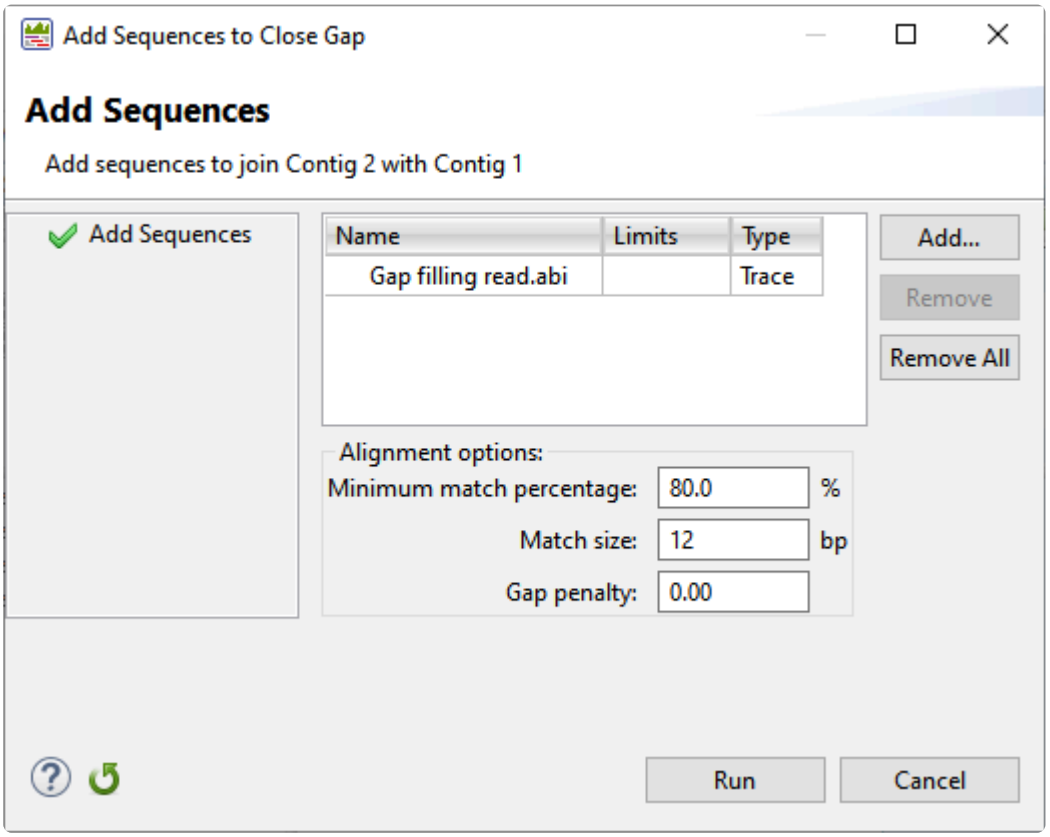
5. Drag the cursor across the ruler area to select a range that spans the gap between the two contigs. There is no need to make a precise selection.

Copyright DNASTAR, Inc. 2022



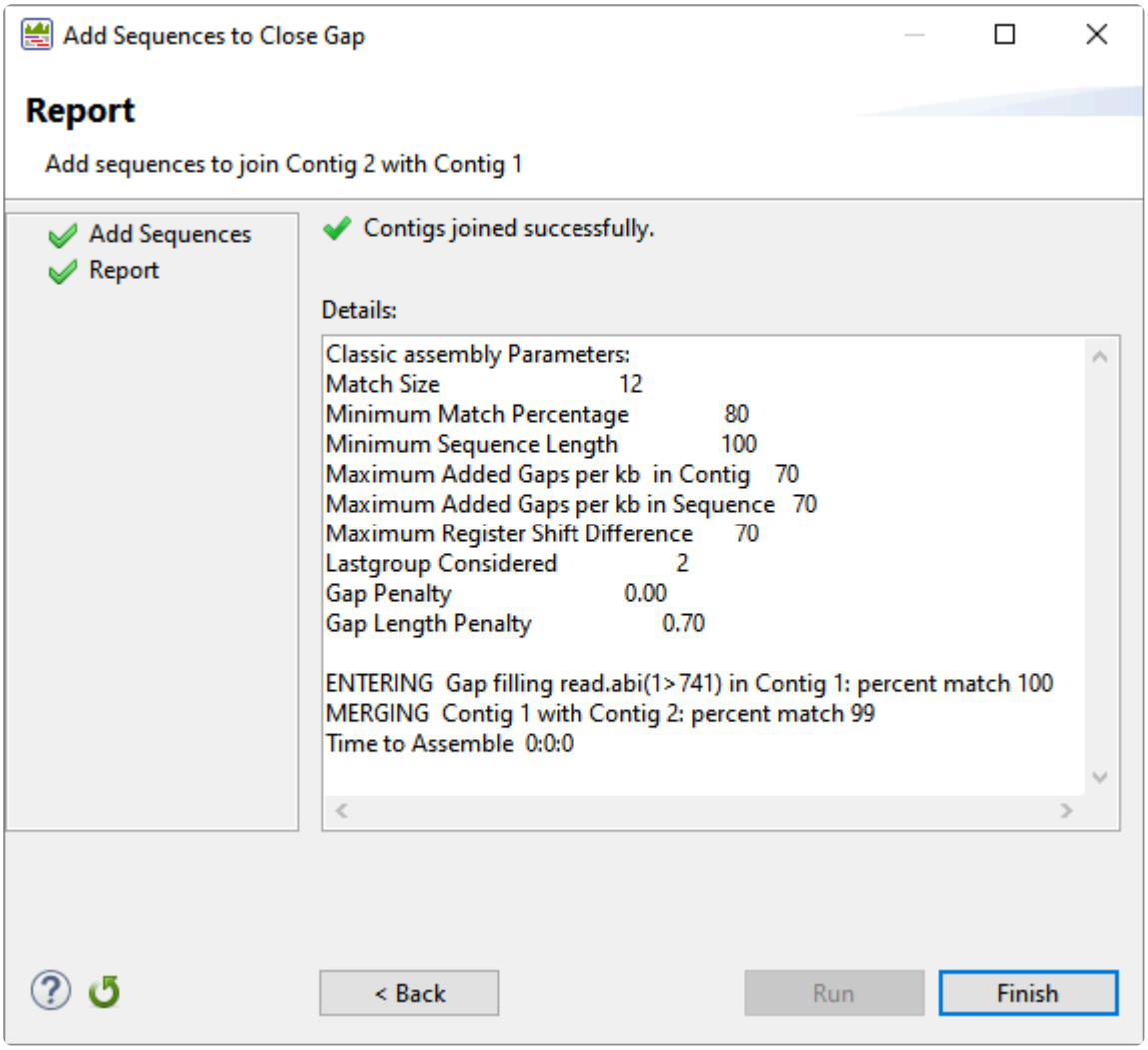
6. Right-click on the selection and choose **Add Sequences to Close Gap**.

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.



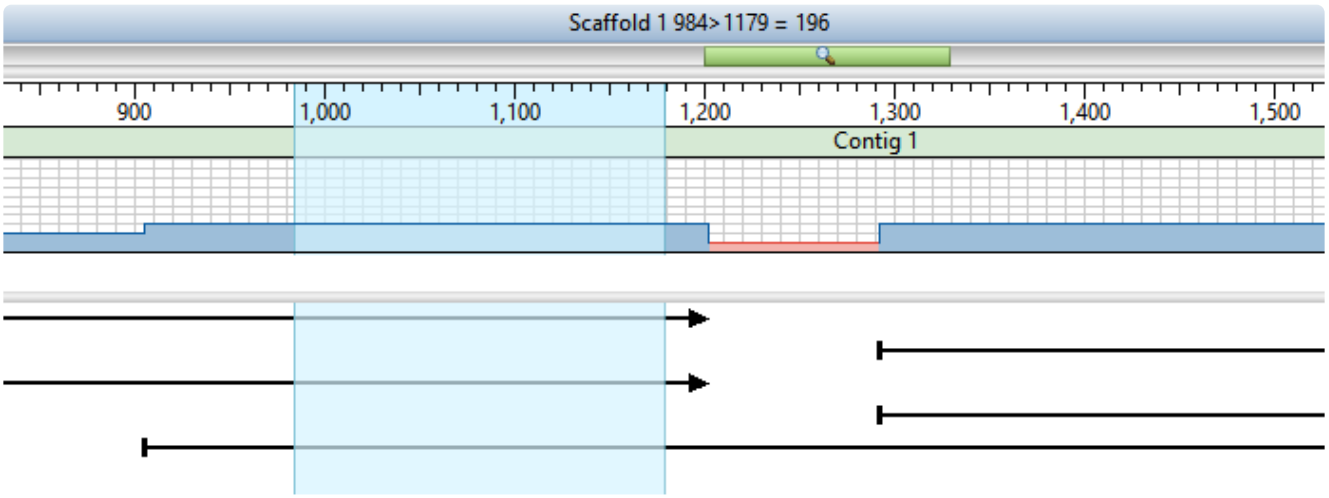
8. Click **Run**. At the bottom of the report, observe the text: **MERGING Contig 1 with Contig 2: percent match 99 (or 100)**.

Copyright DNASTAR, Inc. 2022



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.

Copyright DNASTAR, Inc. 2022

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.

Copyright DNASTAR, Inc. 2022

Getting Started

Did you just launch SeqMan Ultra but have no idea how to start a project? Or you've opened a file but have no idea what to do next? You've come to the right place!

See the following topics to learn how to:

- Use the SeqMan Ultra Welcome screen
- Assemble sequences
- Open an existing assembly
- Use the SeqMan Ultra interface

Copyright DNASTAR, Inc. 2022

SeqMan Ultra Interface

See the following topics to learn about views, projects, panels, and other facets of the SeqMan Ultra interface:

- Welcome screen

- Views
  - Unassembled
  - Project Report
    - Project Overview
    - Project Details
    - Trimming
  - Alignment
  - Strategy
  - Contig Report
    - Contig Information
    - Contig Coverage
    - Contig Structural Variation
  - Features
  - Variants
  - Table
  - Text
  - Pairwise

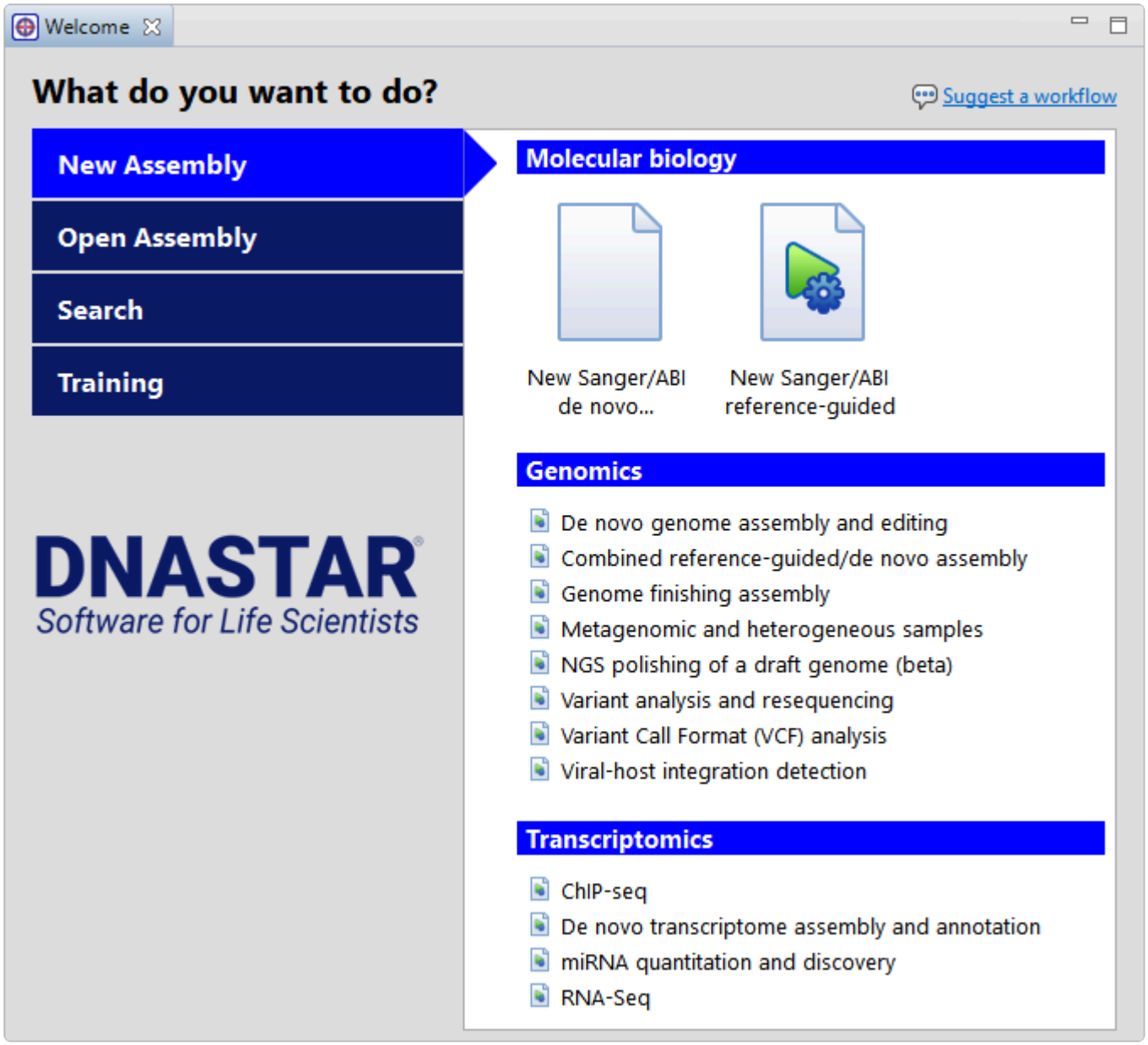
- Panels
  - Explorer
    - Sequences section
  - Style
    - Sequence section
    - Alignment section
    - Strategy section
  - Tracks
    - Layout section
    - Options section
  - Details
  - Jobs

To learn how to optimize the layout and contents of the Protean 3D window, see Customize the look and layout.

Copyright DNASTAR, Inc. 2022

Welcome screen

The Welcome “project” opens when you launch SeqMan Ultra and is a central location for launching and opening assembly projects, searching NCBI databases, and accessing this user guide. As you add projects, the Welcome “project” will remain as the leftmost tab unless you manually move or close it. We recommend leaving the Welcome tab in place, since it has many time-saving shortcuts to useful commands.



Tab on left	Option on right	Description
New Assembly	**Molecular biology**	Choose **New Sanger/ABI de novo** or **New Sanger/ABI reference-guided** to launch the SeqMan NGen application at the applicable tab of its Workflow selection screen. As of Lasergene 17.2, small Sanger assemblies are

Copyright DNASTAR, Inc. 2022

		assembled in SeqMan Ultra, while only large ones are now sent to SeqMan NGen. SeqMan Ultra is used for downstream analysis in both cases.
	Genomics	Choose a genomics workflow by clicking its name. This launches SeqMan NGen application at the applicable tab of its Workflow selection screen. These options are only enabled for users with a current Genomics package license.
	Transcriptomics	Choose a transcriptomics workflow by clicking its name. This launches SeqMan NGen application at the applicable tab of its Workflow selection screen. These options are only enabled for users with a current Genomics package license.
Open Assembly	**Open assembly**	Like the **File > Open** command, this launches a file browser from which you can select an *.sqd* or *.assembly* file. See Create or Open a Project for details.
	Recent documents	Like the **File > Recent Documents** command, this provides a list of recently-opened projects.
Search	**Search**	Identical to using the **Search > Search for Sequences** command. For details, see Search for sequence matches in BLAST.
	Search for text in NCBI nucleotide	Identical to using the **Search > Search for text in NCBI nucleotide** command. For details, see Search for text or accession number matches in Entrez.
	Search for text in NCBI protein	Identical to using the **Search > Search for text in NCBI protein** command. For details, see Search for text or accession number matches in Entrez.
Training	**User Guide**	Opens this User Guide.
	Tutorials	Opens this User Guide to the topic SeqMan Ultra Tutorials.

Copyright DNASTAR, Inc. 2022

Views

The left portion of the SeqMan Ultra window is comprised of one or more “views.”

View types:

To learn more about the types of views and how to change their appearances, click the links below.

- Unassembled view
- Project Report view
    - Project Overview
    - Project Details
    - Trimming
- Alignment view
- Strategy view
- Transcriptome view
- Contig Report view
    - Contig Information
    - Contig Coverage
    - Contig Structural Variation
- Features view
- Variants view
- Table view
- Text view
- Pairwise view

To learn how to select, display and customize views:

See Customize the look and layout.

To access a different instance of a view:

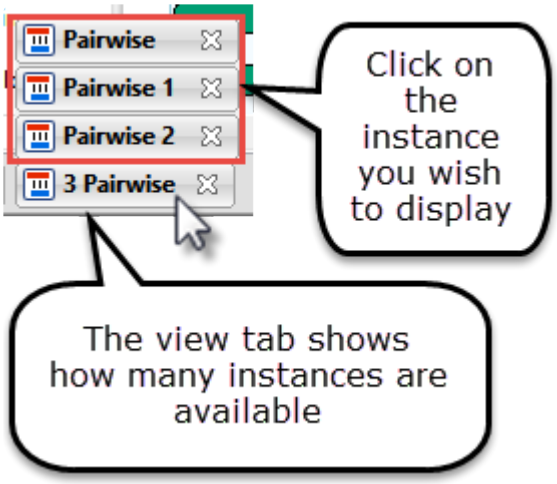
Some DNASTAR applications, including SeqMan Ultra, allow multiple instances of a view to be created. For instance, if you open the results of three BLAST searches, they appear as three instances of the Pairwise view. As shown in the image below, the view tab shows how many views of that type are in that tab folder.

To display a different instance:

1. Hover over the view tab until you see a popup showing each instance by name.

2. Click on the desired instance.

Copyright DNASTAR, Inc. 2022



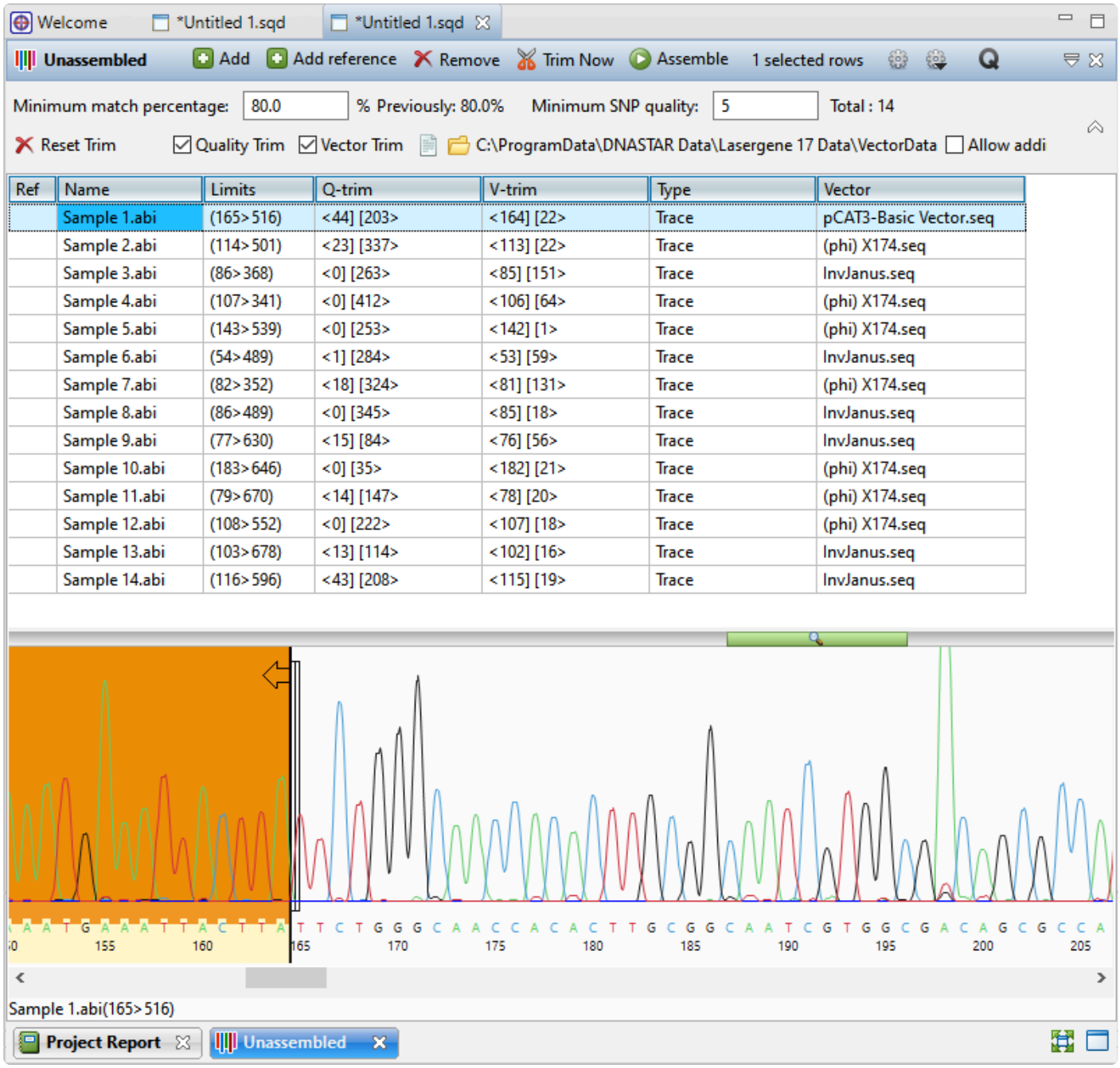
Copyright DNASTAR, Inc. 2022

Unassembled view

The Unassembled view is where you assemble Sanger data directly in SeqMan Ultra; all other data types are assembled using SeqMan NGen. This view lets you load Sanger sample and reference sequences, trim samples based on trace quality or the presence of vector, set assembly parameters and perform the assembly itself.

To launch the Unassembled view from the Welcome window's **New Assembly** tab, click either of the **Molecular biology** options; both options launch Sanger assembly workflows, which are assembled in SeqMan Ultra itself. This causes the Unassembled view to open as the sole view on display. At any other time, you can open the view by clicking on the **Unassembled** view tab or by using the **View > Unassembled > Show** command.

Copyright DNASTAR, Inc. 2022



The view consists of four sections: a blue toolbar, a light gray header, an upper section containing a table, and a lower section that displays trace chromatogram data, if available. The toolbar and header are described in the “how to” table near the bottom of this topic. The upper & lower sections are described separately, below.

Sequence table (upper section):

The top of the Unassembled view consists of a table that is empty until you add sample and/or reference sequences (see the lowermost table in this topic). Once populated, the columns show the following information:

Column Name	Description
Ref	If a sequence was loaded using the **Add reference** tool (see table below), a black check mark will appear to denote that the row contains a reference sequence.
Name	The name of the sequence.
Limits	Displays the coordinates remaining after vector trimming, quality trimming, and/or manual trimming. This is the portion of the sequence that will be used during assembly.
Q-trim	After automatically trimming sequences based on trace data quality, this column shows sequence coordinates that will be omitted from the 5' and 3' ends assembly.
V-trim	After automatically searching for vector sequence, this column shows the coordinates that will be omitted from the assembly due to it being putative vector sequence.
Type	Displays the word **Trace** if the sequence contains trace data. Trace data from Sanger/ABI files are imported and saved in the project along with the sequences themselves.
Vector	After automatically searching for vector sequence, this column shows the name of any vector discovered in a sample.

Chromatogram previewer (lower section):

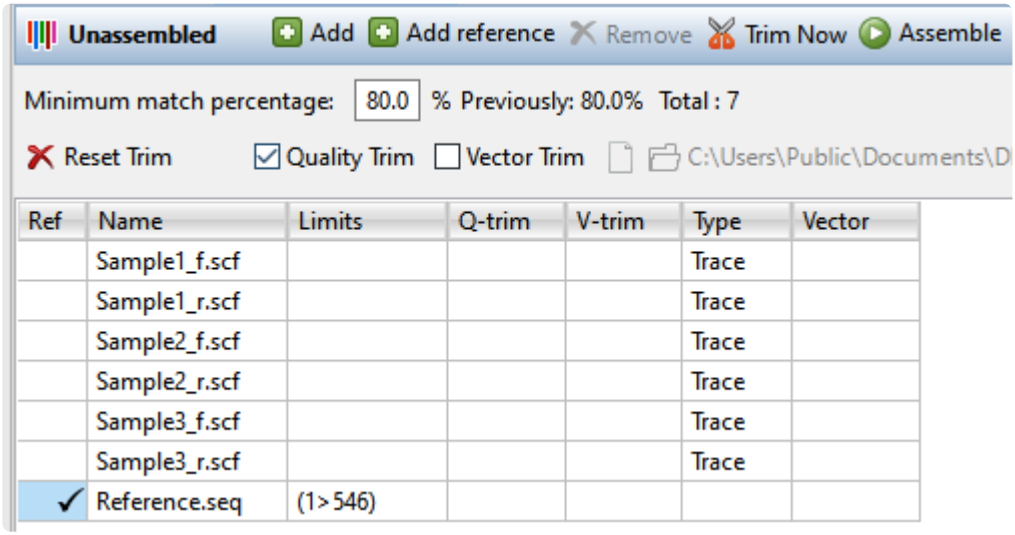
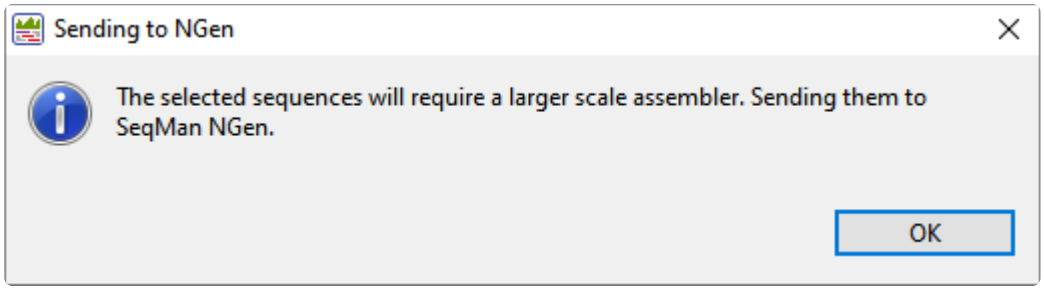
The lower portion of the view consists of a chromatogram previewer with a sequence ruler. The previewer is blank until you double-click a row in the upper section that includes trace data, as shown by the **Type** menu displaying the word **Trace**. Once a trace chromatogram is displayed, any sequence that was removed due to poor quality has a pale yellow background. Trimmed vector sequence is shown with an orange background. You can slide the manual trim bars on the 5' and 3' ends to manually specify the portion of the sequence that should be removed prior to assembly. Manual selection overrides automatic vector and/or quality trimming.

The **Show quality scores** tool ( ) on the right of the header can be used to show or hide different types of quality scores. These scores may assist you in deciding where you would like to manually trim the sequence.

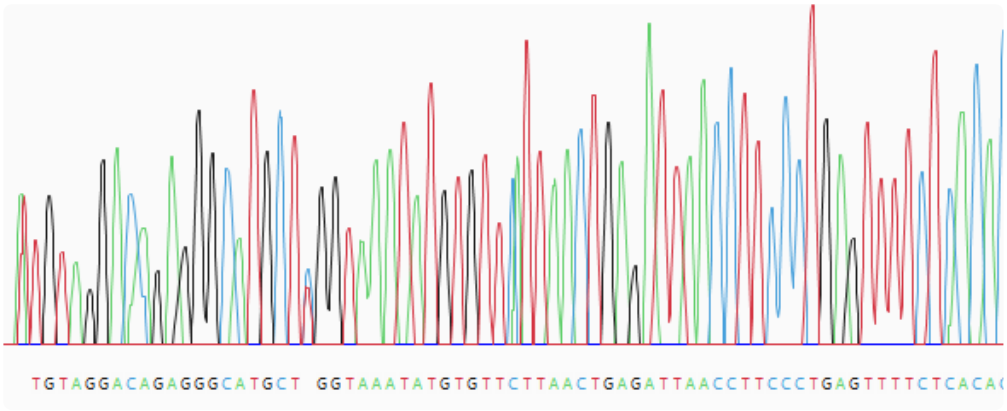
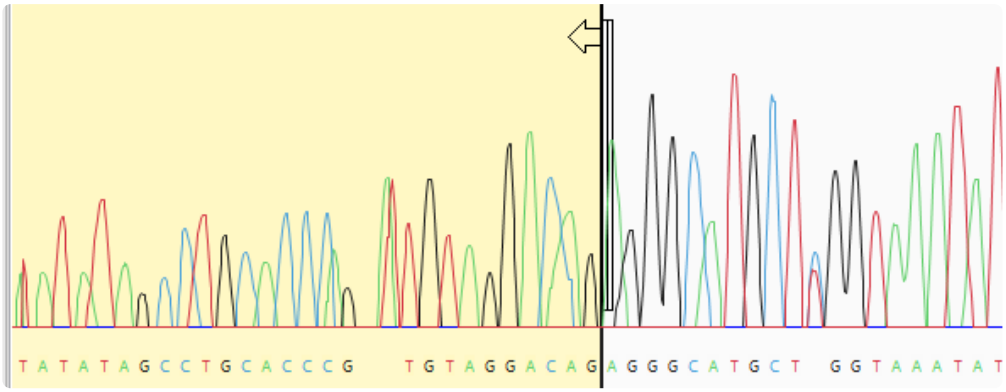
Tasks in the Unassembled view:

Note that the table below shows the tasks without context, and in no particular order. To see how to use these tools and buttons to create a new Sanger assembly, see Assemble Sequences.

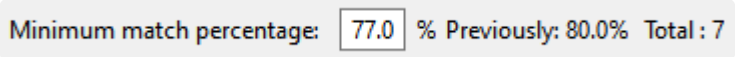
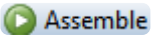
Task	How To
Upload sample and	To upload sample sequences and their trace data (if available), press the **Add** button (Add). To upload reference sequences (optional), press the **Add reference** button (

reference sequences	Add reference). In both cases, select the desired files and press **Open**. Note the files added using the **Add reference** button will appear in the table with a checkmark in the **Ref** column. Sample files that contain trace data will show the word **Trace** in the **Type** column.  Unassembled Add Add reference Remove Trim Now Assemble Minimum match percentage: 80.0 % Previously: 80.0% Total : 7 Reset Trim Quality Trim Vector Trim C:\Users\Public\Documents\D Ref | Name | Limits | Q-trim | V-trim | Type | Vector Sample1_f.scf | Trace Sample1_r.scf | Trace Sample2_f.scf | Trace Sample2_r.scf | Trace Sample3_f.scf | Trace Sample3_r.scf | Trace ✓ Reference.seq | (1>546) **Note:** If you try to add large files or non-Sanger sequences, you will receive a message notifying you that SeqMan NGen must be used for the assembly.  Sending to NGen The selected sequences will require a larger scale assembler. Sending them to SeqMan NGen. OK Press **OK** to launch SeqMan NGen at the **Input Sequences** wizard screen.
Remove sequence rows from the table	Select one or more rows and either choose **Edit > Remove Sequence** or press the **Remove** tool (Remove).
See how many rows are currently selected	This is displayed in the middle of the blue header between the two groups of tools: 5 selected rows.
View trace data for a sample	Double-click on the sample row. This populates the lower portion of the Unassembled view with a trace data chromatogram.

Copyright DNASTAR, Inc. 2022

	
Manually trim sequence ends	Open trace data for a sample by double-clicking its row in the upper table. The chromatogram appears below. Grab the vertical sliders on the 5' and/or 3' ends of the sequence and release them in the desired location. The portion of the sequence that will be removed is shown with a yellow background. The coordinates that will be omitted from the assembly are shown in the bottom left corner of the Unassembled view, as well as in the **Limits** column of the table above. 
Automatically trim sequence ends based on quality	Check the **Quality Trim** box and press the **Trim now** tool (Trim Now). (If the box is not visible, press the tool on the top right). This method uses SeqMan Ultra's proprietary quality score system to remove low-quality data. The coordinates that will be omitted from the assembly are shown in the **Q-trim** column of the table above.
Restore all trimmed sequence	Press the **Reset Trim** tool (Reset Trim). (If the tool is not visible, press the tool on the top right).
Automatically remove vector sequence	Check the **Vector Trim** box. (If the box is not visible, press the tool on the top right). By default, SeqMan Ultra will remove vector sequence from any of the vectors in the *VectorData* catalog installed with Lasergene. If desired, you can override this by using the tools to the right of the checkbox. Press **Choose a vector from a file** to navigate to a specific vector sequence file or **Choose a folder containing vectors** to specify a folder of vectors. Then press the **Trim now** tool (Trim Now). If vector sequence was found and removed, the vector name will be listed in the table's **Vector** column. The coordinates that will be omitted from the assembly are shown in the **V-trim** column.

Copyright DNASTAR, Inc. 2022

Specify the assembly stringency	Enter a number in the box to the right of **Minimum match percentage** in the header. (If the box is not visible, press the tool on the top right). To the right of the box, the header displays the number used the previous time this set was assembled (if an assembly has already taken place) and the number of sample and reference sequences involved in the assembly. If some samples are not being aligned, try changing the **Minimum match percentage** prior to assembling again. 
Assemble the sequences	After performing the desired vector or quality trimming, if any, press the **Assemble** tool () to assemble all the sample sequences in the table. If a reference is present, the assembler will use the reference(s) as a template. Otherwise, the sequences will be *de novo* assembled.
Open a dialog for customizing assembly parameters	Use the **Assembly parameters** tool ( ) or the command **Project > Project Parameters**. See Assembly parameters for information about each of the options.
Display/hide the Minimum match percentage and trimming options in the header	use the **Show quick trimming and assembly options** tool ( ).
Display/hide quality scores in the chromatogram previewer	Press the **Show quality scores** tool ( ) and choose from **Don't show scores**, **Show scores** and **Show averaged scores**. For more information, see Quality score calculations.
Allow or disallow the formation of single-sequence contigs	To allow formation of single-sequence contigs, check the box **Allow adding single sequence contigs**. To prevent single-sequence contigs, leave the box unchecked.

Copyright DNASTAR, Inc. 2022

Project Report view

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

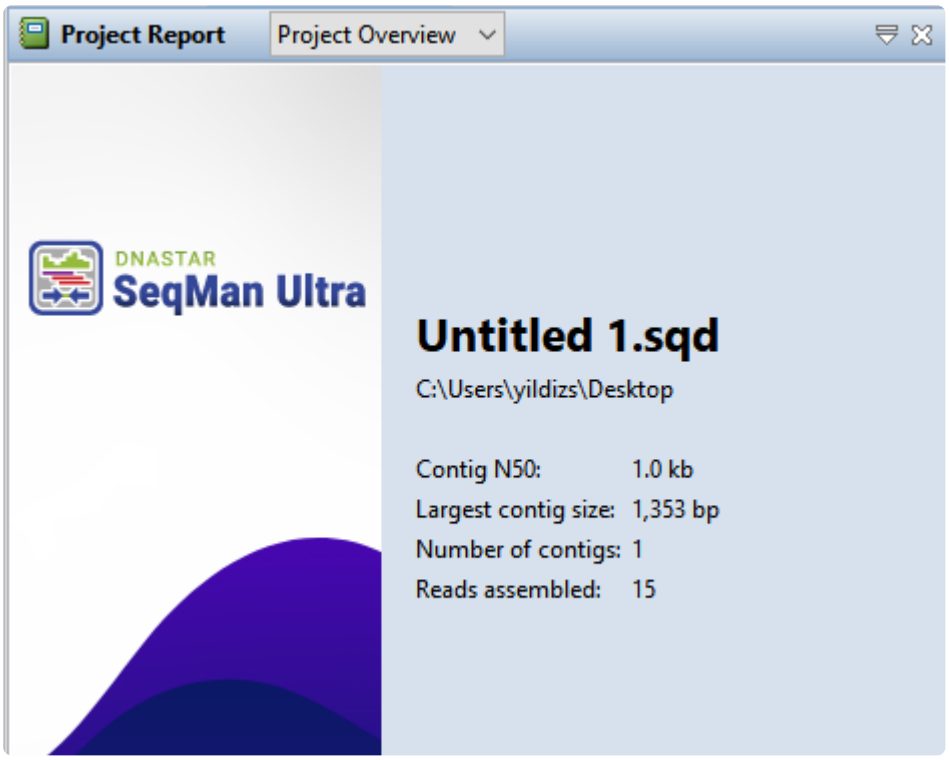
- Project Overview
- Project Details
- Project Statistics
- Trimming

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.

Copyright DNASTAR, Inc. 2022

Project Overview

Project Overview is one of four 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 Project Report view.



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

Parameter	Description
***.sqd* projects**	
Contig N50	A weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value.
Largest contig size	The length of the longest contig in bp.
Number of contigs	The number of contigs in the finished assembly.
Reads assembled	The number of reads that were used in the finished assembly.
***.assembly* projects**	
Average	Average depth of coverage along the assembly.

Copyright DNASTAR, Inc. 2022

coverage	
Number of references	Number of reference sequences used.
Aligned sequences	Number and percentage of sequences used in the finished alignment.
Number of variants	Number of variants found.
Variant assemblies	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.

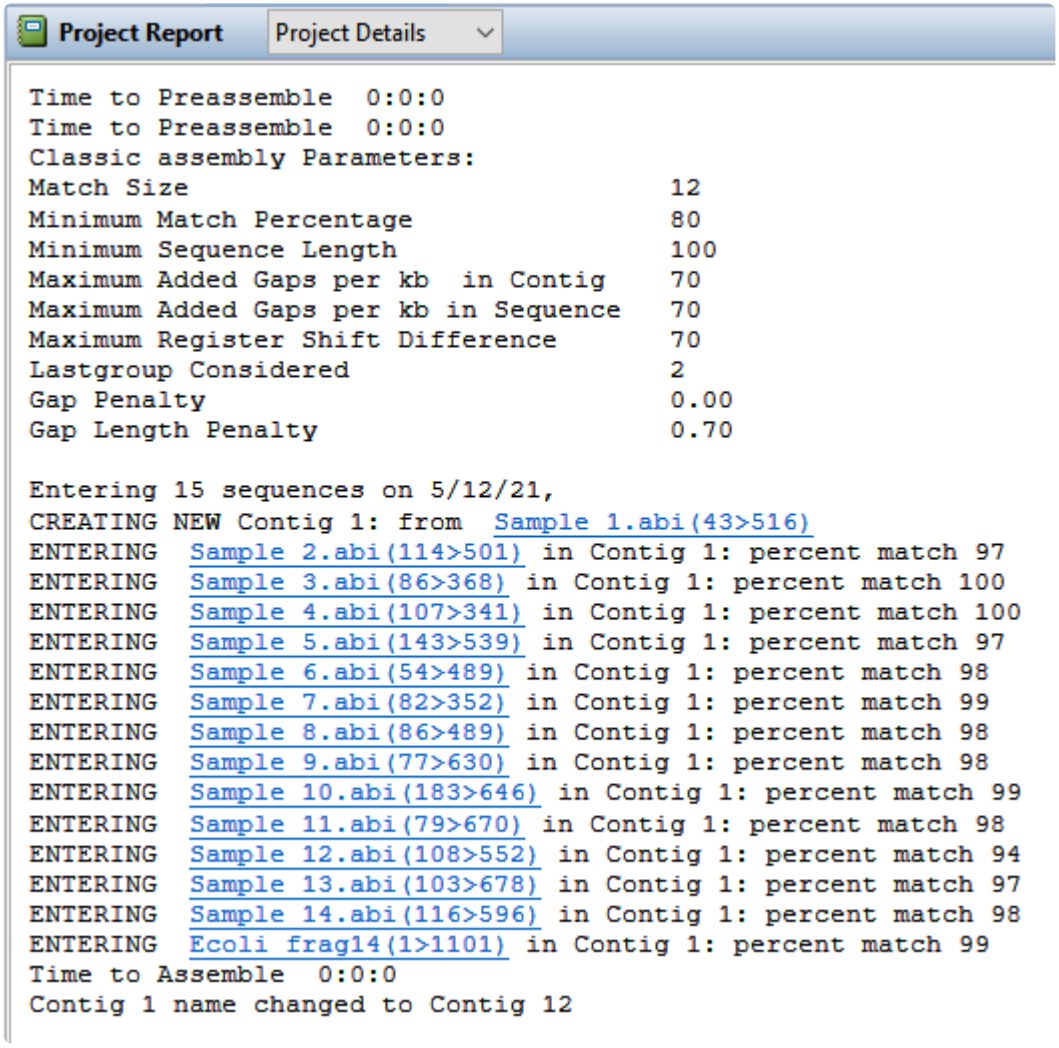
Copyright DNASTAR, Inc. 2022

Project Details

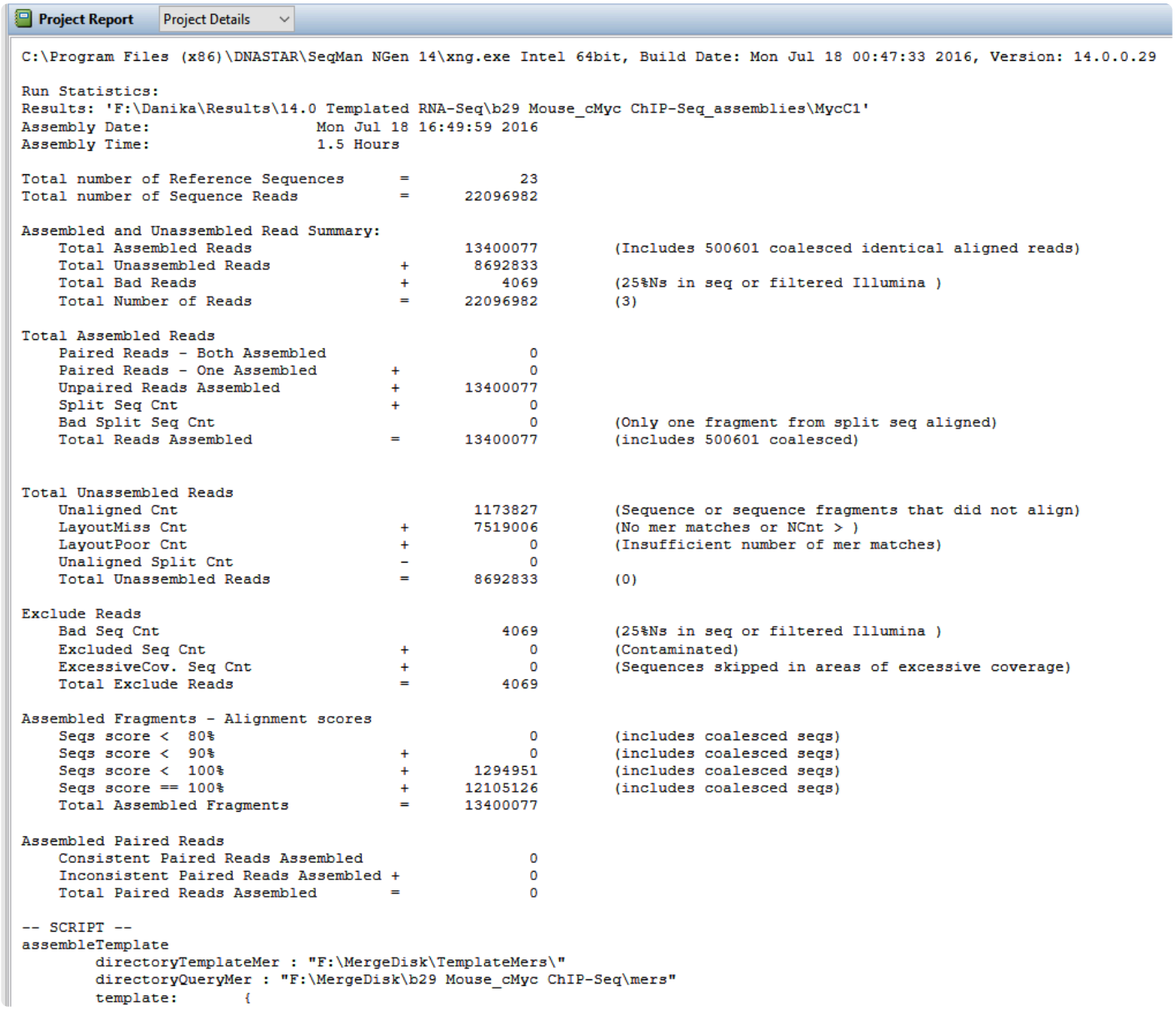
Project Details summarizes assembly time, the parameters used in constructing a contig, and each stage in the progress of an assembly, as well as the median pair distance and standard deviation for each set of a given sized pair library in an assembly. Project Details is one of four Project Report views that may be available, depending on the assembly type that is currently open. For information on accessing or performing tasks in this view, see Project Report view.

The Project Details report varies depending on whether an *.sqd* or *.assembly* is open.

Project Details report for *.sqd* projects:



Project Details report for *.assembly* projects:



Project Report | Project Details

```
C:\Program Files (x86)\DNASTAR\SeqMan NGen 14\xng.exe Intel 64bit, Build Date: Mon Jul 18 00:47:33 2016, Version: 14.0.0.29

Run Statistics:
Results: 'F:\Danika\Results\14.0 Templated RNA-Seq\b29 Mouse_cMyc ChIP-Seq_assemblies\MycC1'
Assembly Date:                  Mon Jul 18 16:49:59 2016
Assembly Time:                  1.5 Hours

Total number of Reference Sequences      =            23
Total number of Sequence Reads           =      22096982

Assembled and Unassembled Read Summary:
    Total Assembled Reads                     13400077        (Includes 500601 coalesced identical aligned reads)
    Total Unassembled Reads              +     8692833
    Total Bad Reads                      +        4069        (25%Ns in seq or filtered Illumina )
    Total Number of Reads                =    22096982        (3)

Total Assembled Reads
    Paired Reads - Both Assembled                   0
    Paired Reads - One Assembled         +          0
    Unpaired Reads Assembled             +   13400077
    Split Seq Cnt                        +          0
    Bad Split Seq Cnt                               0        (Only one fragment from split seq aligned)
    Total Reads Assembled                =   13400077        (includes 500601 coalesced)


Total Unassembled Reads
    Unaligned Cnt                            1173827        (Sequence or sequence fragments that did not align)
    LayoutMiss Cnt                       +   7519006        (No mer matches or NCnt > )
    LayoutPoor Cnt                       +         0        (Insufficient number of mer matches)
    Unaligned Split Cnt                  -         0
    Total Unassembled Reads              =   8692833        (0)

Exclude Reads
    Bad Seq Cnt                              4069        (25%Ns in seq or filtered Illumina )
    Excluded Seq Cnt                     +         0        (Contaminated)
    ExcessiveCov. Seq Cnt                +         0        (Sequences skipped in areas of excessive coverage)
    Total Exclude Reads                  =      4069

Assembled Fragments - Alignment scores
    Seqs score <  80%                               0        (includes coalesced seqs)
    Seqs score <  90%                    +          0        (includes coalesced seqs)
    Seqs score <  100%                   +    1294951        (includes coalesced seqs)
    Seqs score == 100%                   +   12105126        (includes coalesced seqs)
    Total Assembled Fragments            =   13400077

Assembled Paired Reads
    Consistent Paired Reads Assembled               0
    Inconsistent Paired Reads Assembled +           0
    Total Paired Reads Assembled         =          0

-- SCRIPT --
assembleTemplate
        directoryTemplateMer : "F:\MergeDisk\TemplateMers\"
        directoryQueryMer : "F:\MergeDisk\b29 Mouse_cMyc ChIP-Seq\mers"
        template:       {
```

The table below describes some parameters of interest in the *.assembly* version of this report.

Parameter	Description
Total Assembled Reads	Contains information about reads that made it into the assembly. “Coalesced identical aligned reads” indicates the number of reads that were treated as duplicates.
Total Unassembled Reads	Lists reasons why certain reads were not included.
Found SNP Cnt (incl. indel lengths)	(Variant assemblies only) Shows the number of SNPs that were found.
Assembly	(Variant assemblies only) Shows the SNP-calling parameters specified in the SeqMan

Copyright DNASTAR, Inc. 2022

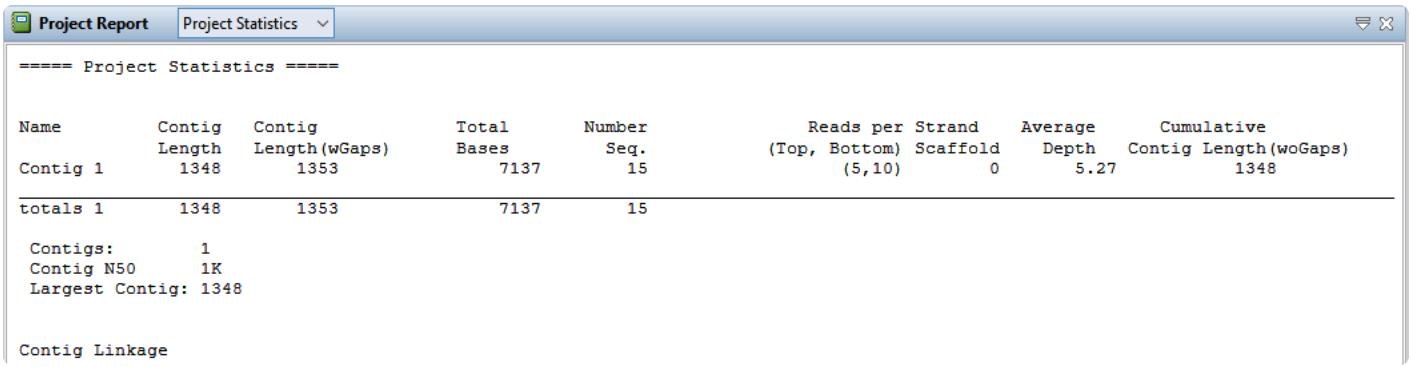
Parameters	NGen wizard for this assembly.
SCRIPT	Below the word “SCRIPT” is the entire assembly script used in SeqMan NGen when creating this assembly.

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.

Copyright DNASTAR, Inc. 2022

Project Statistics

Project Statistics is one of four 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 Project Report view.



Statistic	Description
Located in the table	
Name	Name assigned to the contig during assembly or after renaming it in the Contigs and Scaffolds section.
Contig Length	Length of the contig in bp without counting gaps.
Contig Length (wGaps)	Length of the contig in bp with gaps included.
Total Bases	Total number of bases contained in all sequences used in the assembly.
Number Seq.	Number of sequences assembled into this contig.
Reads per Strand (Top, Bottom) and Scaffold	Number of reads on the top and bottom strands for the scaffold.
Average Depth	Average sequence depth over the length of the contig.
Cumulative Contig Length (woGaps)	Total length of this contig, without gaps, added to the total lengths of the contigs above it in the table.
Located below the table (certain *.sqd* projects only)	
Contigs	Number of contigs in the assembly.
Contig N50	A weighted median statistic such that 50% of the entire assembly is contained in contigs equal to or larger than this value.

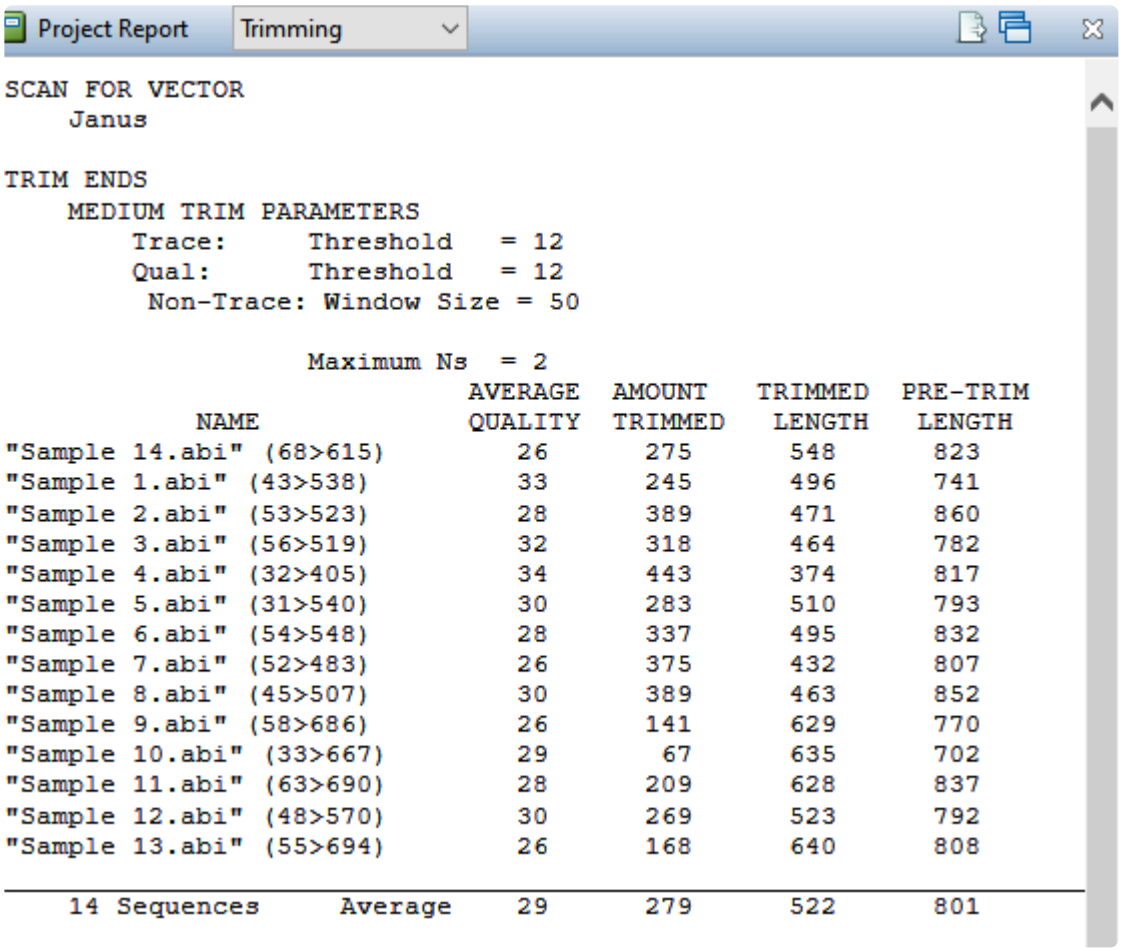
Copyright DNASTAR, Inc. 2022

Largest Contig	Length in bp of the longest contig in the assembly.
Contig Linkage	Shows the paired reads that link contigs together.
PairDistances	For single libraries, the **Median Observed** pair distance is the median distance between paired reads across the data set (only applicable to projects where multiple paired-end libraries are used in a single project).

Copyright DNASTAR, Inc. 2022

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 ot the SeqMan NGen assembly.



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

Data Type	Description
Located above the table	
Trace Threshold	The maximum value of the trace score associated with each base. This is used for end trimming of low quality bases.
Qual Threshold	The maximum value of the quality score associated with each base. This is used for end trimming of low quality bases.
Located in the table	
Non-	SeqMan NGen trims data from sequence ends based on the number of ambiguous residues (Ns)

Copyright DNASTAR, Inc. 2022

Trace Window Size and Max 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.
Name	Name of the read file followed by parentheses showing the position of the sequence on the consensus.
Average Quality	The average quality of bases that remained after trimming this sequence.
Amount Trimmed	Number of bases removed from the sequence, in base pairs.
Trimmed Length	The final length of this sequence after trimming.
Pre-Trimmed Length	The original length of this sequence prior to trimming.

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.

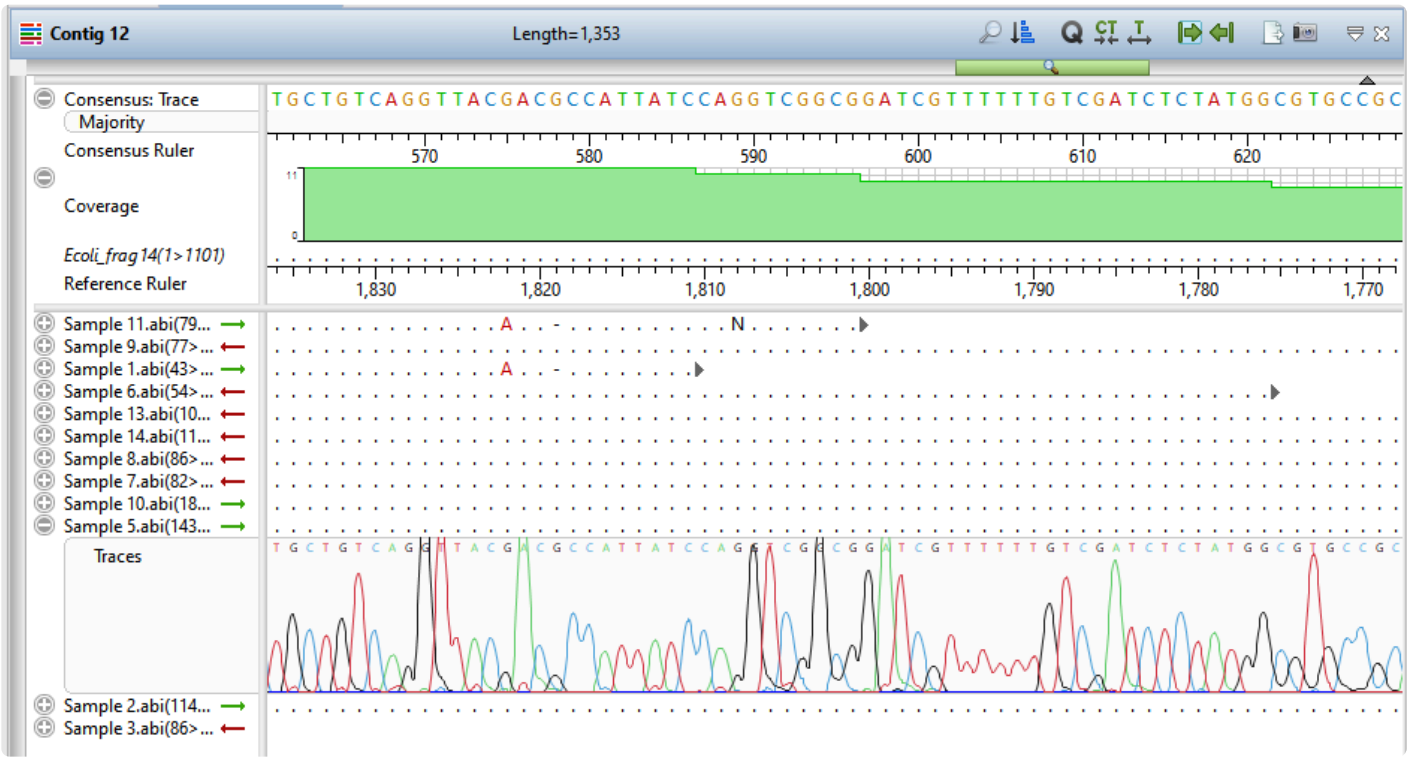
Copyright DNASTAR, Inc. 2022

Alignment View

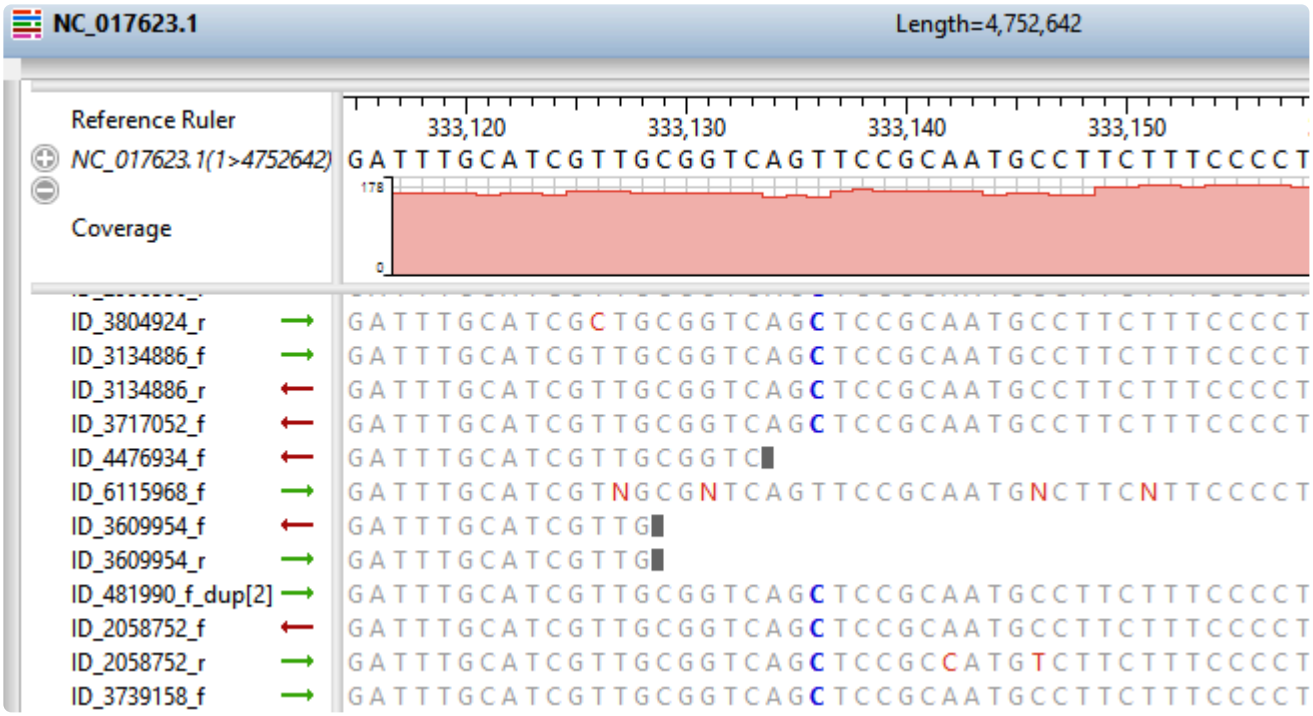
The Alignment view displays the consensus sequence and the alignment of all sequences making up a contig at the nucleotide level. The Alignment view also allows you to edit contigs and contig scaffolds.

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**.

For *.sqd* projects, the default view displays letters only if the base differs from the consensus.



For *.assembly* projects, the default view shows all letters in gray, with variants shown in other colors.



See Default colors in the Alignment view for details. The color schemes and bases shown can be changed using the Alignment section of the Style panel.

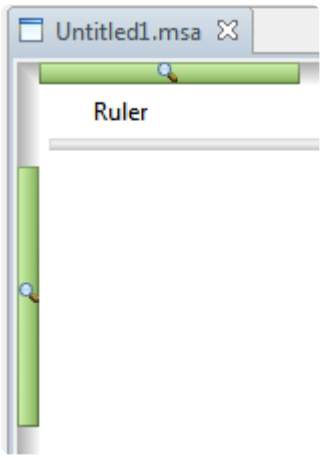
The sequence orientation is shown by the arrow to the left of each sequence string. A right-facing arrow (→) indicates that SeqMan NGen used the input orientation for the alignment. A left-facing arrow (←) shows that SeqMan NGen used the complementary strand.

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 put a sequence in a different order, select its name on the left, then use your mouse to drag and drop it to a new position in the sequence list.

To zoom in or out on the Alignment view, use the green vertical and horizontal zoom in/out sliders. Grab either green slider with your mouse and drag left/right or up/down to zoom in or out in the view.

Copyright DNASTAR, Inc. 2022



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

- Alignment view toolbar
- View multi-sample assemblies in the Alignment view
- Edit a contig
- Search within a contig
- Make a selection
- Customize the look and layout
- Sequence and Alignment sections of the Style panel, along with Available color schemes

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 Variants view, and T's there are not colored as variants.

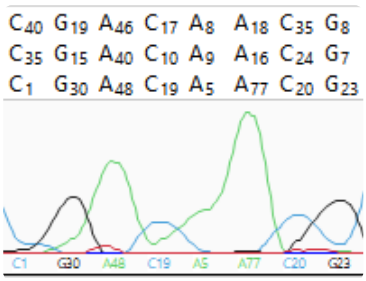
Copyright DNASTAR, Inc. 2022

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 search tools in the upper left of the Alignment view, use the **Find** tool ().
To sort constituent sequences	To sort read files (the files in the left margin of the view) according to the selected criteria, use the **Sort reads** tool (). • **By Name** – Sorts read files in alphanumeric order by name. This type of sorting is normally only used for Sanger reads, which are often user-named. This is a shortcut to the **Sequence > Sort Reads > By Name** command. • **By Position** – Sorts read files by their position along the consensus. This is a shortcut to the **Sequence > Sort Reads > By Position** command. • **By Variant** – Brings all the reads containing variant bases to the top of the alignment stack. This is a shortcut to the **Sequence > Sort Reads > By Variant** command. When you choose this option, SeqMan Ultra automatically loads and calculates the SNPs. Note that if you are working with a large *.sqd* project, this calculation can take a considerable amount of time.
To hide/ show quality scores	Use the **Quality scores** tool (). This tool only functions when a) trace data was used in the assembly. • **Hide scores** – Causes quality scores to be hidden from view. • **Show scores** – Causes quality scores to be displayed as a subscript next to the base. When bases with very high quality scores (>99% of the 95th percentile) are displayed, two asterisks (**) will be shown in place of the usual numerical subscripts. Example: G_{52} A_{84} G** C_{85}.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 *.fas* files, SeqMan NGen automatically checks for *.qual* files with the same name in the same folder. If found, the quality values are displayed. Note that gaps are not assigned quality scores.

Copyright DNASTAR, Inc. 2022

	 • **Show averaged scores** – Displays quality scores averaged over a defined window. This mode is useful for examining trimming. Note that gaps are not assigned quality scores.
To adjust the zoom level	Use the **Compact text** tool ( ) to adjust the zoom level to maximize the number of bases while keeping the text legible. Use the **Restore default zoom** tool ( ) to adjust the zoom level to the (wider) default setting, making the text more comfortable to read.
See how the sequence is anchored	Prior to correcting an *.sqd* alignment with trace data (only), use the **Anchor editing left (insert right)** 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 **Anchor editing right (insert left)** tool ( ) in a similar way.
Slide a Sanger sequence relative to other sequences	Use the **Allow sliding sequences** tool ( ). When this tool is activated, the cursor is shaped like a banana. Use your mouse and the banana-shaped cursor to drag any sequence in the Alignment view to the left or right, as desired. When you are done, click the **Banana** tool again to return the cursor to its normal state. Note that this tool is only active when the alignment contains Sanger trace files.
To force SeqMan Ultra to call bases where ambiguity codes are currently displayed	Use the **Show/hide variant degeneracy** tool ( ). To restore display of ambiguity codes, press the tool again.
Export data	Use the **Export data** tool ( ). This tool acts as a shortcut to the **File > Export Data > Consensus** command and opens the same popup dialog. See Export data to a file for information on using the dialog.
Export image	Use the **Export image** tool ( ). This tool acts as a shortcut to the **File > Export Image > Alignment** command. See Export an image of a view for more information.

Copyright DNASTAR, Inc. 2022

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

Copyright DNASTAR, Inc. 2022

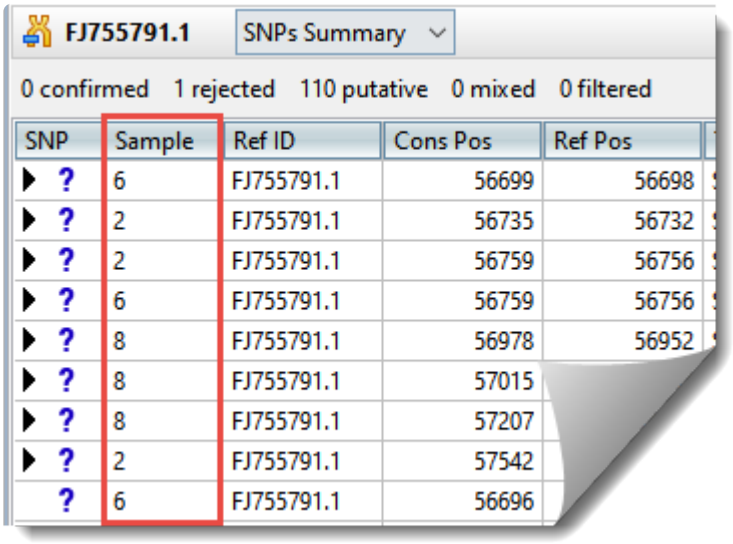
View multi-sample assemblies in the Alignment view

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.

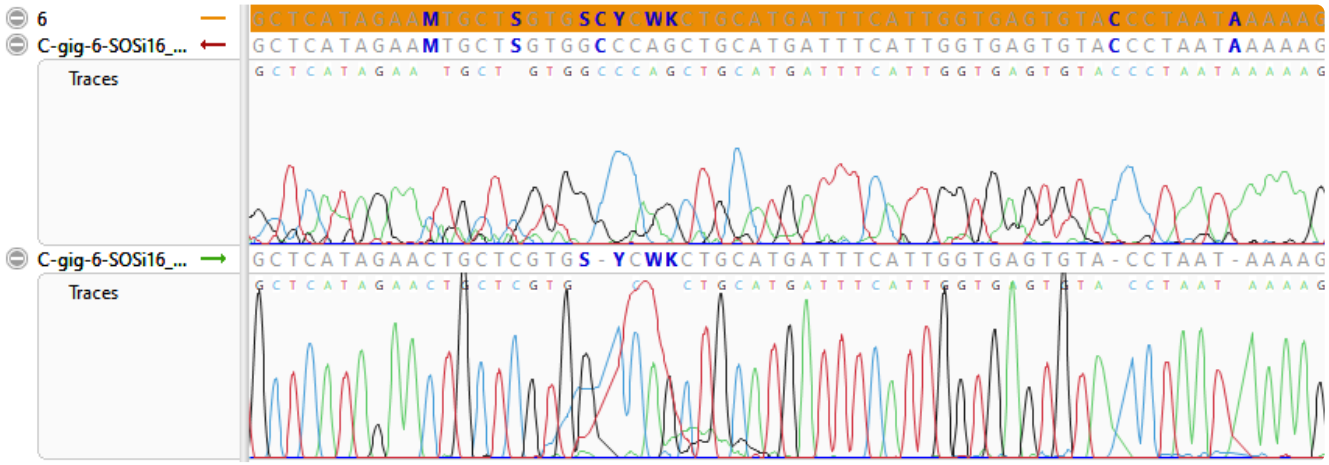


4. Double-click on a variant of interest to launch the Alignment view at that position. The view shows SNPs in blue text on an orange background, allowing you to easily see which samples have variants.

Copyright DNASTAR, Inc. 2022



5. 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.

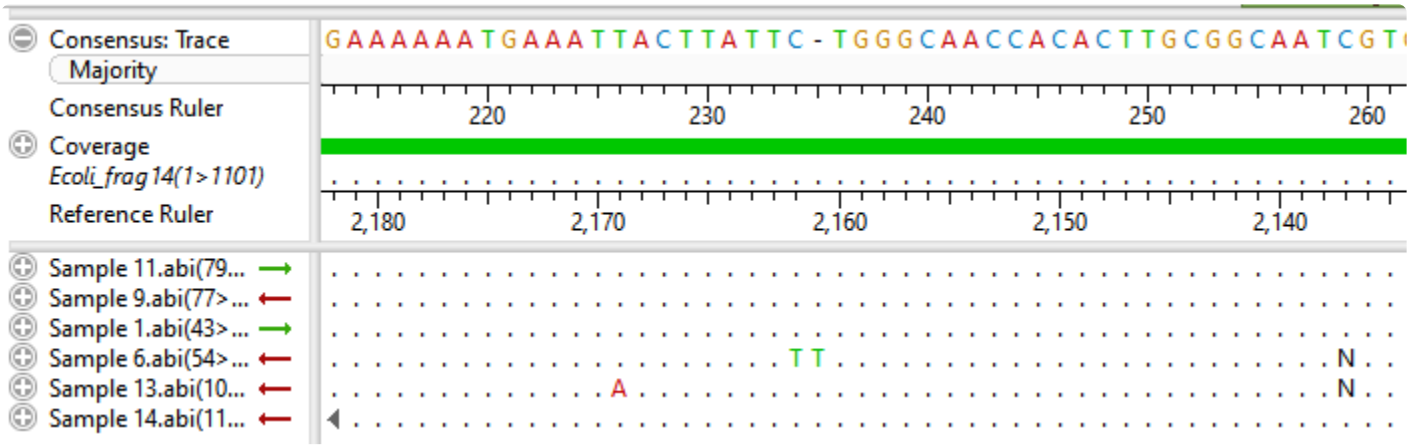


Default colors in the Alignment view

In the Alignment view, colors of SNP and non-SNP bases are determined by the settings in the Alignment section of the Style panel.

Default colors for *.sqd* projects:

The default selection from the **Alignment coloring** drop-down menu is **Show only differences from the consensus**. Ambiguous base differences are shown in black text, while specific base differences are shown, by default, using the **Color by chemistry** color scheme.



Default colors for *.assembly* projects:

The default selection from the **Alignment coloring** drop-down menu is **Color only SNPs**. The table below describes the meaning of different colors in the lower portion of the Alignment view:

SNP Category	Color	Example	Meaning
Not a SNP	Gray	C C T	Did not meet the minimum threshold to be called as a SNP.
Putative SNP	Bold blue	C **G** G	SNP that met or surpassed the minimum variant calling threshold, and is listed in the Variants view with a question mark in the leftmost column. If you apply a more stringent filter in SeqMan Ultra's Variants view, this base will change to "red strikethrough" coloration.
Confirmed SNP	Green	C **G** G	**Validated SNP** – A called SNP that has been manually validated in the leftmost column of the Variants view.
Rejected SNP	Red	T A A	A mismatched base that either 1) was manually rejected by selecting a red 'x' in the leftmost column of the Variants view, or 2) does not meet the minimum variant calling threshold and was filtered by the "hard" filter

Copyright DNASTAR, Inc. 2022

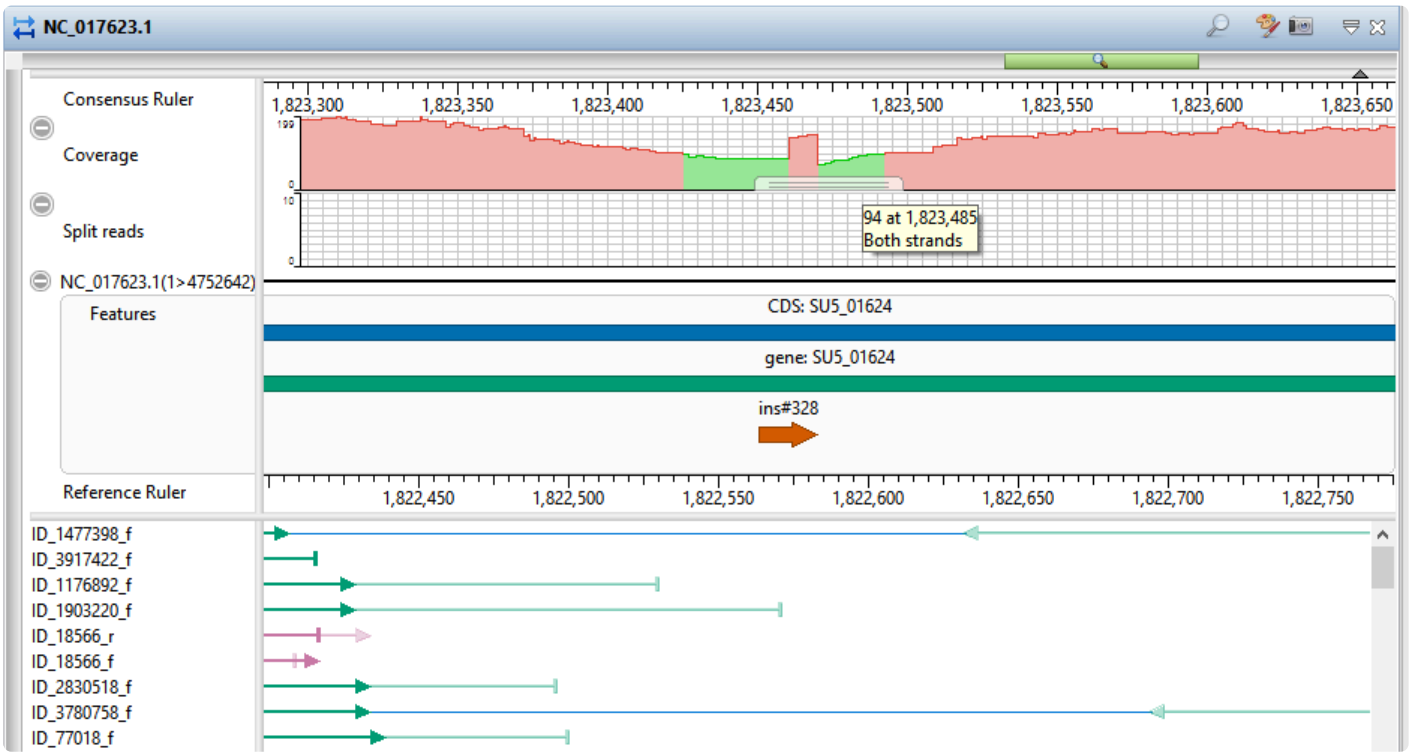
			at the time of assembly using SeqMan NGen. (Note: All assemblies, even those started in SeqMan Ultra, are actually performed “behind the scenes” within the SeqMan NGen application). In the second case, you cannot undo this filter in SeqMan Ultra. If you believe this base to be a SNP, you may wish to rerun the assembly with the variant caller set to a more relaxed stringency setting. In most cases, however, the mismatched base is not a SNP, but a base-calling error; the background “noise” contained by all sequencing data.
Filtered SNP	Red strikethrough		**Filtered SNP** – A base that was called as a SNP during assembly in SeqMan NGen (see note above), but which was later recategorized due to the use of a more stringent filter in SeqMan Ultra’s Variants view. If you relax filter stringency in SeqMan Ultra, this base will return to “bold blue” coloration.

Copyright DNASTAR, Inc. 2022

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**.



Tracks are graphical data displays that run along the length of your sequences in the Strategy and Alignment views. By default, a subset of available data tracks are applied to each of those views. To learn how to apply, show or hide tracks, or to edit track options, see Tracks panel. For information about specific track types, see Tracks.

To zoom in or out on the Strategy view, use the green vertical and horizontal zoom in/out sliders. Grab either green slider with your mouse and drag left/right or up/down to zoom in or out in the view.

The lower half of the Strategy view displays arrows representing the individual sequences in the contig, including color-coded information about paired-end data. 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 coloring 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	Description

Copyright DNASTAR, Inc. 2022

Black	Represents reads for which no paired data information is available or recognized.
Green	Represents paired reads in the same contig whose assembly locations are consistent with the pair specifier parameters.
Dark Blue	Represents paired reads in different contigs in the scaffold whose assembly locations are consistent with pair specifier parameters.
Pale Blue	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.
Orange	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.
Red	Represents paired reads in the same contig whose locations or orientation are inconsistent with pair specifier parameters.
Pink	Represents split reads in the same contig whose locations or orientation are inconsistent with pair specifier parameters.

Paired end sequence data are forward and reverse sequence reads originating at opposite ends of the same 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 two 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 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.

- To see the location of the second member of a pair, hover over the first member for a tooltip with coordinates of its mate pair.

- To move to the location of the second member of a pair, right-click on the first member and select **Show Mate Pair**. 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 arrows use the color coding shown in the table below.

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

- Strategy view toolbar
- Search within a contig

Copyright DNASTAR, Inc. 2022

- Make a selection
- Customize the look and layout
- Change fonts and colors using the Sequence and Strategy sections of the Style panel

Copyright DNASTAR, Inc. 2022

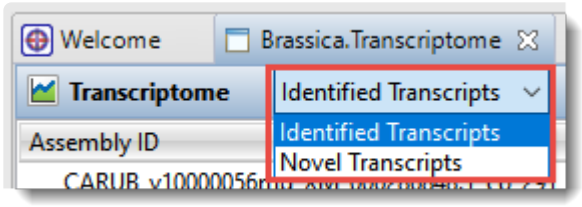
Strategy view toolbar

The tools located in the upper right of the Strategy view are used to accomplish the following tasks.

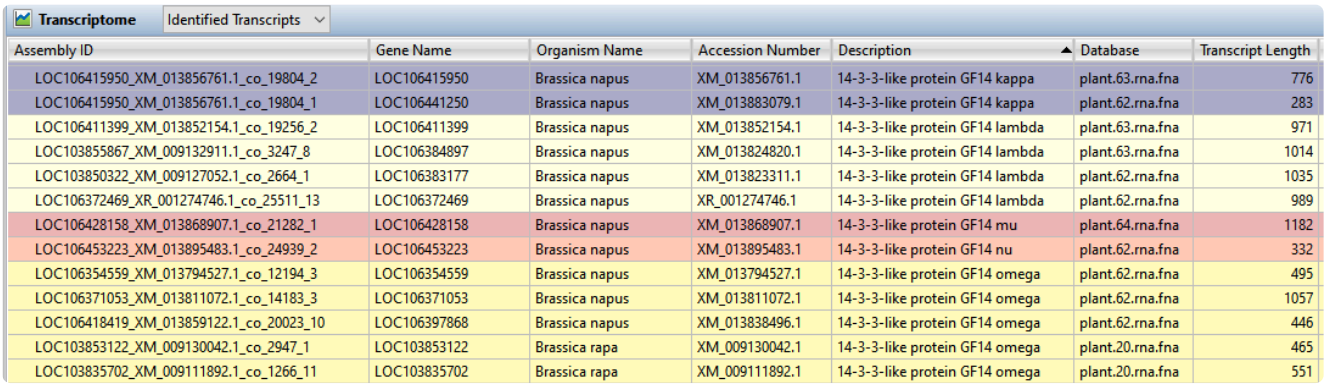
Task	How to
Display the search tools	To display the search tools in the upper left of the Alignment view, use the **Search alignment** tool ().
Customize the appearance of the view	Use the **Show Style panel** tool () to open the Style panel.
Export an image of the view	Use the **Export image** tool ().

Transcriptome view

The Transcriptome view opens automatically when you open a *.Transcriptome* file, the output from a *de novo* RNA-Seq assembly in SeqMan NGen. This view is a two-part report summarizing information about "Identified Transcripts" and "Novel Transcripts." Use the drop-down menu at the upper-left of the view to switch between the two reports.



- **Identified Transcripts** – shows transcripts that had a Transcript Annotation Database match exceeding the specified thresholds. These are labeled with information from the best matching database entry following the default convention: [gene name]_[accession]_co_[assembly ID]_[contig ID]. If you click a column header, the table is not only sorted alphabetically according to the column, but also becomes color-coded for ease of reading.


Transcriptome | Identified Transcripts

Assembly ID	Gene Name	Organism Name	Accession Number	Description	Database	Transcript Length
LOC106415950_XM_013856761.1_co_19804_2	LOC106415950	Brassica napus	XM_013856761.1	14-3-3-like protein GF14 kappa	plant.63.rna.fna	776
LOC106415950_XM_013856761.1_co_19804_1	LOC106441250	Brassica napus	XM_013883079.1	14-3-3-like protein GF14 kappa	plant.62.rna.fna	283
LOC106411399_XM_013852154.1_co_19256_2	LOC106411399	Brassica napus	XM_013852154.1	14-3-3-like protein GF14 lambda	plant.63.rna.fna	971
LOC103855867_XM_009132911.1_co_3247_8	LOC106384897	Brassica napus	XM_013824820.1	14-3-3-like protein GF14 lambda	plant.63.rna.fna	1014
LOC103850322_XM_009127052.1_co_2664_1	LOC106383177	Brassica napus	XM_013823311.1	14-3-3-like protein GF14 lambda	plant.62.rna.fna	1035
LOC106372469_XR_001274746.1_co_25511_13	LOC106372469	Brassica napus	XR_001274746.1	14-3-3-like protein GF14 lambda	plant.62.rna.fna	989
LOC106428158_XM_013868907.1_co_21282_1	LOC106428158	Brassica napus	XM_013868907.1	14-3-3-like protein GF14 mu	plant.64.rna.fna	1182
LOC106453223_XM_013895483.1_co_24939_2	LOC106453223	Brassica napus	XM_013895483.1	14-3-3-like protein GF14 nu	plant.62.rna.fna	332
LOC106354559_XM_013794527.1_co_12194_3	LOC106354559	Brassica napus	XM_013794527.1	14-3-3-like protein GF14 omega	plant.62.rna.fna	495
LOC106371053_XM_013811072.1_co_14183_3	LOC106371053	Brassica napus	XM_013811072.1	14-3-3-like protein GF14 omega	plant.62.rna.fna	1057
LOC106418419_XM_013859122.1_co_20023_10	LOC106397868	Brassica napus	XM_013838496.1	14-3-3-like protein GF14 omega	plant.62.rna.fna	446
LOC103853122_XM_009130042.1_co_2947_1	LOC103853122	Brassica rapa	XM_009130042.1	14-3-3-like protein GF14 omega	plant.20.rna.fna	465
LOC103835702_XM_009111892.1_co_1266_11	LOC103835702	Brassica rapa	XM_009111892.1	14-3-3-like protein GF14 omega	plant.20.rna.fna	551

- **Novel Transcripts** – shows assembled transcripts that either did not have a database match or had a preliminary match that then fell below thresholds upon further processing. The former is labeled following the convention cl_[assembly ID]_[contig ID] and the latter as [gene name]_[accession]_co_[assembly ID]_[contig ID] to give a hint as to the possible identity of that sequence.

Table columns:

Column name	Description
% Gene Match	Length of the matching segment in the database entry x 100, divided by the total length of the database entry.

Copyright DNASTAR, Inc. 2022

% Identity	Total number of identical bases in the matching region x 100, divided by the total number of bases in the matching region.
% Transcript Match	Length of the matching segment in the transcript x 100, divided by the total length of the transcript.
% of Full Length	Length of the assembled sequence x 100, divided by the length of the corresponding database entry. Values greater than 100% indicate that the assembled sequence is longer than the database entry.
Accession Number *	Accession number of the best match.
Assembled Reads	Total number of assembled reads for that sequence.
Assembly ID	Name assigned to the assembled sequence, using the criteria specified in the wizard.
Bit Score	Normalized value calculated from the raw score and expressed in units of “bits,” a common measure in information theory.
Database	Database (e.g. RefSeq, Custom, etc.) from which the best matching gene came.
Description *	Description of the best match.
Gene End	Position in the database entry where the match ends.
Gene Length	Length of the database entry, in bases.
Gene Name *	Best matching gene meeting criteria defined in the wizard.
Gene Start	Position in the database entry where the match begins.
Organism Name *	Organism from which the best matching gene came.
Transcript End	Position in the assembled sequence where the match ends.
Transcript Length	Length of the assembled sequence, in bases.
Transcript Start	Position in the assembled sequence where the match begins.
eValue	“Expectation value,” an estimate of the probability of obtaining the observed alignment score with two random sequences. Expectation values are less sensitive to length than Bit scores and are

Copyright DNASTAR, Inc. 2022

	therefore a better general measure of alignment quality.

* Four columns use default names (e.g., **Gene Name**, **Organism Name**) if one of the default RefSeq databases was used in the SeqMan NGen assembly. However, if you used a custom GREP expression or a custom database that did not include these fields, these columns may have different names or be absent from the table.

About *.Transcriptome* files:

During the assembly stage of SeqMan NGen's *de novo* transcriptome/RNA-Seq workflow, transcript consensus sequences are annotated and named via a search of a transcript annotation database e.g., NCBI's RefSeq database. De novo transcriptome assembly output is saved to a package called [project name].Transcriptome. The package contains three sub-folders, as well as a text file with a high level summary of the results. The three subfolders are:

- Assemblies – Composed of a series of subfolders named sub_0, sub_1, sub_2, … each of which contain editable SQD documents of the Identified Transcripts assemblies. These SQD documents contain one or more contigs that match the same database entry. The Assemblies folder also contains a separate SQD entitled [project name]_NovelTranscripts.sqd, composed of the assembled contigs that did not have matches to the database, as well as a [project name]_AllUnassembled.fastq file containing the unassembled reads from the project in .fastq format.

- Reports – Contains a [project name].AllTranscripts.SearchResults, text file with summary information on both the Identified and Novel transcripts.

- Transcripts – Contains multi-sequence fasta files for the Identified and Novel Transcript consensus sequences.

Assembled transcripts with a database match exceeding the specified thresholds are referred to as "Identified Transcripts," and are labeled with information from the best matching database entry following the default convention: [gene name]_[accession]_co_[assembly ID]_[contig ID]. In cases where a gene name is not provided for a database entry, the name will be truncated to [accession]_co_[assembly ID]_[contig ID].

Assembled transcripts that either did not have a database match or had a preliminary match that then fell below thresholds upon further processing are referred to as "Novel Transcripts". The former is labeled following the convention cl_[assembly ID]_[contig ID] and the latter as [gene name]_[accession]_co_[assembly ID]_[contig ID] to give a hint as to the possible identity of that sequence.

Copyright DNASTAR, Inc. 2022

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 then do any of the following:

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

Note that the **Clone this view** tool () appears in the top right of the view header, but is not currently enabled for this view.

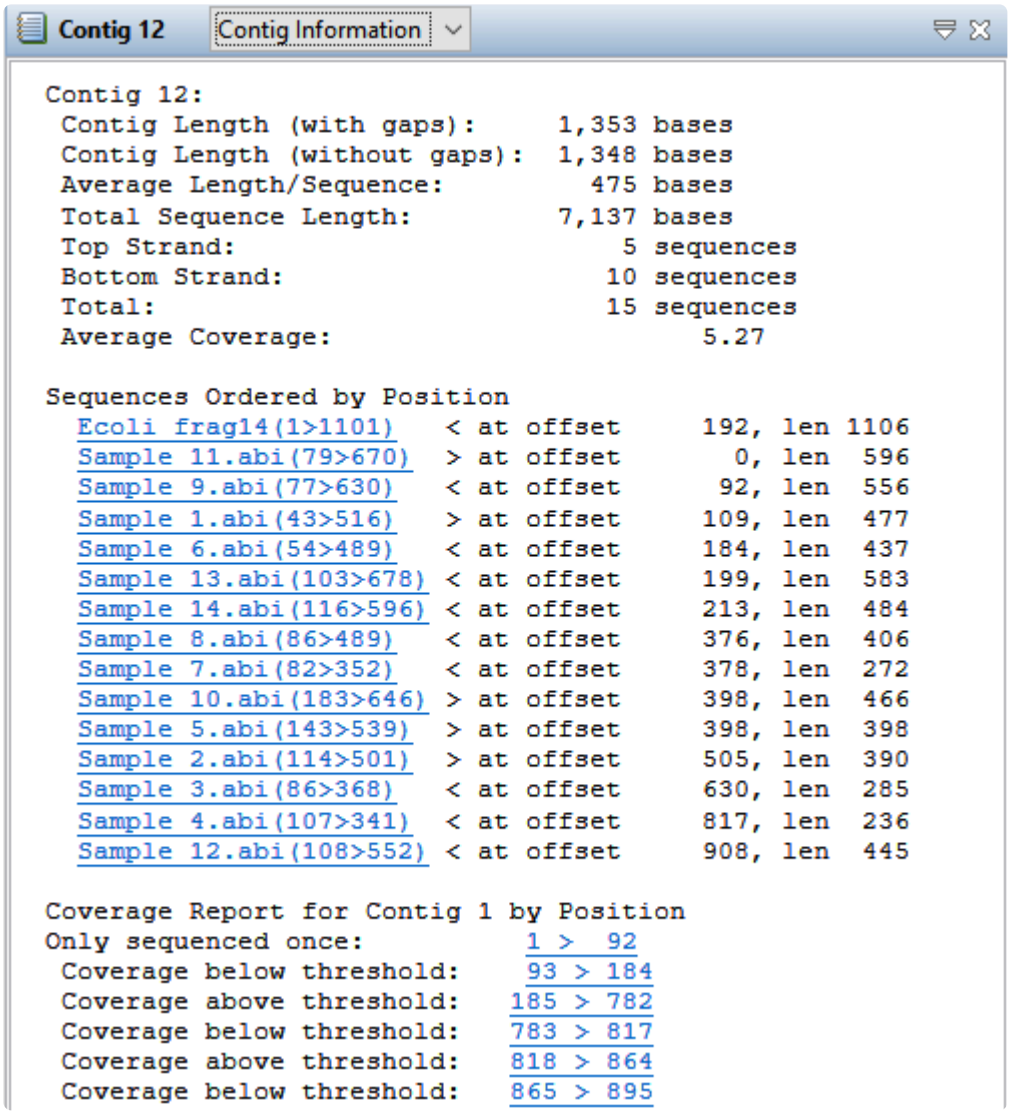
To learn how to customize tables in this view, see Customize tables in the views.

Copyright DNASTAR, Inc. 2022

Contig Information

Contig Information is one of three Contig Report views that may be available, depending on the assembly type that is currently open. 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:

For information on accessing or performing tasks in this view, see Contig Report View.



Parameter	Description
Contig Length (with gaps)	The length of the consensus sequence. This length reflects gaps inserted into the individual sequences and any editing you have done to the contig.
Contig	The length of the consensus sequence without gaps.

Copyright DNASTAR, Inc. 2022

Length (without gaps)	
Average Length/ Sequence	The quotient of **Total Sequence Length** and the number of sequences in a contig.
Total Sequence Length	The sum of all individual sequence lengths. This length reflects gaps inserted into individual sequences.
Top Strand	The number of sequences assigned to the top strand of the consensus sequence.
Bottom Strand	The number of sequences assigned to the bottom strand of the consensus sequence.
Total	The total number of sequences used to construct a contig.
Average (or Median) Coverage	The average (or median) number of reads over the entire length of the consensus.
Sequences ordered by position	Summarizes the offset position and length for each sequence in a contig.
Coverage report for Contig 'n' by Position	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 Strategy Viewing and Coverage parameters.
Coverage Summary (in bases)	Summarizes the number and frequency of bases in each coverage category.

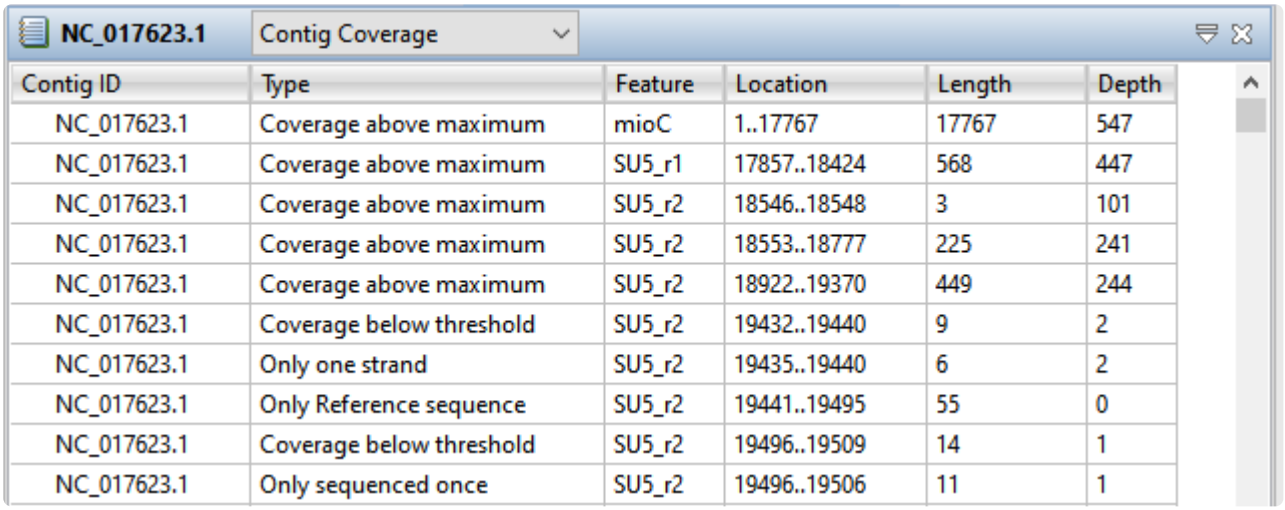
Double click on hyperlinked text to open the Alignment view for the contig in which that sequence appears, with the position or range already selected.

Copyright DNASTAR, Inc. 2022

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.



NC_017623.1 | Contig Coverage

Contig ID	Type	Feature	Location	Length	Depth
NC_017623.1	Coverage above maximum	mioC	1..17767	17767	547
NC_017623.1	Coverage above maximum	SU5_r1	17857..18424	568	447
NC_017623.1	Coverage above maximum	SU5_r2	18546..18548	3	101
NC_017623.1	Coverage above maximum	SU5_r2	18553..18777	225	241
NC_017623.1	Coverage above maximum	SU5_r2	18922..19370	449	244
NC_017623.1	Coverage below threshold	SU5_r2	19432..19440	9	2
NC_017623.1	Only one strand	SU5_r2	19435..19440	6	2
NC_017623.1	Only Reference sequence	SU5_r2	19441..19495	55	0
NC_017623.1	Coverage below threshold	SU5_r2	19496..19509	14	1
NC_017623.1	Only sequenced once	SU5_r2	19496..19506	11	1

This tabular view contains the following columns:

Column Name	Description
Contig ID	The name of the contig.
Type	The type of coverage for the particular region.
Feature	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 **Feature** column shows which feature was used to constrain the analysis area.
Location	The range of sequence corresponding to the region of coverage.
Length	The length of the range of sequence corresponding to the region of coverage.
Depth	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 Alignment view. 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.

Copyright DNASTAR, Inc. 2022

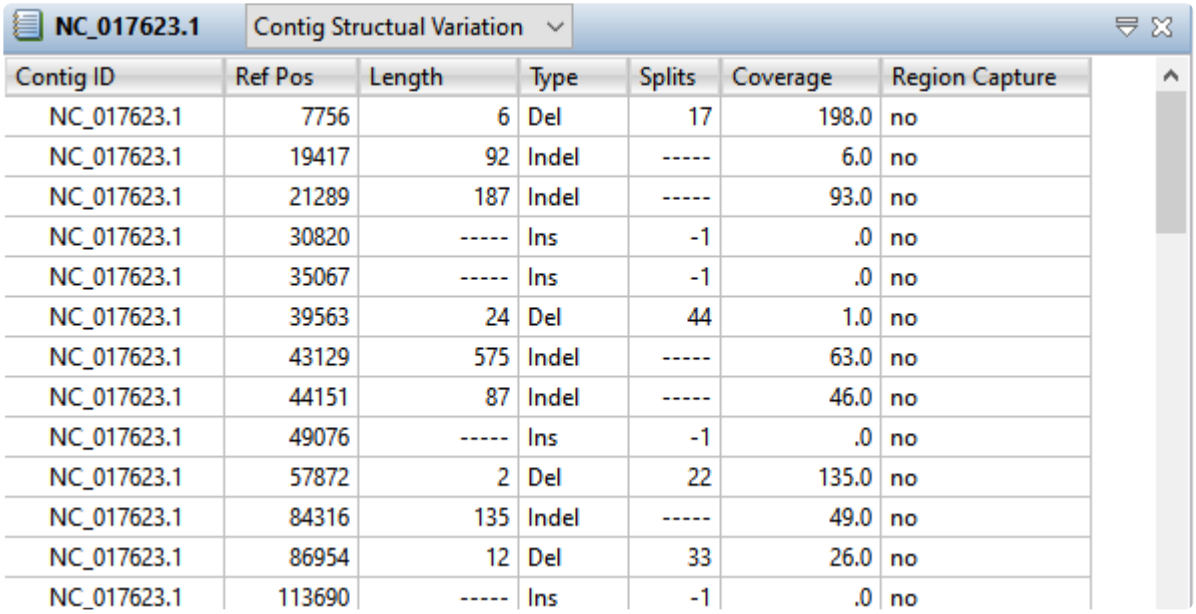
SNP Count	The number of variants in the region. This column is only present if the sequences were assembled to a reference/template.

Copyright DNASTAR, Inc. 2022

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. Note that this report is only available for *.assembly* projects.



Contig ID	Ref Pos	Length	Type	Splits	Coverage	Region Capture
NC_017623.1	7756	6	Del	17	198.0	no
NC_017623.1	19417	92	Indel	-----	6.0	no
NC_017623.1	21289	187	Indel	-----	93.0	no
NC_017623.1	30820	-----	Ins	-1	.0	no
NC_017623.1	35067	-----	Ins	-1	.0	no
NC_017623.1	39563	24	Del	44	1.0	no
NC_017623.1	43129	575	Indel	-----	63.0	no
NC_017623.1	44151	87	Indel	-----	46.0	no
NC_017623.1	49076	-----	Ins	-1	.0	no
NC_017623.1	57872	2	Del	22	135.0	no
NC_017623.1	84316	135	Indel	-----	49.0	no
NC_017623.1	86954	12	Del	33	26.0	no
NC_017623.1	113690	-----	Ins	-1	.0	no

This tabular view contains a subset of the following columns:

Column Name	Description
Contig ID	The name of the contig in which the SV was found.
Contig Pos	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.
Ref Pos	the position where the SV was detected in ungapped reference coordinates.
Length	Total length of a deletion in bases.
Type	The type of SV: Ins (insertion), Del (deletion), or Indel (substitution).
Splits	The number of "split reads" defining the breakpoints of a deletion.
Library	A number corresponding to the library with information in that row. This column is present only for

Copyright DNASTAR, Inc. 2022

	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 **Contig > Contig Info** command.
Pair Dist	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.
Pair Count	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.
Coverage	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 Alignment View column just prior to the left edge of the deletion. However, **Coverage** does not measure the depth of coverage of that column, as the column is excluded from the range over which the mean is taken. **Coverage** FAQ: Q: In the SV Report for a gap-closure workflow, why does the **Coverage** column display low values (e.g., 0-7) for the majority of reference positions? 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. Q: Why doesn't the Alignment view for an SV with a **Coverage** of **4** have four sequences of coverage? 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.
Feature	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. *polB…ilvH*). To see a complete list of affected genes for a given SV, select the SV and use the **Features > Show Feature Table** 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.
Region Capture	Whether or not the identified region overlaps a segment in the region capture file (i.e., the BED file).

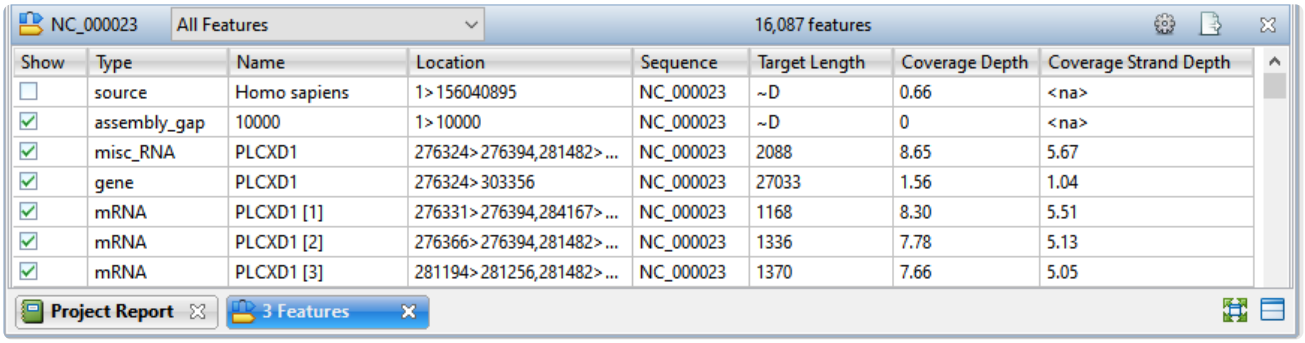
Note: For multi-sample data, structural variation is reported per MID-tag, per sample.

Copyright DNASTAR, Inc. 2022

Features view

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 **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
☐	source	Homo sapiens	1>156040895	NC_000023	~D	0.66	<na>
☑	assembly_gap	10000	1>10000	NC_000023	~D	0	<na>
☑	misc_RNA	PLCXD1	276324>276394,281482>...	NC_000023	2088	8.65	5.67
☑	gene	PLCXD1	276324>303356	NC_000023	27033	1.56	1.04
☑	mRNA	PLCXD1 [1]	276331>276394,284167>...	NC_000023	1168	8.30	5.51
☑	mRNA	PLCXD1 [2]	276366>276394,281482>...	NC_000023	1336	7.78	5.13
☑	mRNA	PLCXD1 [3]	281194>281256,281482>...	NC_000023	1370	7.66	5.05

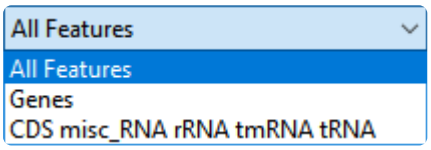
If you select multiple contigs, a separate Features view will open for each of them. If you select “Unlocated Contigs,” individual tables will open for each member of the group. You may wish to avoid selecting “Unlocated Contigs” if you are working with a large number of such contigs, as with many transcriptome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

The following table describes each of the columns in this view:

Column	Description
Show	Boxes with/without checkmarks are visible/hidden in the other views. Add or remove check marks by clicking the boxes.
Type	Feature type.
Location	The sequence range where the feature is located.
Sequence	The name of the contig containing the feature.
Target Length	The length of the feature in bases.
Coverage Depth	The average depth of sequence coverage along the length of the feature.

The table below shows tasks that can be done in the Features view:

Copyright DNASTAR, Inc. 2022

Task	How To
Choose which features appear in the Features view	Use the drop-down menu on the left of the Features view header to show increasingly specific tables of features: **All Features**, **Genes** (only) and the subset of feature types chosen on the Input Reference Sequences page of the SeqMan NGen wizard (e.g. **CDSs and ncRNAs** only). 
Select all features of a specified type	Select a type of feature from the Features view (e.g., a CDS feature) and then use the menu commands **Features > Select > Select All Selected** (or **Select All of Type** or **Select Only of Type**).
Choose which features appear in the Strategy view	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 the following commands to add or remove checkmarks in groups. • To check (or uncheck) all rows currently selected in the Features view, use **Features > Show (or Hide) > All Selected**. • To check (or uncheck) all rows of the same feature type whose row is currently selected in the Features view, use **Features > Show (or Hide) > All of Type**. • To check (or uncheck) all rows of the type whose row is currently selected in the Features view while unchecking all other rows, use **Features > Show (or Hide) > Only of Type**.
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.
Export data from the table	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.
View features for constituent sequences	Display the Features track in the Alignment or Strategy views. If you are unable to see an expected feature, it may have been on part of the sequence that was trimmed as vector.
View features on the consensus sequence	Display the Features track in the Strategy view.

Copyright DNASTAR, Inc. 2022

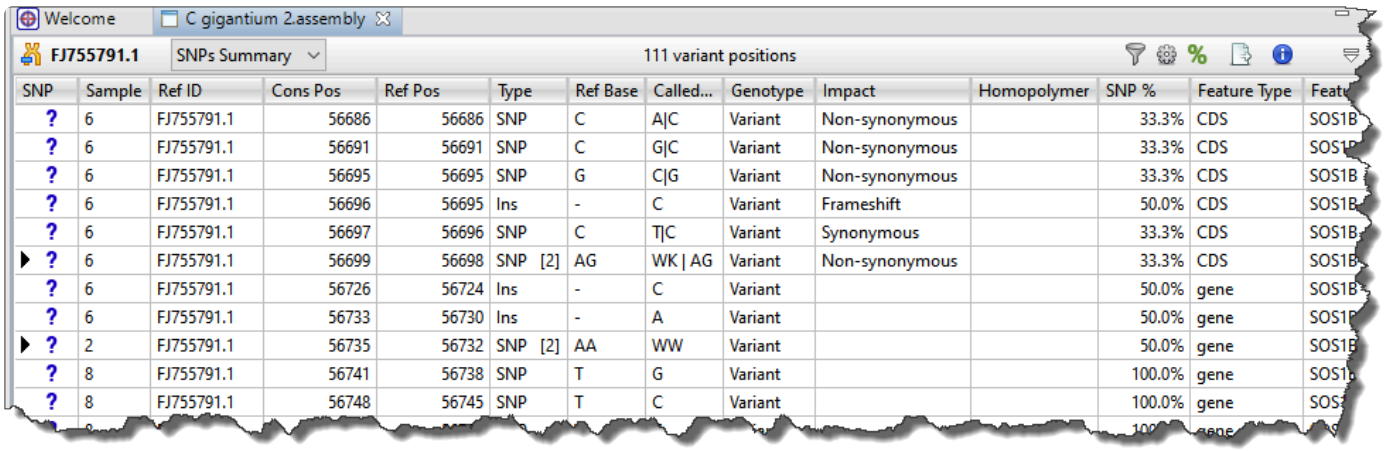
Simultaneously select a feature in multiple views.	When you select a feature in the Features view, the range of the feature will also be selected in the Alignment view, for features on constituent sequences; or the Strategy view, for features on the consensus sequence.

Copyright DNASTAR, Inc. 2022

Variants view

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	Featu
?	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	SOS1
?	8	FJ755791.1	56748	56745	SNP	T	C	Variant			100.0%	gene	SOS1

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.

The following table describes the various columns in this table:

Column Name	Description
SNP	A manual evaluation score for the SNP, with a question mark being the default. See **Change the status of a selected variant** in the table below for a legend and instructions.
Ref ID	The base at this position on the reference chromosome.
Cons Pos	The position on a gapped contig corresponding to this chromosome (for SeqMan Ultra and SeqMan NGen assemblies only).
Ref Pos	The position on the reference chromosome.
Type	Specifies the variation type as SNP, Del (deletion) or Ins (insertion). For .assembly files and certain SQD files (e.g., from de novo or special templated workflows), pink typeface may be used to indicate non-synonymous variants.
Ref Base	The base at this position on the reference chromosome.

Copyright DNASTAR, Inc. 2022

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 (-).
Genotype	Whether this SNP call indicates the reference or not, and (for diploids) whether this calls a homozygous or heterozygous genotype.
Impact	The impact of the variant or indel on the genome, displayed as one of the following values: • **Synonymous** – No amino acid changes. • **Non Synonymous** – Amino acid substitution only. • **Nonsense** – Amino acid to translational stop. • **Frameshift** – An indel within a coding region and which is not a multiple of 3, thereby changing the reading frame. • **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 Insertion** – An insertion within a coding region whose length is divisible by 3. The type is followed by the word Conservative if the insertion occurs between two codons, and Disruptive if it occurs with a codon. • **Inframe Deletion** – A deletion within a coding region whose length is divisible by 3. The type is followed by the word Conservative if the insertion occurs between two codons, and Disruptive if it occurs within two codons. If sorting by the Impact column, the column is ordered by severity. For example, a Frameshift is more severe than a Nonsense change.
Homopolymer	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 may not list all homopolymeric indels. When possible, insertions or deletions are placed at the 5’ end (top strand) of the run during alignment.
P Not Ref	The probability that this position does not match the reference. For combined SNPs and indels, P not ref will be the minimum of the P not refs in the used columns.
Q Call	The Phred-like quality score of the called genotype. It is a measure of the confidence that the SNP is present in the sample on a 0-60 log10 scale. For combined SNPs and indels, Q call will be the minimum of all available columns at that reference position.
SNP %	The percentage of the sequence at this position in the assembly which varied from the

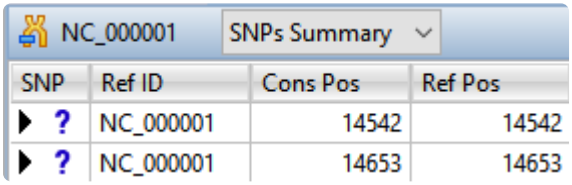
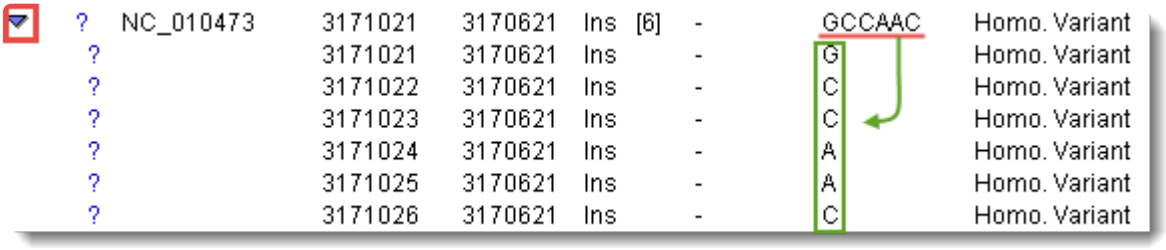
Copyright DNASTAR, Inc. 2022

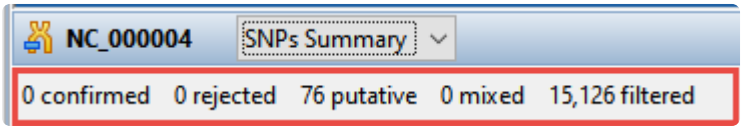
	reference.
Feature Type	The type of feature as annotated in the template sequence for the gene (e.g. CDS, attenuator, C_region, etc.).
Feature Name	
Transcript ID	
DNA Change	The change(s) in the DNA sequence using the nomenclature established by the Human Genome Variation Society (HGVS).
Amino Acid Change	The change(s) in the protein sequence using the nomenclature and conventions established by the Human Genome Variation Society (HGVS). The cell text 'p.(=)' signifies a synonymous change, while an empty cell denotes a non-coding region.
Depth	The number of reads overlapping the aligned column. Since this calculation disregards bases below the quality threshold, the Alignment View may show a greater number of sequences than the Depth shown in the Variants Summary Report. 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.
Deletion	The number of deleted bases in the Indel.
A (C, G, T) Cnt	The number of bases of this type called in the aligned column. A dash (-) represents the reference base.

The table below describes tasks related to the Variants view:

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 Alignment view 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 **Variants** menu to select a large group of variants at once. Regardless of the current selection, you can select **Variants > All** or **Variants > None**. 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. **Variants > Select > Select All of Type 'CDS'**). To select everything that is not currently selected, use **Variants > Invert Selection**. Alternatively, you can use the right-click menu options of the same names.
Change the status of selected	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. • ? – Putative (the initial status for all variants in the table)

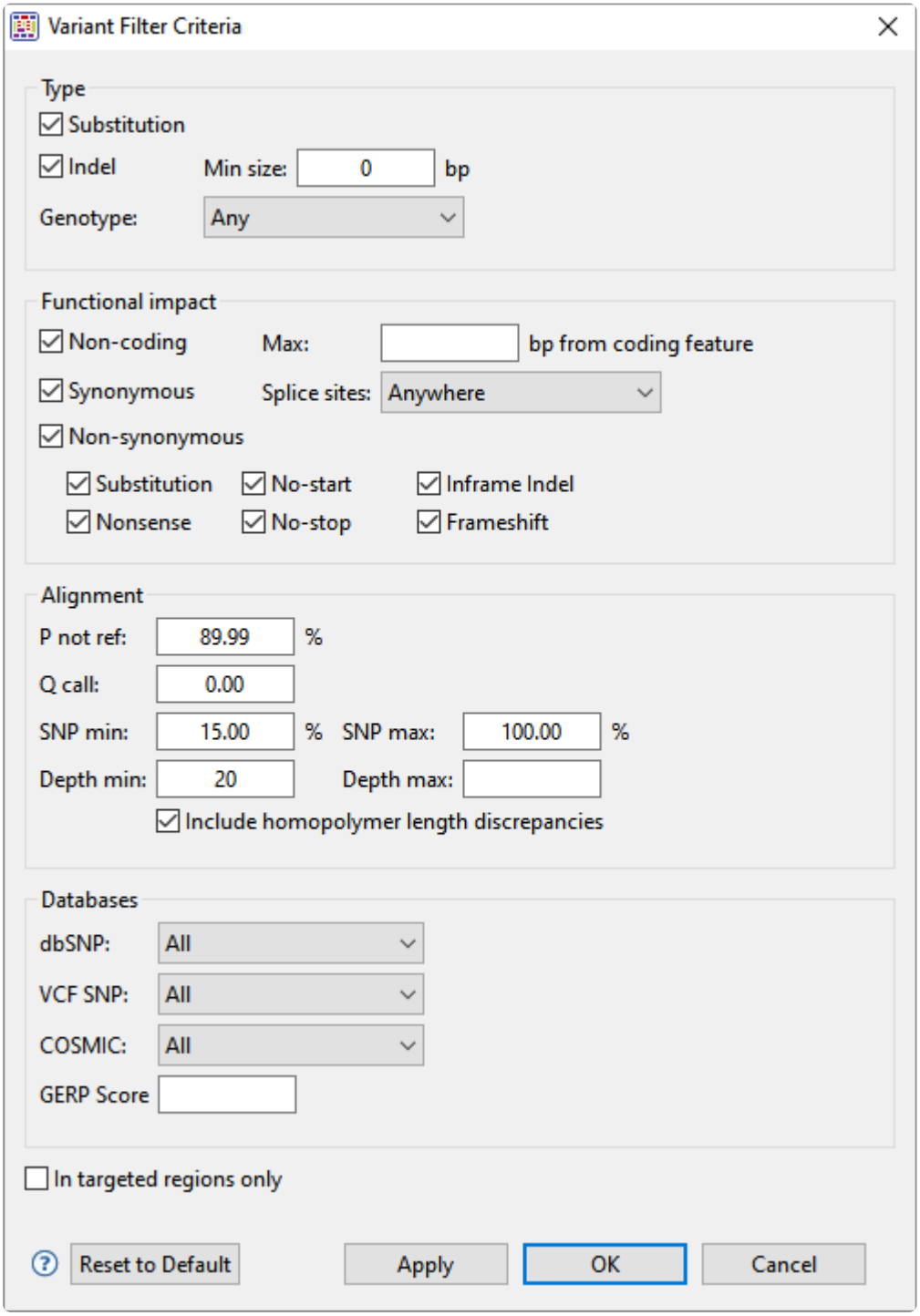
Copyright DNASTAR, Inc. 2022

variants	• ✓ – 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 following menu commands: **Variants > Confirm (Reject, Question) Selected Variants**.
Hide/show multiple references bases	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.  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. 
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	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.

columns	
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 the table	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. See Export data from the Variants view for details.
Toggle between showing/ hiding the table header	Use the **Show quick details** tool ( ) in the top right of the view. An example of the table header appears below boxed in red: 

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**.



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.

Copyright DNASTAR, Inc. 2022

- **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 vilter 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**.

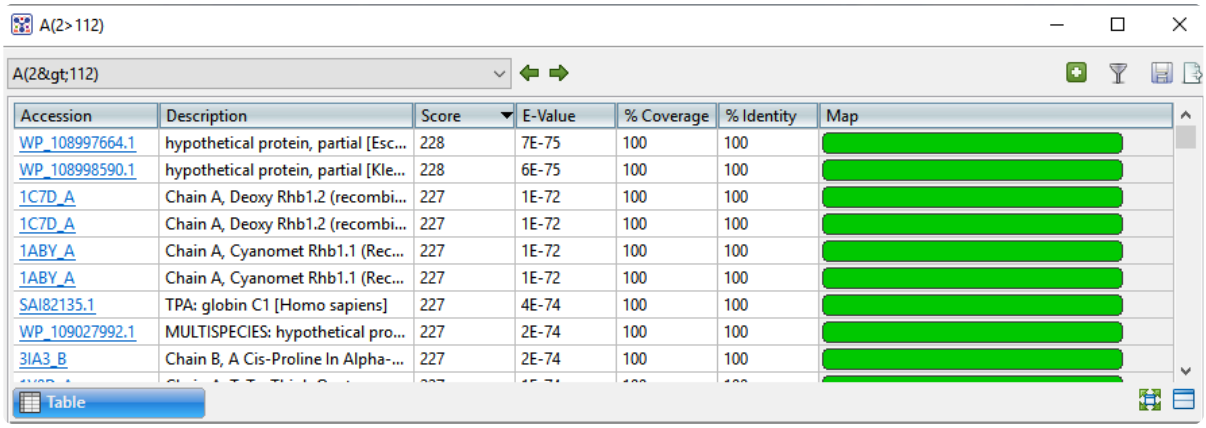
To save your settings, click **Apply** or **OK**, or simply press **Enter**.

Copyright DNASTAR, Inc. 2022

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:



Columns in the view’s table vary according to the search type. Some of the possible columns are described below:

Column	Description
Accession	The NCBI “Accession” number for the match. Press the link to open the corresponding page on the NCBI website.
Description	The NCBI “Definition” field text.
Length	The length of the sequence in the match.
Organism	The name of the organism associated with this match.
Score	The similarity score for the match. (See the NCBI definition for **raw score**). In general, a higher score denotes a better match.
E-Value	The chance that the match is expected to occur by chance. (See the NCBI definition for **E value**). In general, a lower score denotes a better match.
%Coverage	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. **%Coverage** = 565/596 = 95%.
%Identity	The percentage of bases that match when the query and result sequence are aligned. (NCBI definition).

Copyright DNASTAR, Inc. 2022

	Example: Query has 565 bases that align to the match and 3 bases that don't match. **%Identity** = 562/565 = 99.5%.
Map	Image showing areas of agreement between the query and the match.

Tasks pertaining to the Table view:

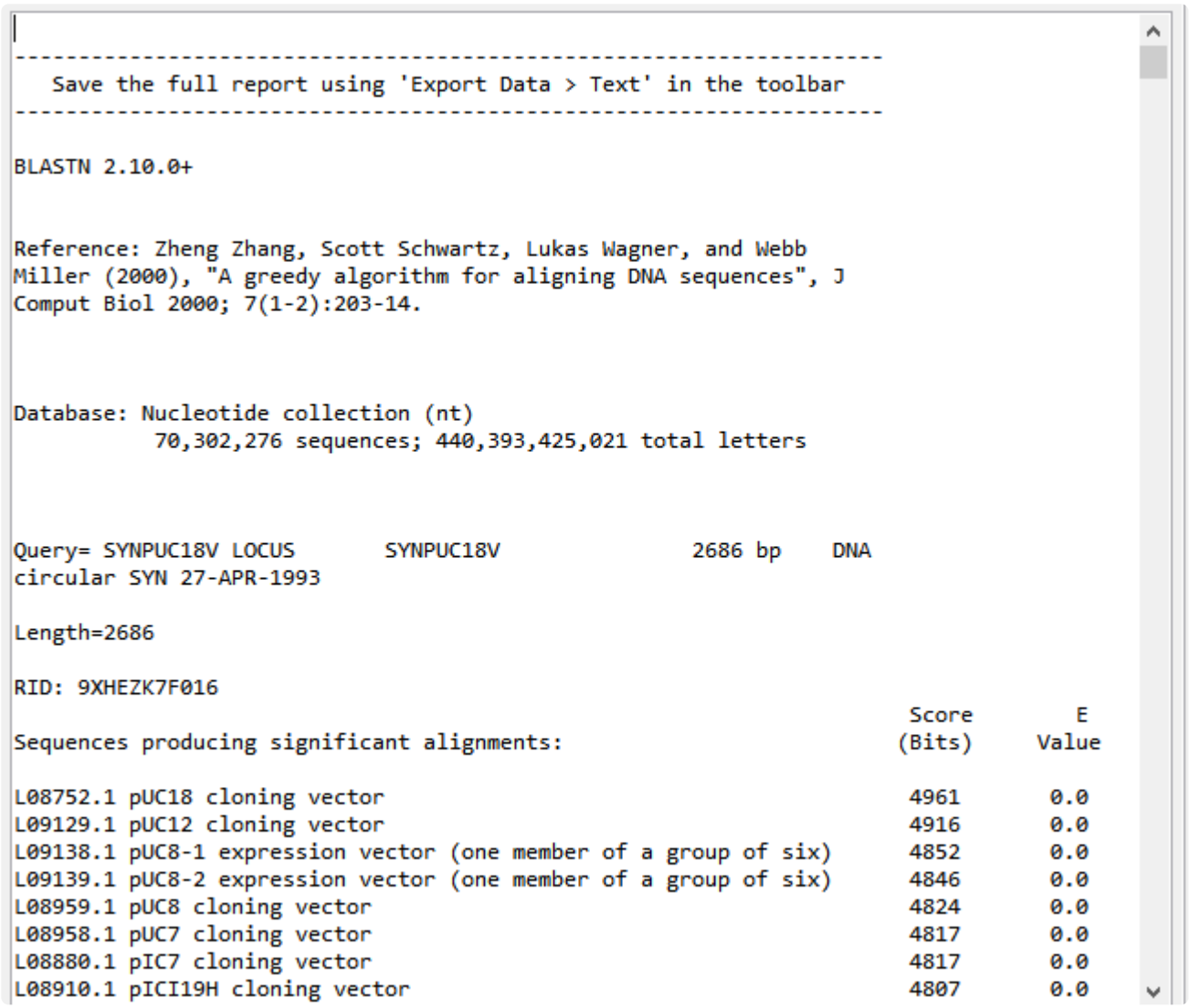
Task	How To…
If a BLAST or Entrez search was performed	
Launch the NCBI page for matching database entry	Click the hyperlink in the table's Accession column.
Copy all information from a search result row	Right-click on the row and choose **Copy**.
Save the sequence match in *.gbk* format	Right-click on the row and choose **Download sequences**. 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.
Add sequences or matches	Use the **Add sequences or matches** tool ( ). Sequences are added back into the host application, not the job's document.
Filter results from a sequence or text search	Use the **Filter results** tool ( ). Learn more about sequence search or text search filtering.
Export data from the view	Use the **Export data** tool ( ), described in the topic Export data from the Text, Table or Pairwise views.
Only if a BLAST search was performed	
To save results	Click the **Save as** tool ( ). A save dialog opens and allows you to save the search results as a *.search* format file. This option is not available for Entrez searches.
To display a pairwise alignment using a particular search result	If both the Table and Pairwise 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 **with** drop-down menu.

Copyright DNASTAR, Inc. 2022

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:



```
------------------------------------------------------------------
   Save the full report using 'Export Data > Text' in the toolbar
------------------------------------------------------------------

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.



Database: Nucleotide collection (nt)
           70,302,276 sequences; 440,393,425,021 total letters



Query= SYNPUC18V LOCUS       SYNPUC18V               2686 bp    DNA
circular SYN 27-APR-1993

Length=2686

RID: 9XHEZK7F016
                                                                  Score        E
Sequences producing significant alignments:                      (Bits)     Value

L08752.1 pUC18 cloning vector                                     4961       0.0
L09129.1 pUC12 cloning vector                                     4916       0.0
L09138.1 pUC8-1 expression vector (one member of a group of six)  4852       0.0
L09139.1 pUC8-2 expression vector (one member of a group of six)  4846       0.0
L08959.1 pUC8 cloning vector                                      4824       0.0
L08958.1 pUC7 cloning vector                                      4817       0.0
L08880.1 pIC7 cloning vector                                      4817       0.0
L08910.1 pICI19H cloning vector                                   4807       0.0
```

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.

Copyright DNASTAR, Inc. 2022

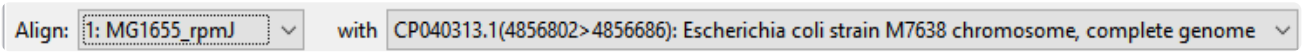
- Information about the Query sequence, including its length and the Request ID (RID) number.
- A table of sequences producing significant alignments.
- Pairwise alignments of each significant match with summary information on the match.

Copyright DNASTAR, Inc. 2022

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 1>117	CP040313.1 1>117	%Identity	%Gaps	Identical	Gap Count	Gap 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.

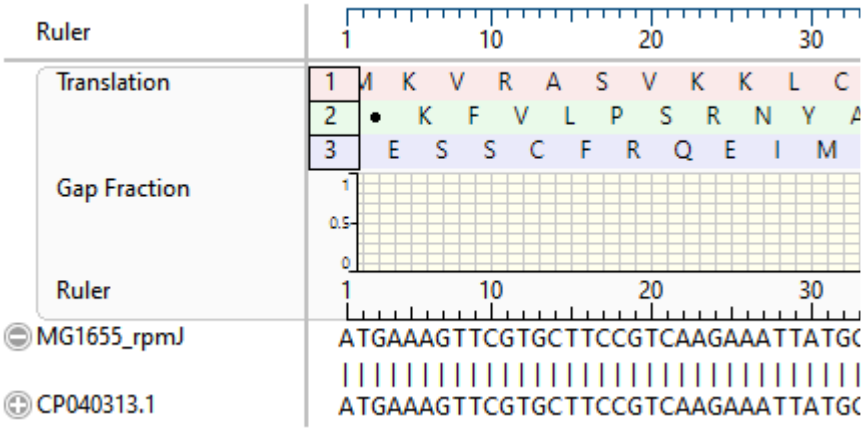
Value	Description
% Identity	Number of matching residues divided by the sum of all residues and gap characters. The quotient is multiplied by 100 to give the similarity as a percent.
% Gaps	The gap length divided by the consensus length. The quotient is multiplied by 100 to give the similarity as a percent.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

Copyright DNASTAR, Inc. 2022

Gap Length	Number of gap characters in an alignment.
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.

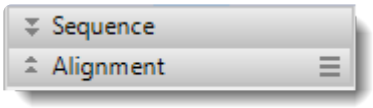


Panels

Settings panels allow you to optimize the look, contents and layout of the SeqMan Ultra window. To learn what can be done within each panel and the expandable “sections” within each panel, click the links below.

- Explorer panel
  - Sequences section
- Style panel
  - Sequence section
  - Alignment section
  - Strategy section
- Tracks panel
  - Layout section
  - Options section
- Details panel
- Jobs panel

The expand bars used to open panel areas have a “hamburger menu” on the right end of the bar.



Hamburger Menu Command	Description
Available for all expand bars	
Collapse Others	Collapses all expand bars in the panel except the one whose menu was used to choose this command.
Collapse All	Collapses all expand bars in the panel.
Available only for expand bars in the Style panel	
Apply Sequence Style	Lets you choose and apply a file of custom settings for the section controlled by this expand bar. See Style panel for more information.
Save Sequence Style	Saves the current settings in the section controlled by this expand bar. See Style panel for more information.
Save Sequence Style as Default	Saves the current settings in the section controlled by this expand bar to use as default settings in future projects. See Style panel for more information.
Restore DNASTAR Default Sequence Style	Restores the default DNASTAR settings for the section controlled by this expand bar. See Style panel for more information.

Copyright DNASTAR, Inc. 2022

Explorer panel

Basic contig and constituent sequence information can be viewed in the Explorer panel, represented by a compass icon ().

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)

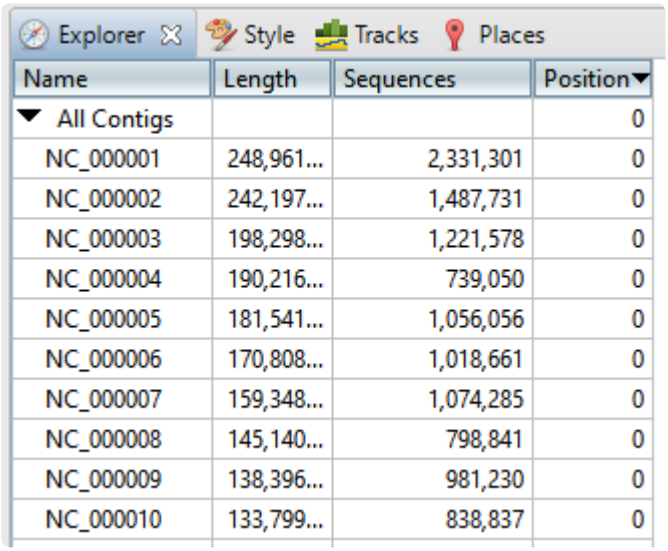
The Explorer panel consists of two expandable sections that are discussed separately:

- Contigs and Scaffolds section
- Sequences section

Copyright DNASTAR, Inc. 2022

Contigs and Scaffolds section

To view the Contigs and Scaffolds section, click on the Explorer panel expand bar entitled **Contigs and Scaffolds**. If you can't see the Explorer panel, choose **View > Explorer** to make it visible.



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

This section consists of 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.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
Name	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 **Contig > Split at Insertion** command, the new 5' contig keeps the original name and coordinates. The new 3' contig is given the identical name appended with "[2]".
Length	The length of the individual contig, including gaps and/or edits. Note that subtracting one **Name** coordinate from the other may yield a result that does not match the number shown in the **Length** column. This is because the **Name** coordinates are based on the ungapped and unedited

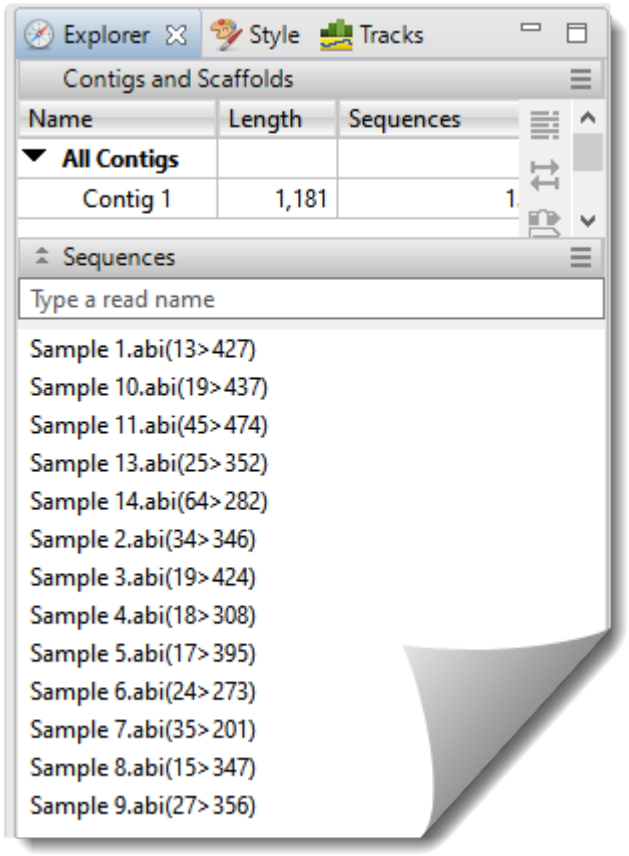
Copyright DNASTAR, Inc. 2022

	reference sequence, while the **Length** value includes gaps and edits.
Sequences	The number of sequences in the contig.
Position	The position of each ordered contig or “scaffold. A “0” indicates that contigs or scaffolds have not been ordered.

Copyright DNASTAR, Inc. 2022

Sequences section

To view the Sequences section, click on the Explorer panel expand bar entitled **Sequences**. If you can't see the Explorer panel, choose **View > Explorer** to make it visible.



Sequences appear alphabetically by name and are followed with the sequence range that was used in the assembly. Portions that were trimmed (e.g., due to poor quality or the presence of vector) prior to the assembly are omitted from the range.

To open the Alignment view for the contig in which the sequence appears, double-click on the name of the sequence in the list. If there are numerous sequences on the list, navigate to the sequence of interest by starting to type its name in the search bar at the top of the section.

Copyright DNASTAR, Inc. 2022

Style panel

The Style panel has expandable sections with options affecting the Alignment and Strategy views.

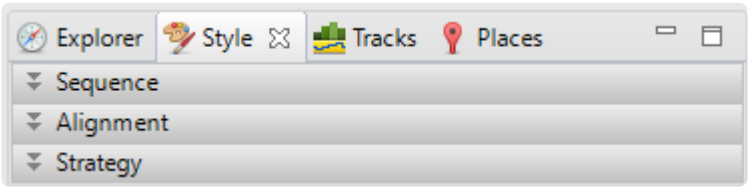
To open the Style panel:

Click on the **Style** tab (

). If the panel is not already visible, reveal it using any of the following methods:

- Choose **View > Style > Style** or **View > Reset Panel Layout**.
- Press **Ctrl/Cmd+Alt+Y**.
- Right-click on a track and choose **Show Style Panel**.

To change settings in the Style panel:



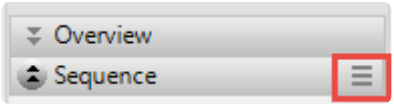
Click the links below to learn about each of the sections in the Style panel:

- Sequence section – affects both the Alignment and Strategy views
- Alignment section – affects the Alignment view
- Strategy section – affects the Strategy view

To save styles and apply them to future projects:

Once you have set up styles, you can save them and then apply them to future projects, either by default or only when you specify that they be applied. You can choose whether to save and apply styles for just one particular view or for all views.

Note that some of the instructions below refer to the “hamburger” menu. To access commands on these menus, click on the expand bar of interest in the Style panel. Then click on the tool consisting of three horizontal lines.



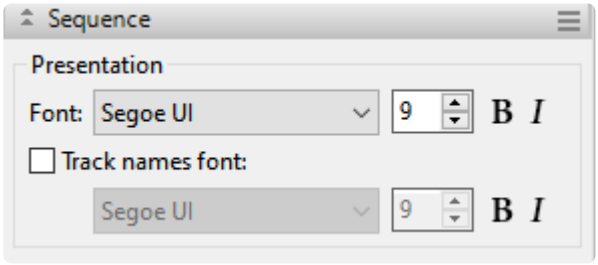
Copyright DNASTAR, Inc. 2022

Task	How To
Save all current styles to a file that can be applied to future projects	Use **View > Document Style > Save Style**.
Save the styles from one particular view to a file that can be applied to that view in future projects	Click the “hamburger” menu on the right of a Style panel expand bar and choose **Save (Expand bar name) Style**.
Save all current styles to a file that will be applied as the default for future projects	Use **View > Document Style > Save Style as Default**.
Save the current styles from one particular view to a file that will be applied as the default for that view in future projects	Click the “hamburger” menu on the right of a Style panel expand bar and choose **Save (Expand bar name) Style as Default**.
Apply styles that were saved earlier to the current project	Use **View > Document Style > Apply Style**
Apply styles that were saved for this view previously and apply them to the same view in the current project	Click the “hamburger” menuon the right of a Style panel expand bar and choose **Apply (Expand bar name) Style**.
Restore all DNASTAR default styles to the current project	Use **View > Document Style > Restore DNASTAR Default Style**.
Restore the DNASTAR default style for one particular view to that view in the current project	Click the “hamburger” menu on the right of a Style panel expand bar and choose **Restore DNASTAR Default (Expand bar name) Style**.

Copyright DNASTAR, Inc. 2022

Sequence section

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

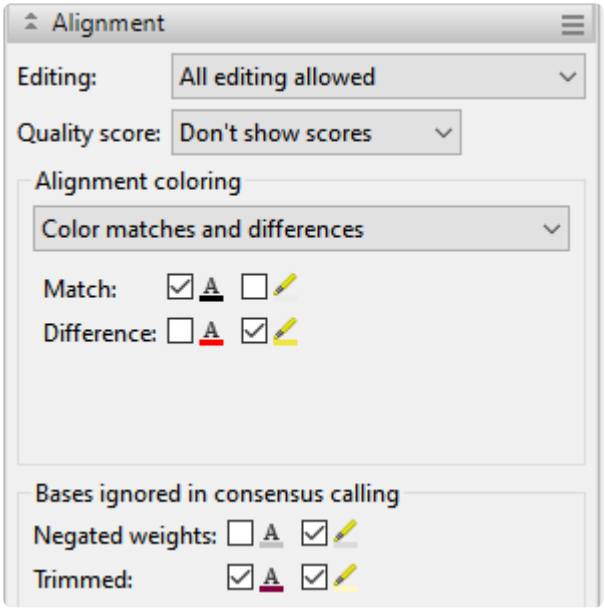


Task	How to…
Select font options for sequence names	Select a font from the **Font** drop-down menu. Recently selected fonts (if any) appear at the top of the menu, and other available fonts appear below the row of dashes. Enter the preferred font size by typing a number between 4-72 in the box at right, or by using the corresponding up and down arrows ( ). To make the font **Bold** and/or *Italic*, press the **B** and/or ***I*** buttons, respectively.
Select font options for tracks	Check the box next to **Track names font**, 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 ( ). To make the font **Bold** and/or *Italic*, press the **B** and/or ***I*** buttons, respectively.

Copyright DNASTAR, Inc. 2022

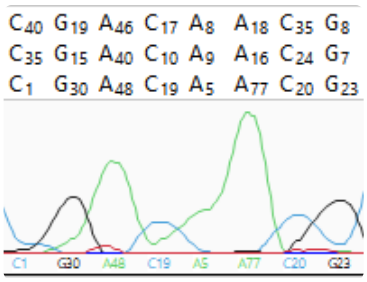
Alignment section

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

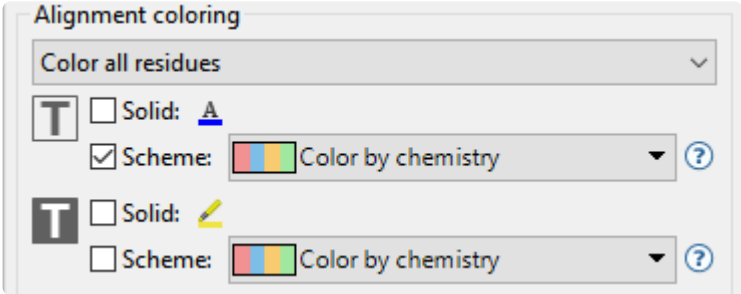


Task	How to…
Specify the type(s) of editing allowed in an *.sqd* assembly	Make a selection from the **Editing** drop-down menu. Choices are: • **All editing allowed** – Allows you to type in new nucleotides or gaps into the consensus sequence, or to delete the same. To add gaps, use dashes (-). • **Only gap editing allowed** – Allows you to add gaps using dashes (-) or to remove gap-indicating dashes. Does not allow nucleotide characters to be added or deleted. • **No editing allowed** – 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 **Yes**, the **Editing** menu will be switched to **All editing allowed**. • **No editing allowed permanently** – This option is similar to **No editing allowed**. However, the only way to override the option is to manually make a different selection from the **Editing** menu.
Limit color-coding to the	Check the **Color only in targeted coverage regions** box to cause coloring options selected elsewhere in this section to only be applied to nucleotides in the targeted region of the assembly.

Copyright DNASTAR, Inc. 2022

targeted coverage region only	
Display quality scores	To display quality scores as subscripts next to each nucleotide, check the **Quality** box and then make a selection from the drop-down menu to the right: • **Show scores** – Causes quality scores to be displayed as a subscript next to the base. When bases with very high quality scores (>99% of the 95th percentile) are displayed, two asterisks (**) will be shown in place of the usual numerical subscripts. Example: G_{52} A_{84} G** C_{85}.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 *.fas* files, SeqMan NGen automatically checks for *.qual* files with the same name in the same folder. If found, the quality values are displayed. Note that gaps are not assigned quality scores.  • **Show averaged scores** – Displays quality scores averaged over a defined window. This mode is useful for examining trimming. Note that gaps are not assigned quality scores.
Specify where background and foreground coloring should be applied	To show where background and/or foreground coloring should be applied, make a selection from the **Alignment coloring** drop-down menu. This option is only active if at least one of the **Color sequence foreground** or **Color sequence background** boxes are checked in the **Sequence** section described above. Options include: **No special coloring**, **Color only differences from consensus**, **Color only matches to consensus**, **Show only differences from consensus**, **Color matches and differences**, and **Color only SNPs**. Choosing any of the first four menu options enables the following options.

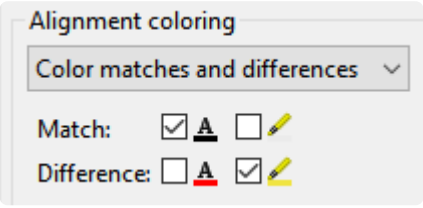
Copyright DNASTAR, Inc. 2022



To specify the color of sequence characters, use the “foreground color” tools to the right of T. To display the letters in a single solid color of your choice, check the box next to **Solid**. Then click on the ‘A’ tool to the right of the checkbox and choose a color from the color picker. To instead choose a preset color scheme, check the box next to **Scheme** and make a selection from the drop-down menu. Color scheme choices are described in Available color schemes.

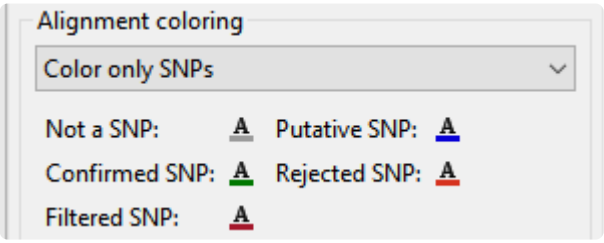
To specify the background color (boxes around each letter), use the “background color” tools to the right of T. To use a single solid color of your choice, check the box next to **Solid**. Then click on the ‘highlighter pen’ tool to the right of the checkbox and choose a color from the color picker. To instead choose a preset color scheme, check the box next to **Scheme** and make a selection from the drop-down menu. Color scheme choices are described in Available color schemes

Choosing **Color matches and differences** enables the following options.



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.

Choosing **Color only SNPs** enables a different set of options.

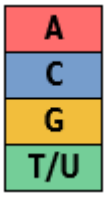
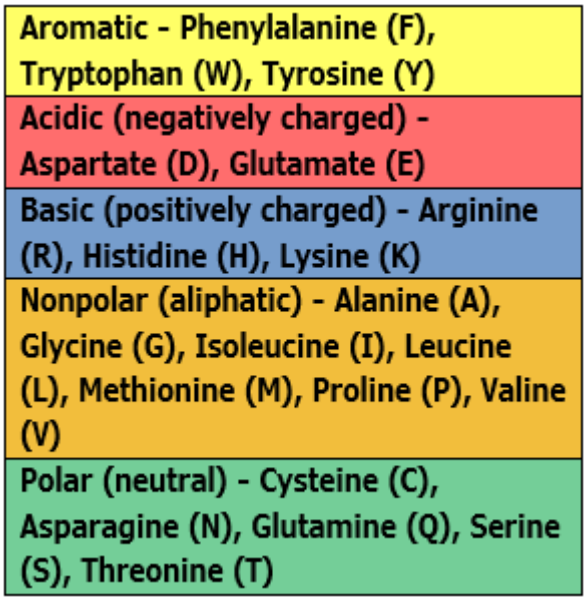
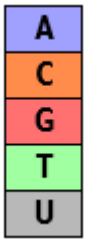
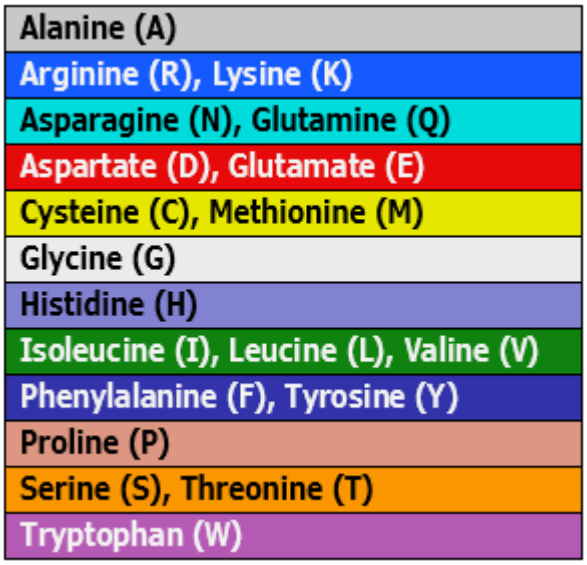


- **Not a SNP** – A base that did not meet the minimum threshold to be called as a SNP

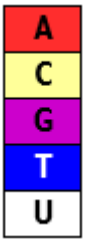
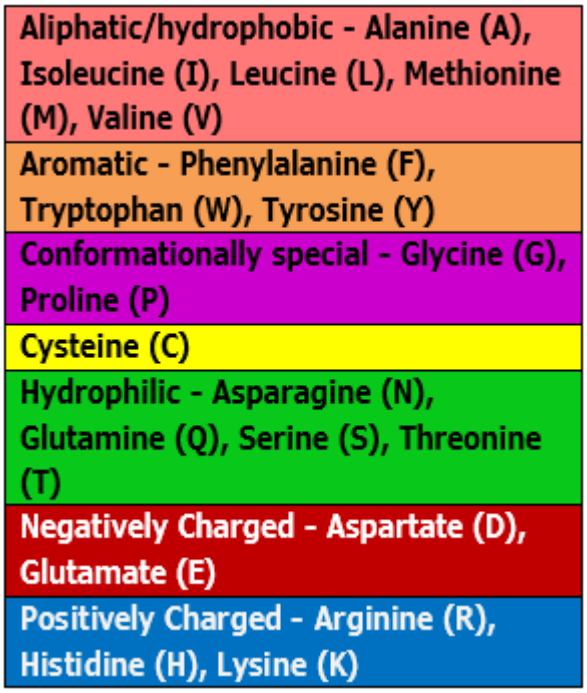
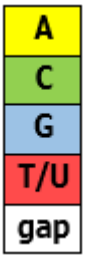
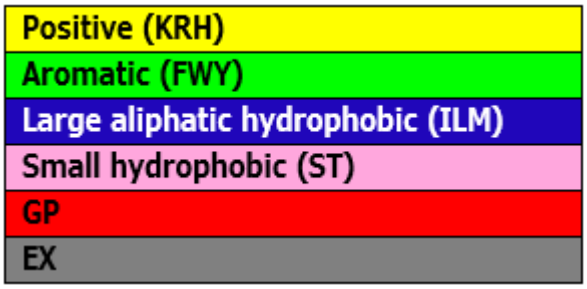
	and that does not appear in the Variants view. Default color is gray (). • **Putative SNP** – A base that met or surpassed the minimum SNP calling threshold, and therefore contains a question mark in the Variants view **SNP** column. Default color is blue (). If you apply a more stringent filter in the Variants view, this base will change to “red strikethrough” coloration. • **Confirmed SNP** – A called SNP that was manually validated and therefore has a checkmark in the Variants view **SNP** column. Default color is green . • **Rejected SNP** – A base that was called as a SNP during assembly in SeqMan NGen, but which was was manually rejected and therefore has a red ‘x’ in the Variants view **SNP** column. Default color is red (). The same red color is used by default to indicate a mismatched base that was filtered at the time of assembly using SeqMan NGen and thus never appeared in the Variants view. In most cases, the mismatched base is not a SNP, but a base-calling error; the background “noise” contained by all sequencing data. If you believe this base to be a SNP, you cannot undo this filter in SeqMan Ultra. Instead, you would need to rerun the assembly with the variant caller set to a more relaxed stringency setting. • **Filtered SNP** – A base that was initially called as a SNP but was filtered out by setting more stringent filtering criteria in the Variants view. Default color is red strikethrough (). Relaxing the filtering criteria can cause the base to revert back to the blue coloring of a putative SNP.
Display negated weights or trimmed sequence with a custom font color or highlighter	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.

Available color schemes

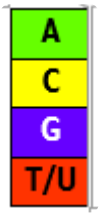
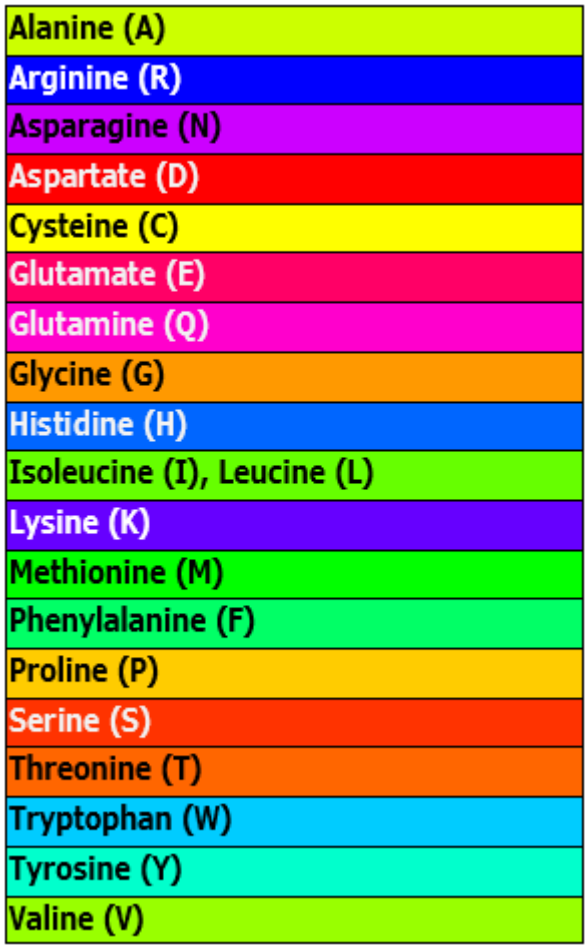
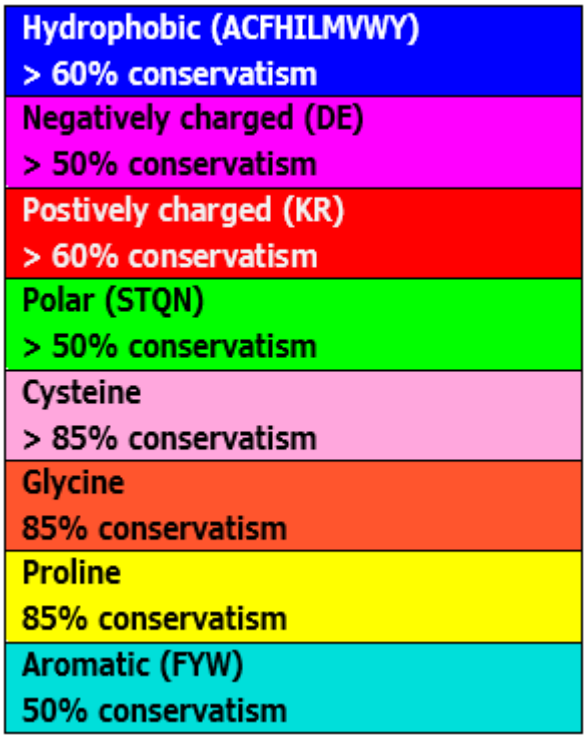
The Sequence section allows you to select color schemes for sequences in the Alignment view. Each available color scheme is described below:

Color Scheme	Description	Nucleotide Legend	Amino Acid Legend
Color by Chemistry	The default color scheme used by multiple DNASTAR applications. Amino acids are colored according to their side chain chemistry. While Color by Chemistry can be used for nucleotide sequences too, we recommend instead using a **Solid** background color for nucleotides and then selecting a **Comparison** like **Color only nonconserved residues**, **Color only differences from reference**, etc.		
Physiochemical: Shapely	Matches the RasMol amino acid and RasMol nucleotide color schemes, which are, in turn, based on Robert Fletterick's "Shapely models."		

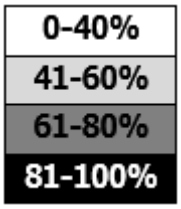
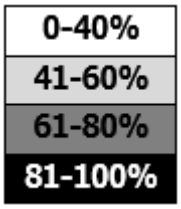
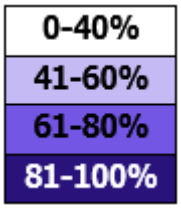
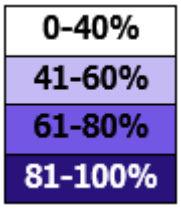
Copyright DNASTAR, Inc. 2022

Physiochemical: Zappo	Colors residues according to their physico-chemical properties, and is also used in JalView (Waterhouse et al., 2009).		
Physiochemical: UGene	Unipro UGENE coloring scheme.		

Copyright DNASTAR, Inc. 2022

Spectral Colors: Taylor	Taken from Taylor (1997) and also used in JalView (Waterhouse et al., 2009).		
Clustal X	Taken from Thompson et al. (1994). This color scheme is only available for protein alignments.	N/A	

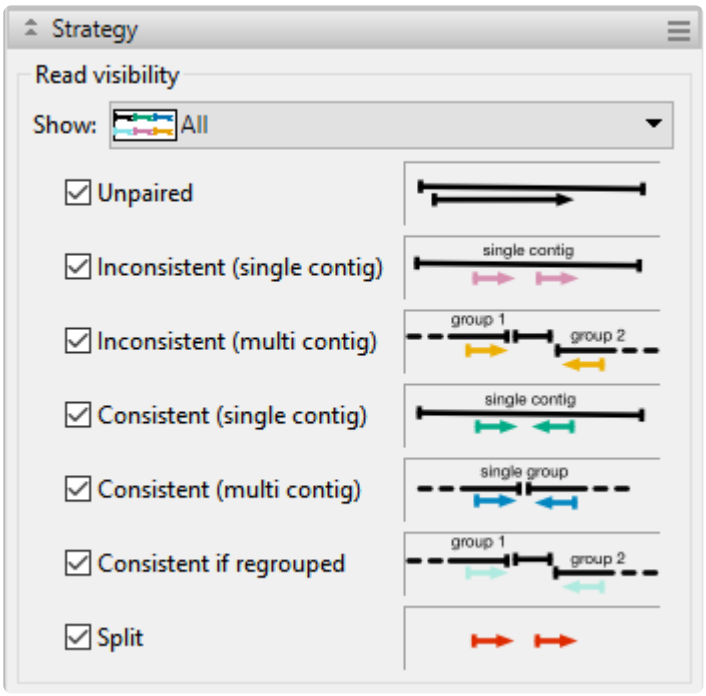
Copyright DNASTAR, Inc. 2022

Grayscale – Percentage Identity	This option uses only black/white/gray and is suitable for publications requiring grayscale images.		
Percentage Identity	This option uses only shades of slate blue.		

Copyright DNASTAR, Inc. 2022

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**.



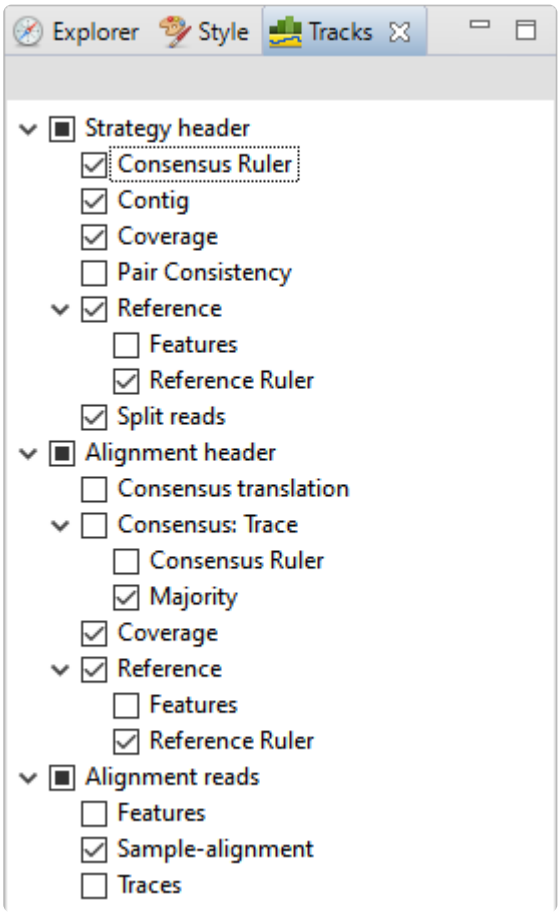
Task	How to…
Show or hide specific types of paired reads in the Strategy view	Choose an option from the **Show** drop-down menu or manually check or uncheck the individual boxes: **Unpaired**, **Inconsistent (single contig**, **Inconsistent (multi contig)**, **Consistent (single contig)**, **Consistent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

Copyright DNASTAR, Inc. 2022

Tracks panel

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.

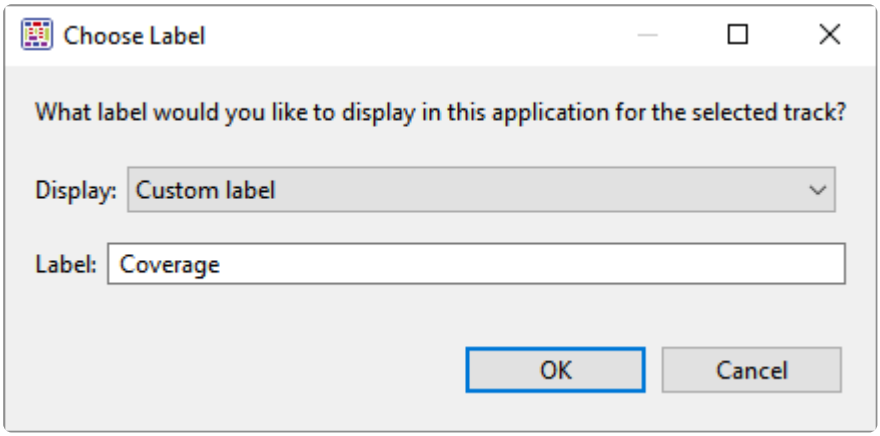
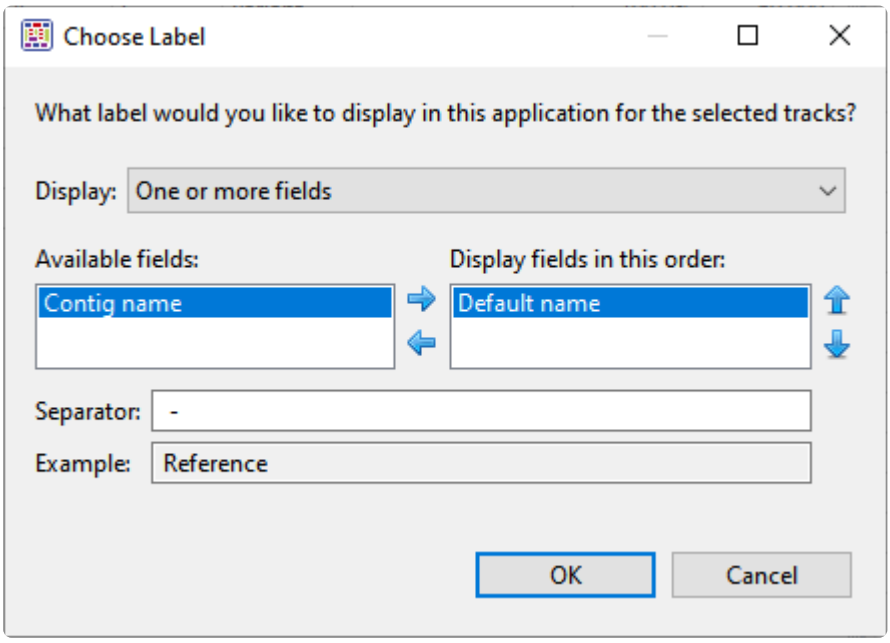
The following table shows tasks that affect the display of tracks:

Task	How to…
To display a track	To display a track in the Alignment and/or Strategy views: 1. Open the Tracks panel by clicking on the **Tracks** tab (Tracks). If the tab is not visible, reveal it using Choose **View > Tracks > Tracks** or **View > Reset Panel Layout**. 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. 

Copyright DNASTAR, Inc. 2022

	2. 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. 3. In the Alignment or Strategy view(s) click on the plus sign next to the track name to reveal it.
To expand/ collapse selected tracks	Click the plus/minus signs (/) to the left of the sequence name. In the Alignment view, you can also right-click on a track and choose **Expand** or **Collapse**.
To expand/ collapse all tracks in the Alignment view	Right-click anywhere in the Alignment view and choose **Expand All** or “Collapse All*.
To hide tracks	Do one of the following: • To hide a specific track from all views, uncheck the box next to the name of the track in the Tracks panel (see top row of this table). • To hide a specific track in one view only, right-click on the track within that view and choose **Hide Track**. • To remove a selected track as if unchecking its box in the Tracks panel, use **Edit > Delete** or press the **Delete** key.
To change the order of tracks in a view	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.
To change options for a track	See Edit track options.
To change track height or spacing between tracks	See Change track height or spacing between tracks.

Copyright DNASTAR, Inc. 2022

To manually rename a single selected track	Select a track in the Alignment or Strategy view, then right-click on it and choose **Rename**. Alternatively, select the track name and choose **Edit > Rename**. The Choose Label dialog opens with the **Display** drop-down menu set to **Custom label**.  Type the desired name into the **Label** text box, then click **OK**.
To automatically rename tracks using specified data fields	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**.  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:

Copyright DNASTAR, Inc. 2022

- 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.

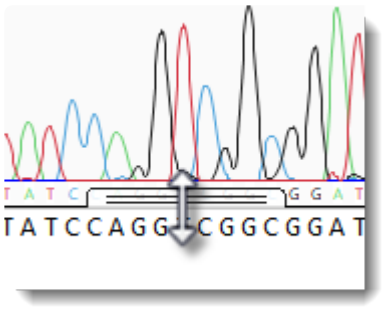
- By default, a hyphen (-) is included between each of the displayed fields. If desired, type in a different **Separator**. Or remove all text in the **Separator** box to include nothing between the fields.

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

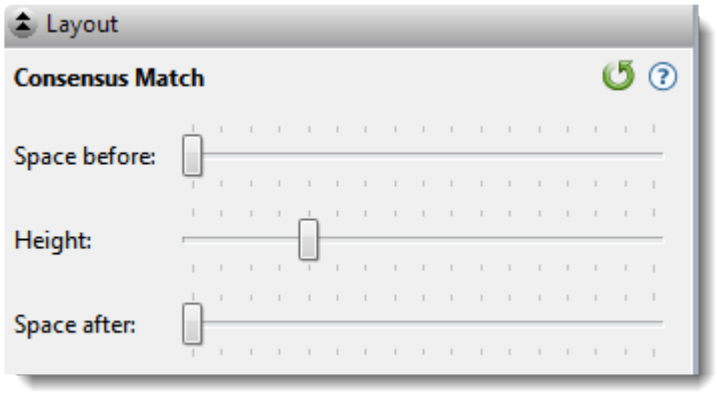
  When you are satisfied with the choices made in this dialog, click **OK**

Layout section

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.

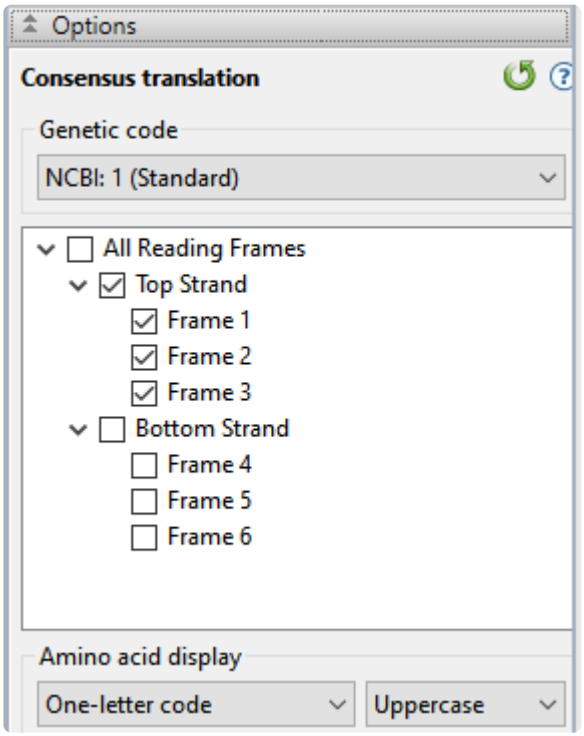
If you want to return the sliders in this dialog to their default settings+, click the **Restore default settings** tool ( ). |

Copyright DNASTAR, Inc. 2022

Options section

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 (Options).



- 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. |

Copyright DNASTAR, Inc. 2022

Details panel

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 Welcome screen is open	**Open an assembly** launches a file chooser **Watch a training video** opens the **Training** page on the DNASTAR website. The **Recent Documents** section provides links for opening a recent document, if any exist.
A single read sequence **Selected Read** **NC_002973 selection:** ID_1530469 NC_002973: 1 read	Contig name, read ID, and the number of reads selected (i.e. "1").
Multiple read sequences **Selected Reads** **NC_002973 selection:** NC_002973: 2 reads	Contig name and the number of reads selected.
A portion of the consensus sequence **Selected Region** **NC_002973 selection:** 16526> 19276 = 2751	Contig name and the inclusive coordinates of the beginning and end of the selection
A portion of a read sequence **Selected Region** **Sequence:** ID_1396536_r1__dup[40] **Sequence selection:** 8<23 = 16	Read ID and the inclusive coordinates of the beginning and end of the selection
One or more tracks	See View information about a selected track and View information about multiple selected tracks.

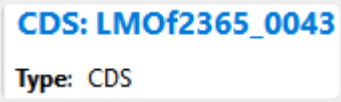
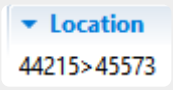
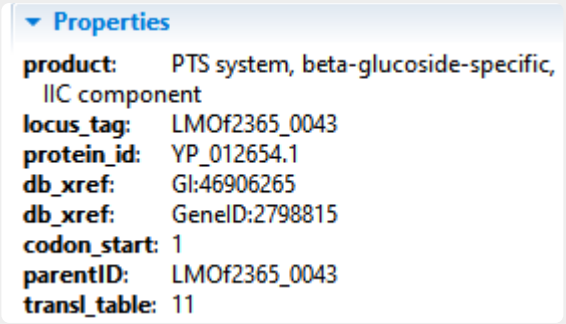
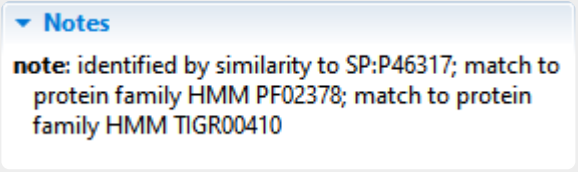
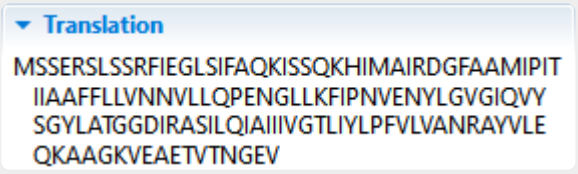
Copyright DNASTAR, Inc. 2022

A single feature	See View information about a selected feature.
Multiple features **2 features** **Type:** CDS, gene	The number of features selected and the feature type(s).

Copyright DNASTAR, Inc. 2022

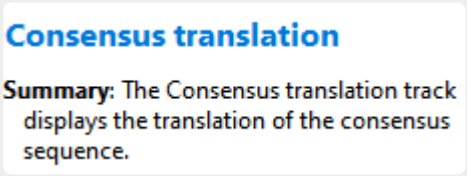
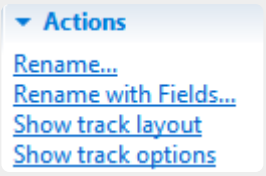
View information about a selected feature

If a single feature has been selected in a Feature track, the Details panel contains the following information:

Heading 	Feature name (in blue) and type.
Location 	Inclusive range of the feature in nucleotides.
Properties 	Properties shown vary by feature and may include the organism, gene name NCBI feature qualifiers, etc.
Notes 	The NCBI feature qualifier “/note”.
Translation 	The translated sequence shown in IUPAC amino acid codes.

View information about a selected track

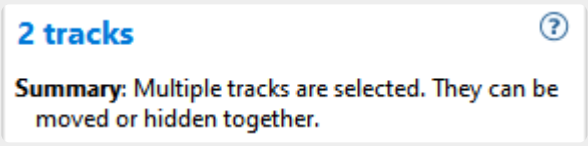
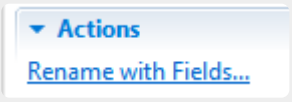
If a single track has been selected in a view or panel, the Details panel contains the following information.

Heading 	Type of track (in blue). **Summary** – Basic information about what is displayed in the selected track.
Actions 	Only a subset of links may be available, depending on the selected track. • **Rename** – Shortcut to **Edit > Rename** (See Work with Data Tracks). • **Rename with Fields** – Shortcut to **Edit > Rename with Fields** (See Work with Data Tracks). • **Show track layout** – Shortcut to **View > Tracks > Layout**. (See Change track height or spacing between tracks). • **Show track options** – Shortcut to **View > Tracks > Options**. (See Edit track options).

Copyright DNASTAR, Inc. 2022

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.

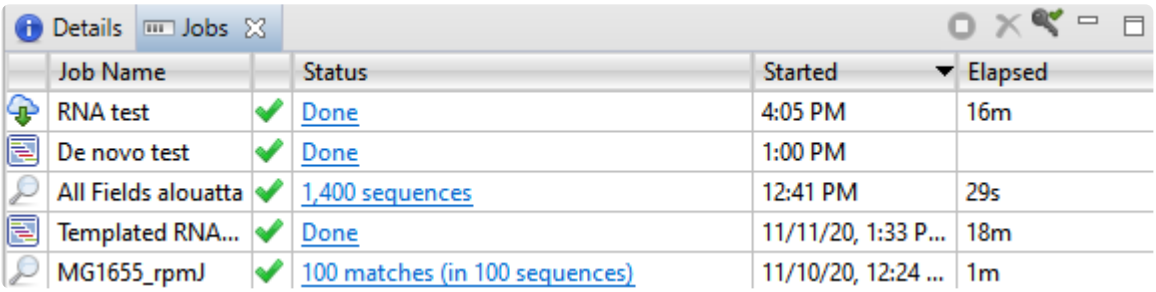
Heading 	Number of tracks currently selected (in blue). **Summary** – Description of options available for the multiple track selection.
Actions 	**Rename with Fields** – Shortcut to **Edit > Rename with Fields** (See Work with Data Tracks).

Jobs panel

After initiating an assembly in SeqMan NGen or SeqMan Ultra or an NCBI BLAST or Entrez search in SeqMan Ultra, you can monitor its progress and open its eventual output from the Jobs panel.

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 Jobs**.



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

Column	Description
(icon column)	– Local assembly that started in SeqMan Ultra or SeqMan NGen. – Cloud assembly that started in SeqMan Ultra or SeqMan NGen. – Text or sequence search of the NCBI database.
Job Name	Assemblies originating in SeqMan Ultra or SeqMan NGen are named as specified during assembly setup. Text searches are all named with the field and term of the first search criterion followed by "plus # terms," where # represents the number of additional criteria. Sequence searches use the name specified in the Job screen of the Search wizard. The same **Job Name** can be used for more than one prediction.
(icon column)	An icon indicating the status of the search (see below).
Status	A word or phrase indicating the status of the search. • **Started** – **Run** or **Run Job** were pressed in the Search wizard • **Queued** – – Submitted and placed into the queue • **Running** – – Running • **Done** – – Job completed successfully; click the link to view the results. Instead of **Done**, some successful searches will instead display the link as **'n' matches (in 'y' sequences)**. • **Failed** – – Job failed

Copyright DNASTAR, Inc. 2022

	• **Stopped** – – Job stopped by user • **Expiring** – – 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. • "The assembly has failed" message, followed by a description • "Too many matches found" message, followed by a description
Started	The date and time the job began.
Elapsed	The total elapsed time between starting and completing the job.

Tasks that pertain to the Jobs panel:

Task	How To
Open the results of one or more jobs	Click on the **Done** (Done) or **Expiring** links for that job. See below this table for more information on job expiration. You can also open a saved *.search* file using **Search > Open Search File**.
Stop one or more jobs that are in progress	Select one or more rows with a Status of **Started**, **Queued** or **Running**. Then press the **Stop** tool ( ) in the header above the table. Once stopped, a job cannot be restarted from the Jobs panel. You must instead initiate a new search.
Delete a finished or failed job	Select one or more rows with a Status of **Done**, **Failed** or **Stopped**. Then press the **Delete** tool ( ) in the header above the table.
Log in to Cloud Assemblies	If you see a key icon with a yellow triangle ( ), this indicates that you are not logged into Cloud Assemblies and that any Cloud Assemblies will not show up in the Jobs list. Click the icon to enter your login credentials. If you are logged in, the key icon is accompanied by a green check mark ( ).
Open/ reopen the progress dialog for a job	Select one or more rows with a Status of **Started**, **Queued** or **Running**. Then right-click and choose **Open Progress dialog**.

Copyright DNASTAR, Inc. 2022

Assemble Sequences

Starting in Lasergene 17.2, you can use the Welcome screen to launch any type of assembly from within SeqMan Ultra or SeqMan NGen, as desired.

If you begin in SeqMan Ultra, and have a small-to-medium sized Sanger data set, assembly will take place in SeqMan Ultra itself. If you are working with a large Sanger data set, multi-sample Sanger data, or any type of NGS data, you will be seamlessly transferred to SeqMan NGen to set up and run the assembly. Use the links below to see the steps for each type of assembly.

- Assemble small-to-medium sets of Sanger data
- Assemble multi-sample Sanger data
- Assemble NGS data or large sets of Sanger data

Copyright DNASTAR, Inc. 2022

Assemble small-to-medium sets of Sanger data

You can assemble small-to-medium sized Sanger data sets in SeqMan Ultra itself. (Other types of data need to be assembled in SeqMan NGen.)

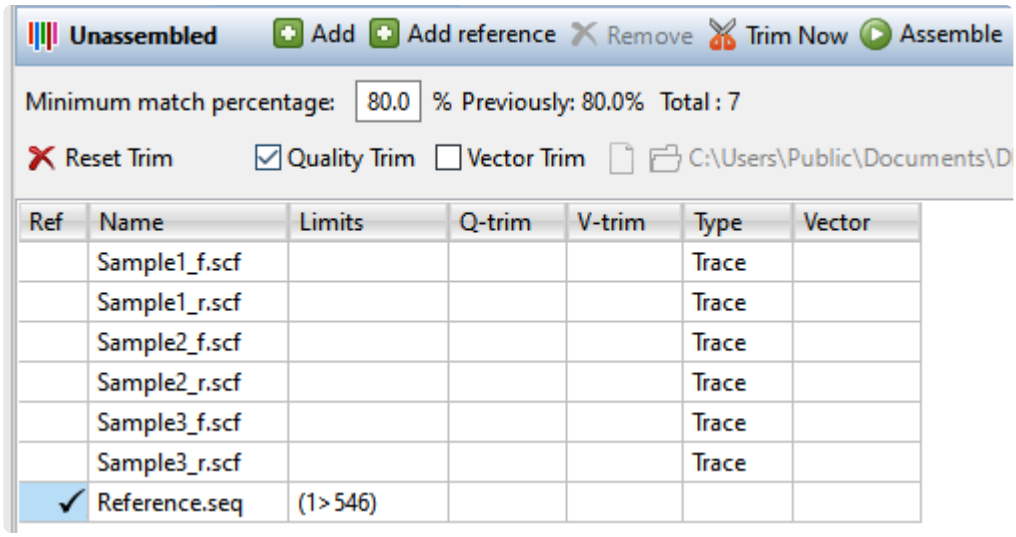
To assemble small-to-medium sets of Sanger data:

1. Do any of the following:

   - Click on the **Welcome** project tab or choose **View > Welcome** to open the Welcome screen. Click on **New Assembly** on the left. Then choose **New Sanger/ABI assembly** from the **Molecular biology** section.
   - Choose **File > New Sanger/ABI Assembly**.
   - Press **Ctrl/Cmd+N**.

   In each case, the Unassembled view will open.

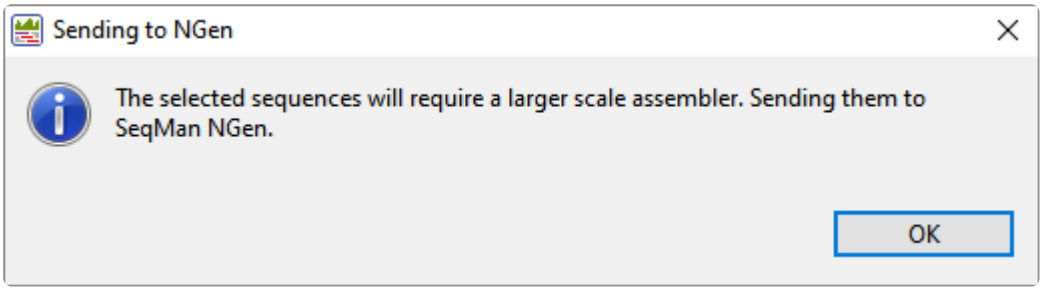
2. In the Unassembled view, upload the sample sequences and, optionally, a reference sequence. To upload sample sequences and their trace data (if available), press the **Add** button (Add) or choose **Edit > Add**. To upload reference sequences (optional), press the **Add reference** button (Add reference). In both cases, select the desired files and press **Open**. Files added using the **Add reference** button will appear in the table with a checkmark in the **Ref** column. Sample files that contain trace data will show the word **Trace** in the **Type** column.



Ref	Name	Limits	Q-trim	V-trim	Type	Vector
	Sample1_f.scf				Trace	
	Sample1_r.scf				Trace	
	Sample2_f.scf				Trace	
	Sample2_r.scf				Trace	
	Sample3_f.scf				Trace	
	Sample3_r.scf				Trace	
✓	Reference.seq	(1>546)				

p(banner tip). **Note:** If you try to add large files or non-Sanger sequences, you will receive a message notifying you that SeqMan NGen must be used for the assembly.

Copyright DNASTAR, Inc. 2022



Press **OK** to launch SeqMan NGen at the **Input Sequences** wizard screen.

3. (optional) If you need to remove rows from the table, select them and press the **Remove** tool (Remove). Prior to removing rows, you can verify how many rows are selected by looking in the middle of the blue header between the two groups of tools.

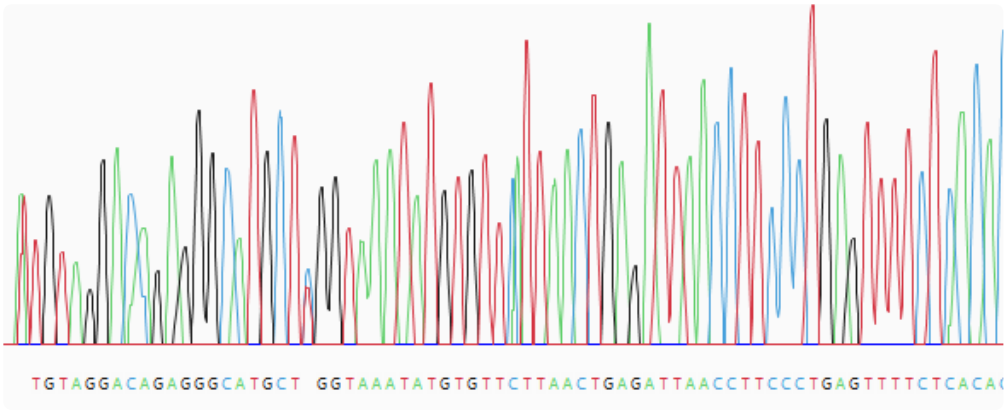
   5 selected rows.

4. (optional) If you wish to perform automatic vector removal, start by checking the **Vector Trim** box. By default, SeqMan Ultra will remove vector sequence from any of the vectors in the *VectorData* catalog installed with Lasergene. If desired, you can override this by using the tools to the right of the checkbox. Press **Choose a vector from a file** to navigate to a specific vector sequence file or **Choose a folder containing vectors** to specify a folder of vectors. Then press the **Trim now** tool (Trim Now). If vector sequence was found and removed, the vector name will be listed in the table's **Vector** column. The coordinates that will be omitted from the assembly are shown in the **V-trim** column.

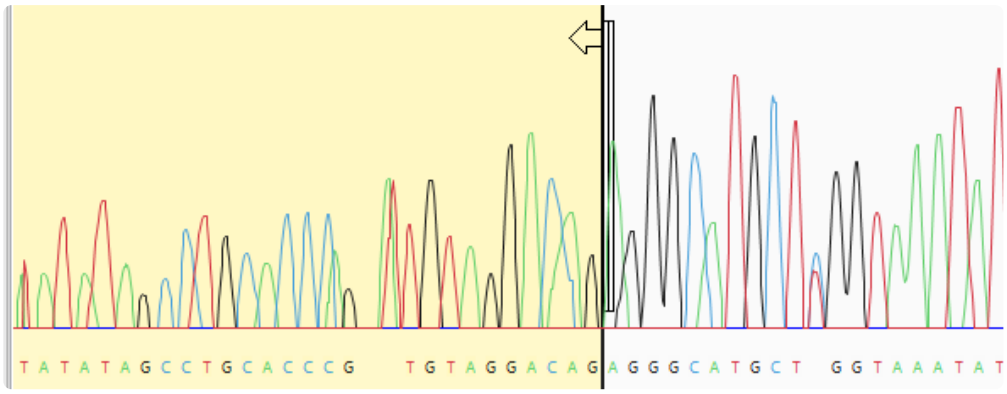
5. (optional) Automatically trim sequence ends based on quality. Start by checking the **Quality Trim** box and press the **Trim now** tool (Trim Now). This method uses SeqMan Ultra's proprietary quality score system to remove low-quality data. The coordinates that will be omitted from the assembly are shown in the **Q-trim** column of the table above. See Quality score calculations and End-trimming based on averaged quality scores for more information.

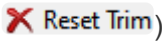
6. (optional) Manually trim sequence ends. Start by double-clicking on any sample row containing trace data. This populates the lower portion of the Unassembled view with a trace data chromatogram.

Copyright DNASTAR, Inc. 2022



Grab the vertical trim sliders on the 5' and/or 3' ends of the sequence and release them in the desired location. The portion of the sequence that will be removed is shown with a yellow background. If you already performed automated quality or vector trimming in the steps above, a trimmed portion may already be visible. You can still use the sliders, if desired, to override the trim location and choose a different one. The coordinates that will be submitted for assembly are shown in the bottom left corner of the Unassembled view, as well as in the **Limits** column of the table above.

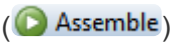


7. (optional) If you change your mind and wish to restore all the trimmed sequence, press the **Reset Trim** tool (

).

8. Specify the assembly stringency by entering a number in the box to the right of **Minimum match percentage** in the header. To the right of the box, the header displays the number used the previous time this set was assembled (if an assembly has already taken place) and the number of sample and reference sequences involved in the assembly. If some samples are not being aligned, try changing the **Minimum match percentage** prior to assembling again.



9. To assemble all the sample sequences in the table, press the **Assemble** tool (

). If a reference is present, the assembler will use the reference(s) as a template. Otherwise, the sequences will be *de novo* assembled.

Copyright DNASTAR, Inc. 2022

The status of the assembly can be checked in the Jobs panel.

Once assembly is complete, results will open in the Project Details view and any contigs will appear in the Explorer panel.

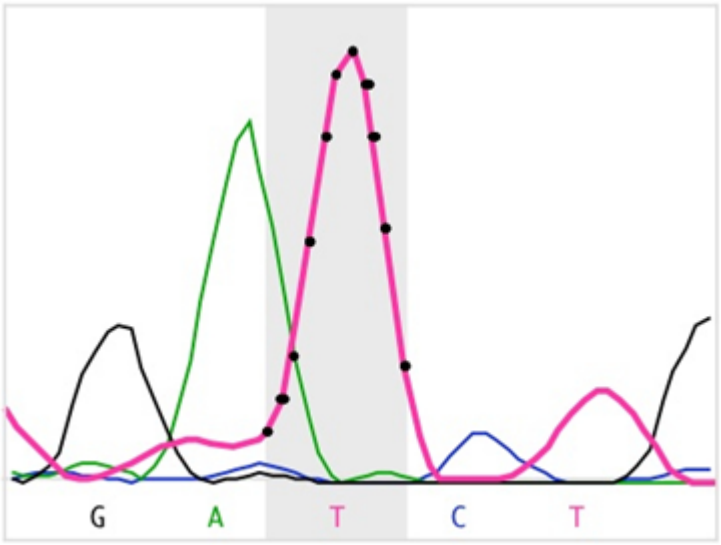
Copyright DNASTAR, Inc. 2022

Quality score calculations

When Sanger/ABI trace data are used in an assembly, SeqMan Ultra analyzes the shape and intensity of peaks to calculate quality scores (Q), and averaged quality scores (Q/n). In quality score calculations:

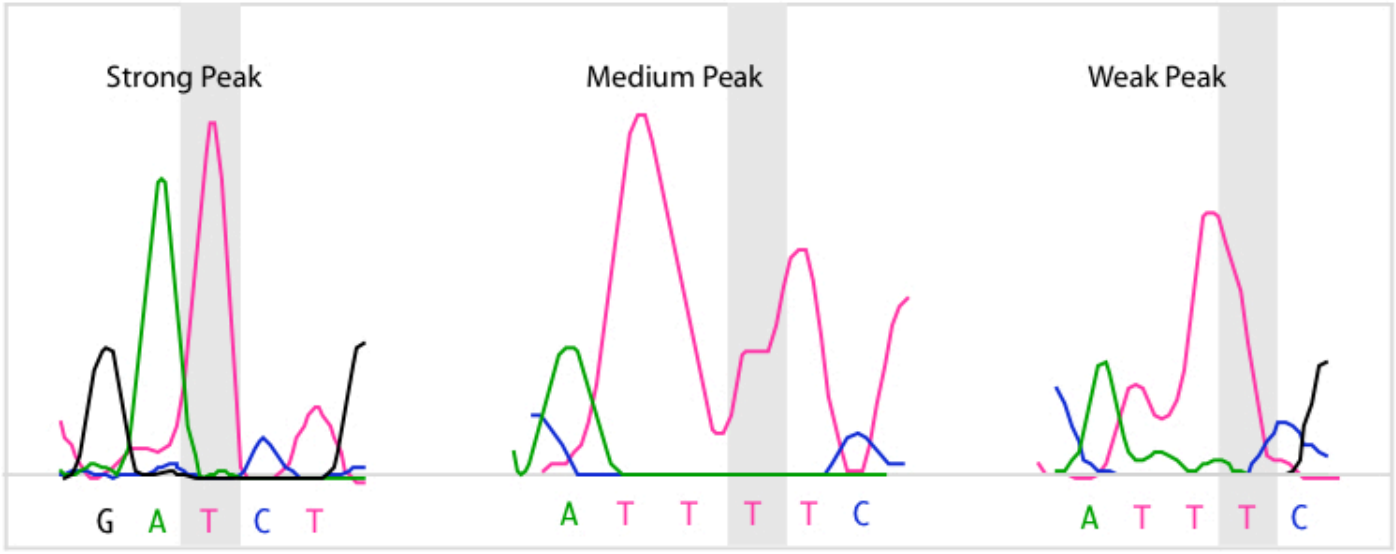
- Taller, sharper peaks receive higher scores than less distinct peaks. The heights of any underlying peaks are subtracted from the highest peak's score during the calculation.
- The further a peak is from the location at which the base was called, the lower the quality score.

The trace data for a DNA sequence comprises four sets of traces—one each for A, C, G, and T. Each trace contains a sequence of intensity values that can be plotted to form a graphical display of trace data. The portions of the four traces associated with a single base call each contain about ten to twelve data points. Only the trace from which the base call is derived is used to calculate a quality score (e.g. if the base call is a T, only the T trace is analyzed to calculate a quality score). The figure below shows data plotted for five base calls. The data points associated with the center base—a T—are marked with black dots.



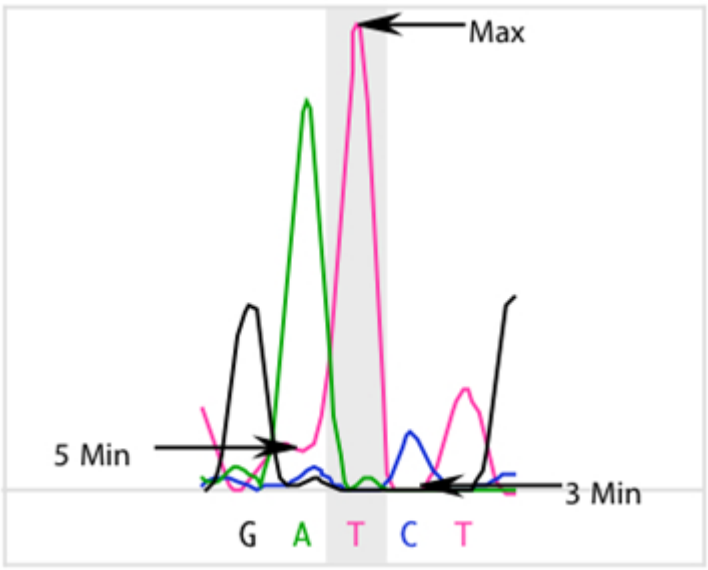
SeqMan Ultra calculates each of the peaks in the trace data. A peak is defined as trace data that exhibits negative curvature. Slope is used to differentiate between three kinds of peaks: strong, medium, and weak. Strong peaks exhibit a change in the sign of the slope, medium peaks contain a shoulder with a slope of zero, and weak peaks have neither a change in sign nor a shoulder. If the trace data for a base call do not contain a peak, its quality score is zero. The figure below contains examples of the three kinds of peaks for the highlighted T base.

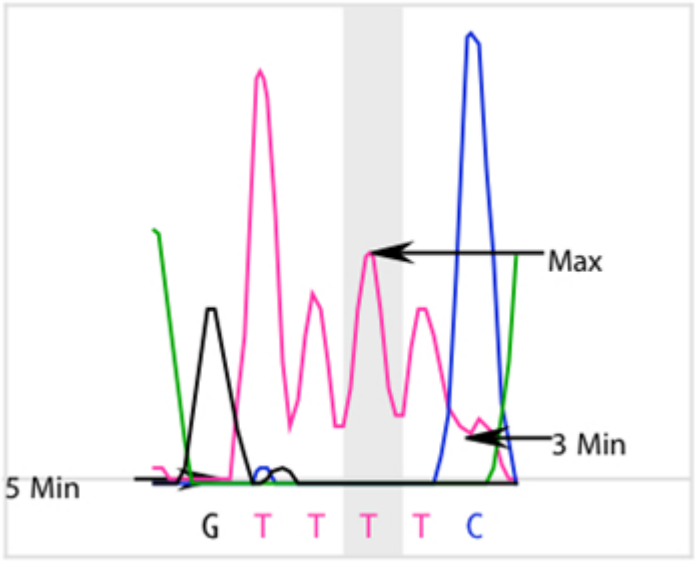
Copyright DNASTAR, Inc. 2022

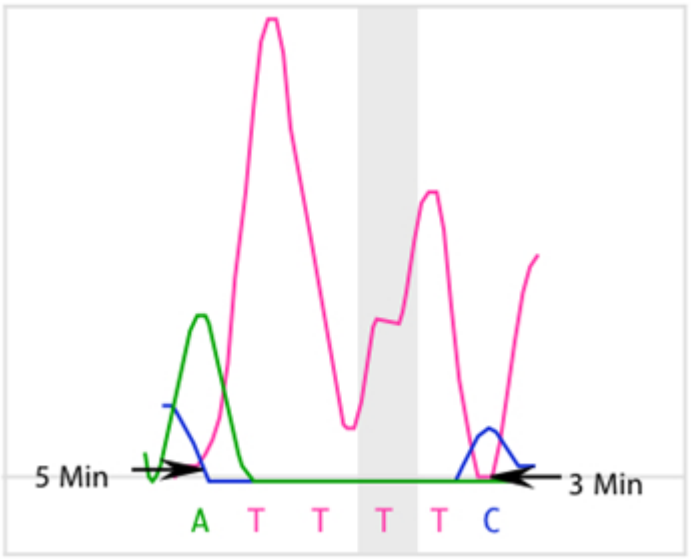


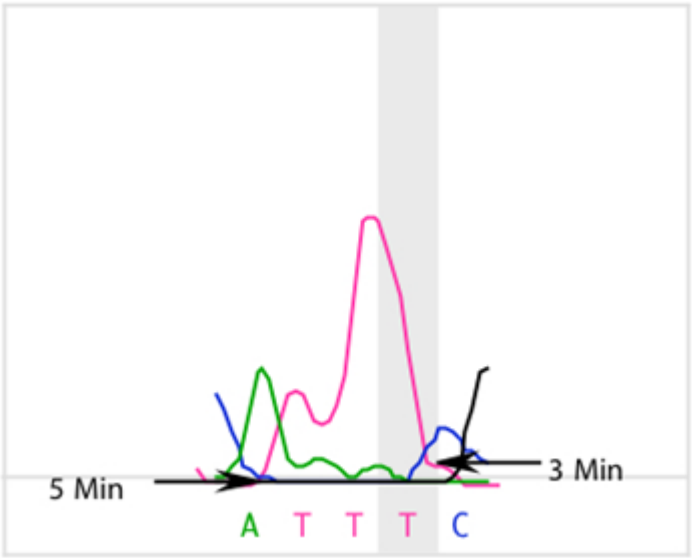
Quality score calculations take into account several parameters:

- Three extreme intensity points: 5Min (5’ minimum), 3Min (3’ minimum), and Max. 5Min and 3Min are the intensity values to either side of the base call that are the minimum values of the data for that base. If a run of identical base calls occurs, then the minimums are taken from either side of the homopolymeric run. Max is the intensity value of the peak.



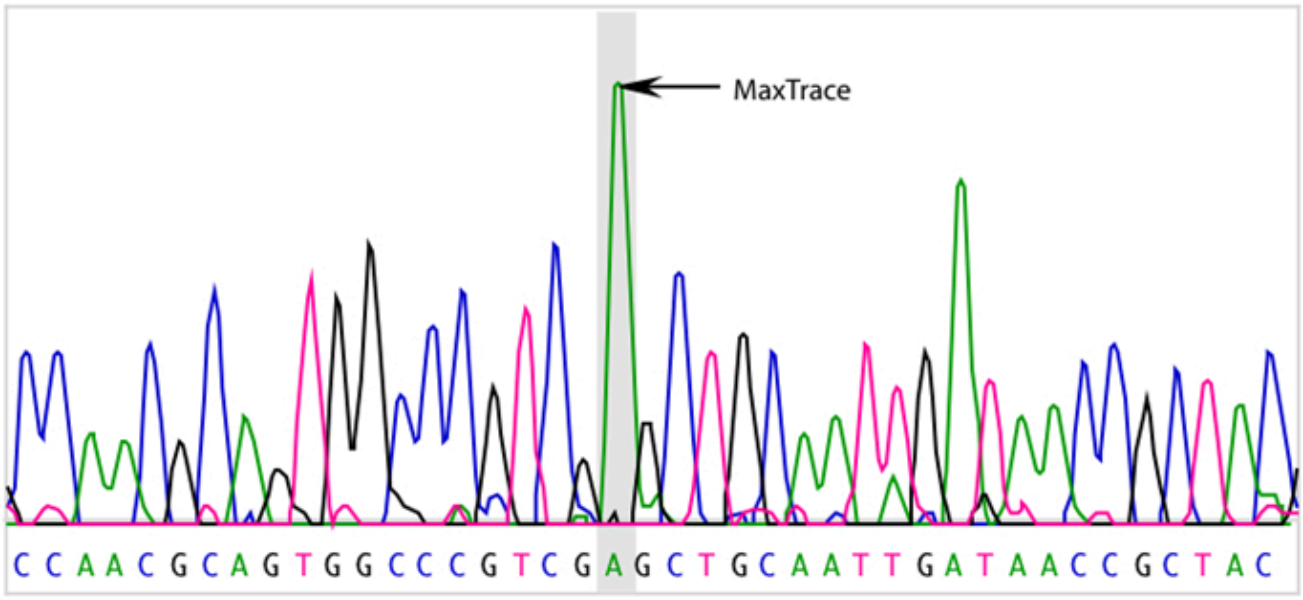




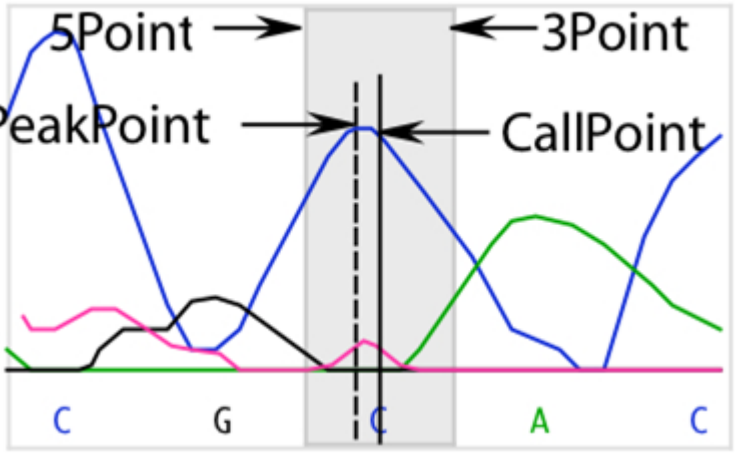


- Each quality score calculation includes division by the maximum intensity of all four traces for an

entire sequence. This assigns higher scores to higher peaks. In this example, an A peak has the highest intensity value. Its intensity value, MaxTrace, is used in the quality score calculation for all bases in the sequence.



- Trace data files identify the point in the trace data where the base was called, or “distance weight.” In high quality data, this usually coincides with the point where SeqMan Ultra detects a peak. In poorer quality data, the peak can be offset significantly. Each quality score is adjusted to reflect the distance from the detected peak to the point where the base was called.



The fraction of the number of points in the offset to the total number of points is the Dist weight used in the quality score calculation. It is calculated as follows:

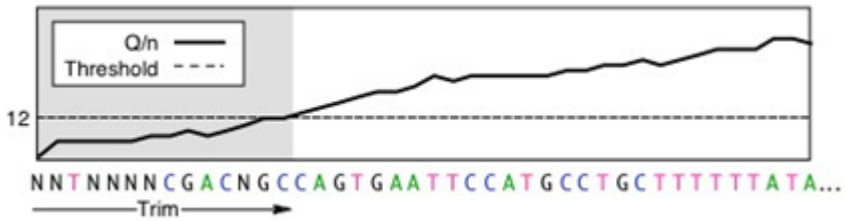
$$Dist = \frac{\text{abs}(PeakPoint - CallPoint) + 1}{(3Point - 5Point) + 1}$$

Copyright DNASTAR, Inc. 2022

End-trimming based on averaged quality scores

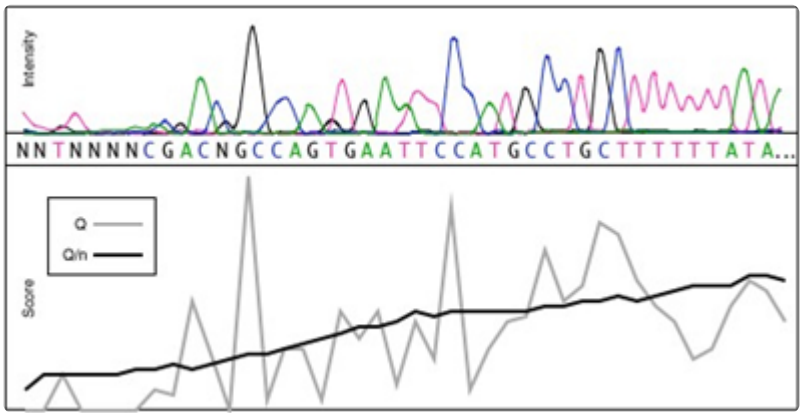
SeqMan Ultra uses averaged quality scores (Q/n) to identify regions of poor quality data at the end of sequences. Averaged quality scores are calculated as the average of the quality scores, Q, over a window of 21 bases. The average score is assigned to the base in the center of the window. Averaging the scores smooths out the quality scores and quantifies the general quality of data in a region. To perform quality end-trimming, a threshold is set and the longest sequence of bases with all Q/n meeting the threshold is identified. Below-threshold ends to either side of the high-quality region of the sequence are trimmed off before assembly.

In the example below, the quality scores, Q, and averaged quality scores, Q/n, are graphed for the 5' end of a 794 base pair sequence. A dashed horizontal line marks the quality end-trimming threshold. The average scores, Q/n, are compared to the threshold and the first 14 bases are trimmed from the 5' end of this sequence.



Poor quality data on the ends of sequences often contain miscalled bases that produce mismatches in alignments with other sequences. If the number of mismatches is high enough that SeqMan Ultra's **Minimum Match Percentage** threshold (set in Assembly parameters) is not met, sequences will not be assembled in the same contig. Trimming the poor quality data from the ends of sequences allows better and more complete assembly.

The image below compares Q to Q/n scores for the 5' end of a sequence.



Assemble multi-sample Sanger data

When you elect to assemble multi-sample Sanger data from within SeqMan Ultra, you will be seamlessly transferred to SeqMan NGen to set up and run the assembly. Once assembly is complete, you can then press a button to launch the results in SeqMan Ultra for editing and analysis.

To assemble multi-sample Sanger data:

Do any of the following:

- Click on the **Welcome** project tab or choose **View > Welcome** to open the Welcome screen. Click on **New Assembly** on the left. On the right, choose **New multi-sample Sanger/ABI assembly**.

- Choose **File > New Multi-Sample Sanger/ABI Assembly**

SeqMan NGen will open at the Workflow screen. Use the links below to learn about each screen you will encounter in the SeqMan NGen wizard:

- Reference Sequence
- Input Sequences
- Set Up Experiments
- Assembly Options
- Analysis Options
- Assembly Output
- Run Assembly Project
- Assembly Summary.

Once assembly has finished, click the **Open assembly** button from the Assembly Summary screen to open the results in SeqMan Ultra for editing and analysis.

Copyright DNASTAR, Inc. 2022

Assemble NGS data or large sets of Sanger data

When you elect to assemble multi-sample Sanger data from within SeqMan Ultra, you will be seamlessly transferred to SeqMan NGen to set up and run the assembly. Once assembly is complete, you can then press a button to launch the results in SeqMan Ultra for editing and analysis.

To assemble NGS data or large sets of Sanger data:

Do any of the following:

- Choose **File > New Assembly with Options** or press **Ctrl/Cmd+Shift+N**.
- Click on the **Welcome** project tab or choose **View > Welcome** to open the Welcome screen. Click on **New Assembly** on the left. On the right, choose **New assembly with options** or any of the **Genomics** or **Transcriptomics** options.

> **Note:** A padlock icon represents a workflow that is not available for your license type. A Lasergene Genomics license allows access to all workflows, while a Lasergene Molecular Biology license only allows you to access Sanger assembly workflows.

SeqMan NGen will open at the Workflow screen. Refer to the SeqMan NGen User Guide for detailed information about using the wizard. Help topics for some of the wizard screens you might encounter are listed below.

- Reference Sequence
- Input Sequences
- Assembly Options
- Analysis Options
- Assembly Output
- Run Assembly Project
- Assembly Summary.

Once assembly has finished, click the **Open assembly** button from the Assembly Summary screen to open the results in SeqMan Ultra for editing and analysis.

Copyright DNASTAR, Inc. 2022

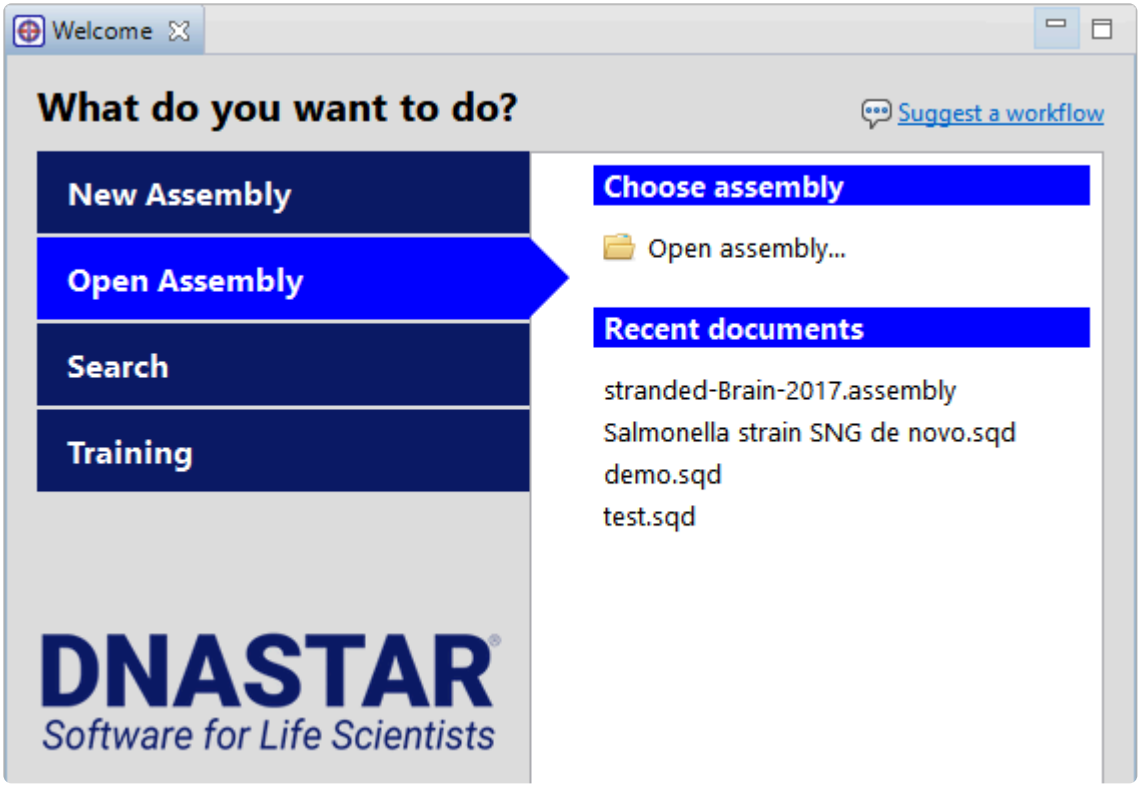
Open an Existing Assembly

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.

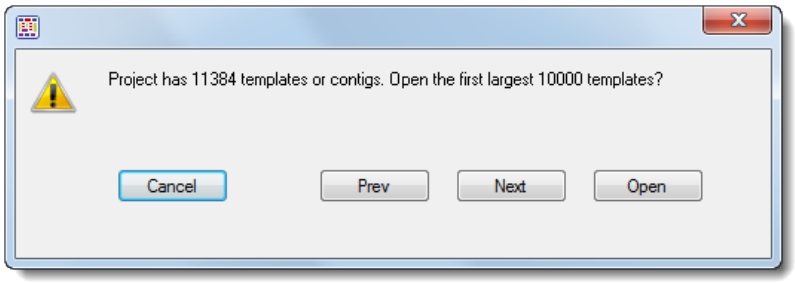


- With any window active, use the **File > Open** command.
- Drag & drop an assembly file/package onto the SeqMan Ultra window.
- Double-click on an *.sqd* or *.ace* file. Macintosh also supports double-clicking on an *.assembly* package.

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**

Copyright DNASTAR, Inc. 2022

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.

Copyright DNASTAR, Inc. 2022

Navigate between open assemblies

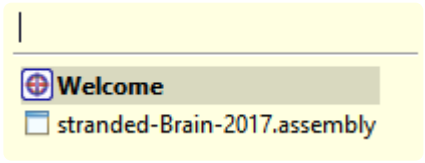
One instance of SeqMan Ultra may include multiple open projects. Each window is displayed in a separate tab.

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.

Copyright DNASTAR, Inc. 2022

Contigs

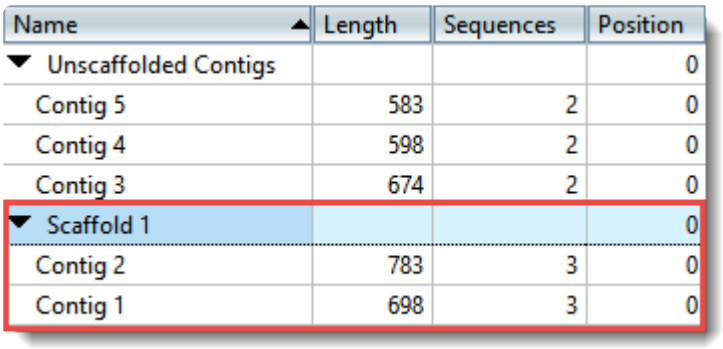
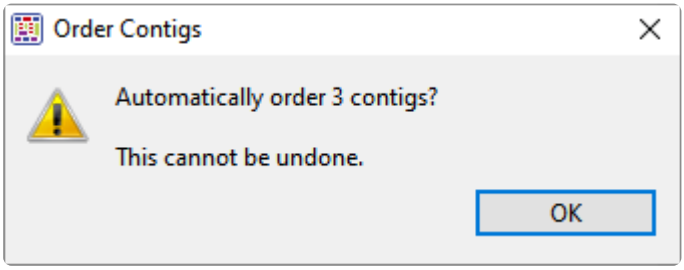
The following topics contain information on viewing, editing and searching within contigs.

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

Copyright DNASTAR, Inc. 2022

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.

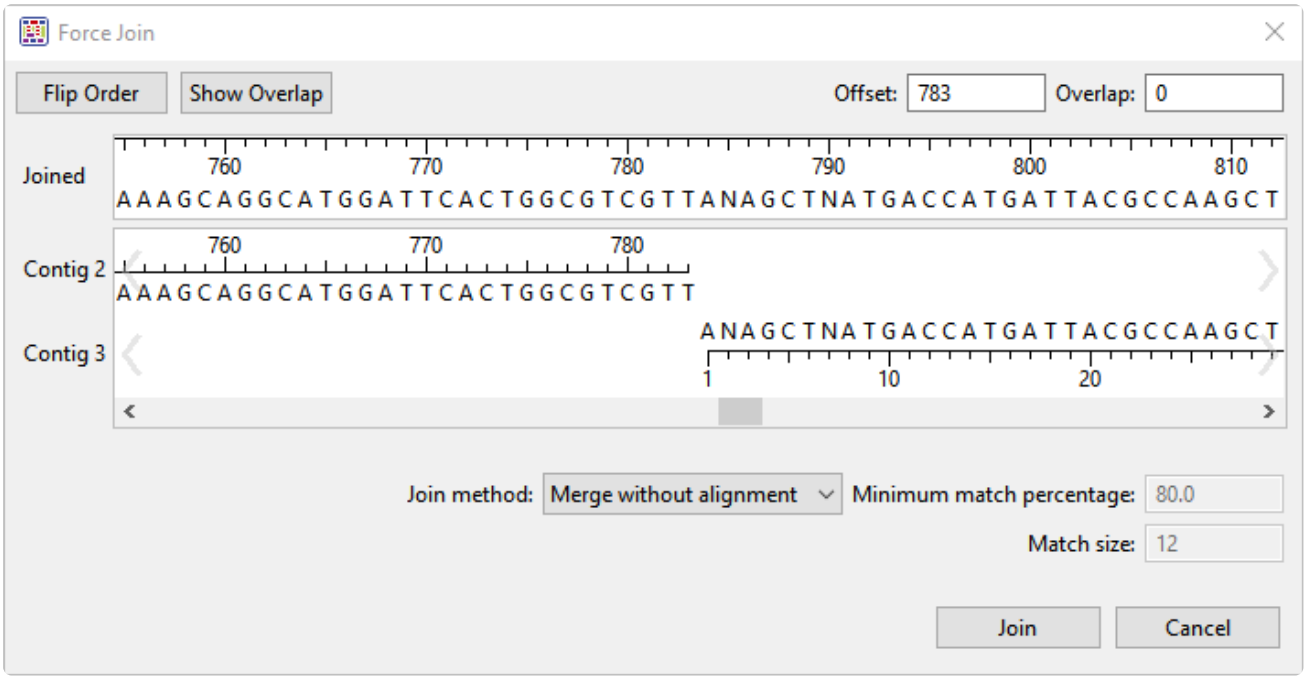
Task	How To
Delete a contig	Select the contig in the Explorer panel, then use **Contig > Delete Contig**.
Place contigs on a scaffold	A “contig scaffold” is a collection of contigs that are related based on their known spatial relationships rather than their sequence similarity. SeqMan NGen uses paired 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 **Shift+click** or **Ctrl/Cmd+click** to select the contigs you wish to place together on a scaffold. Next, select **Contig > New Scaffold** or right-click on a contig and choose **New Scaffold**. The contigs are now placed together in a scaffold at the bottom of the table.  Once at least one scaffold has been created, all other contigs are placed in the **Unscaffolded Contigs** scaffold. All positions in the **Unscaffolded Contigs** scaffold are initially set to 0. Note that the order of contigs in the **Unscaffolded Contigs** scaffold does not imply that adjacent contigs should be joined.
Order contigs in a scaffold automatically	In the Explorer panel, select the contigs using **Shift+click** or **Ctrl/Cmd+click**. Contigs must already be in the same scaffold. Next, choose **Contig > Order Contigs** or right-click on the selection and choose **Order Contigs**. A warning appears. 

Copyright DNASTAR, Inc. 2022

	To continue with contig ordering, press **OK**. SeqMan Ultra will evaluate or reevaluate all contigs, including those already in scaffold.Note that changing the **Position** of a contig can influence whether pairs in different contigs are consistent.
Order contigs in a scaffold manually	Contigs in a scaffold can be manually ordered to reflect known spatial arrangements between pairs. In the Explorer panel, edit the **Position** cell for each contig in the scaffold by clicking on it and typing in an integer representing the order in which the contig should be placed. To view the contigs in order, click once or twice on the **Position** header.Note that changing the **Position** of a contig can influence whether pairs in different contigs are consistent.
Reverse complement contigs	Select the contig(s) in the Explorer panel, then choose **Contig > Reverse Complement (Contig Name)**.
Split a contig	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 Alignment view, then place the cursor in the consensus at the position where you wish to split the contig. Finally, select **Contig > Split at Insertion**. 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.
Place unlocated contigs based on paired-end information	Select the contigs in the Explorer panel, then use **Contig > Place Unlocated Contigs**. 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 **Contigs > Order contigs** command, SeqMan Ultra will only attempt to place contigs from the unscaffolded pool.
Join two or more adjacent contigs	Select the contigs in the Explorer panel, then use **Contig > Align contigs end-to-end**. 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.
Replace a portion of the consensus sequence with a portion of a constituent sequence	In the Alignment view, highlight the region of the constituent sequence you wish to use. Select **Contig > Force Consensus**. 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 **Contig > Compute Consensus**. 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.
Merge two contigs	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

Copyright DNASTAR, Inc. 2022

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.

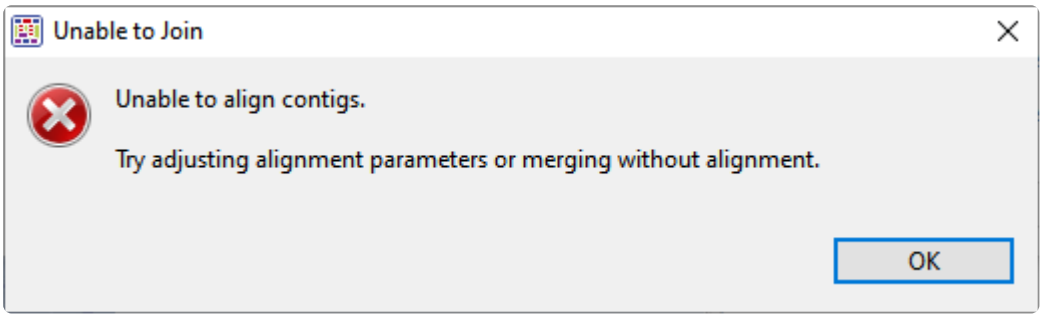
4. (optional) To center the overlapping area for a better view, press **Show Overlap**.

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."

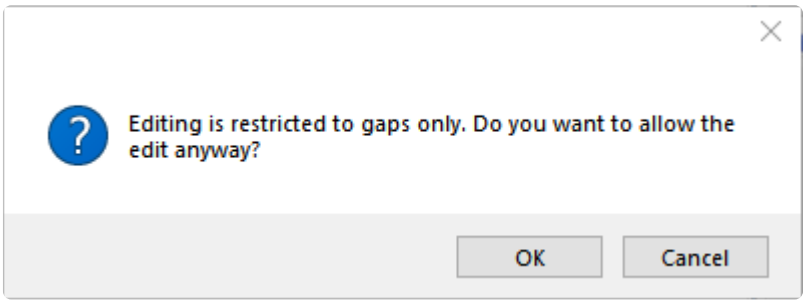
Copyright DNASTAR, Inc. 2022

	
Restore the trimmed end(s) of a contig automatically	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 **Contig > Extend Contig Ends > (Subcommand)**. Choose between: **Extend 3' Sequences ends**, **Extend 3' and 5' Sequences ends** and **Extend and Align 3' and 5' Sequences ends**. 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.
Restore the trimmed end(s) of a contig manually	Select a contig in the Explorer panel and open it in the Alignment view. Scroll to the 3' or 5' end of the sequence. Click the plus sign next to the sequence name to show the **Traces** data track. 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.

Copyright DNASTAR, Inc. 2022

Edit a contig

In the Alignment view, you may edit constituent or consensus sequences in an *.sqd* project contig by removing or adding nucleotides or gaps.

Task	How To
Remove or add gaps	Insert the cursor into an individual sequence and type one or more dashes (-) or delete the same. The consensus sequence will be recalculated to reflect your changes.
Remove or add nucleotides using the keyboard	To remove nucleotides, insert your cursor in the desired location and use the **Backspace** or **Delete** keys. To add nucleotides, insert your cursor and begin typing. When you add new nucleotides, they will appear in lower case to signify a manual addition. If you edit one or more constituent sequences, the consensus sequence will be recalculated to reflect your changes. If you add nucleotides to or delete them from the consensus and then restore a trimmed constituent sequence in the same region, the restored sequence will not reflect the insertion/deletion. If you want to ensure that all sequences reflect the changes to the consensus sequence, you should restore trimmed data (see row below) before editing the consensus sequence. Note: If you attempt to add or remove nucleotides using this method, you may receive the following warning:  If you still wish to proceed, click **OK**.
Restore trimmed sequence or trim it even more	Trimmed constituent sequences appear in the view with a small triangle at one or both ends of the sequence. Drag the triangle to either side to reveal previously-trimmed sequence or to trim the sequence even more. If trace data is available for the sequence, you can view this by clicking the plus sign (⊕) to the left of the sequence name. Having the trace data displayed can assist you in determining where the trim point should be.
Cut or copy	Select some or all of the sequence, then right-click and choose **Cut** or **Copy**.

Copyright DNASTAR, Inc. 2022

a selected portion of the sequence	
Slide a Sanger sequence relative to other sequences	Use the **Banana** tool (). in the Alignment view header. When this tool is activated, the cursor is shaped like a banana. Use your mouse and the banana-shaped cursor to drag any sequence in the Alignment view to the left or right, as desired. When you are done, click the **Banana** tool again to return the cursor to its normal state. Note that this tool is only active when the alignment contains Sanger trace files.
Remove a selected sequence from the contig	Select a sequence name at the left of the Alignment view, then choose **Sequence > Remove Sequence**.

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.

To reassemble one or more contigs after editing them, see Reassemble contigs.

Copyright DNASTAR, Inc. 2022

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.

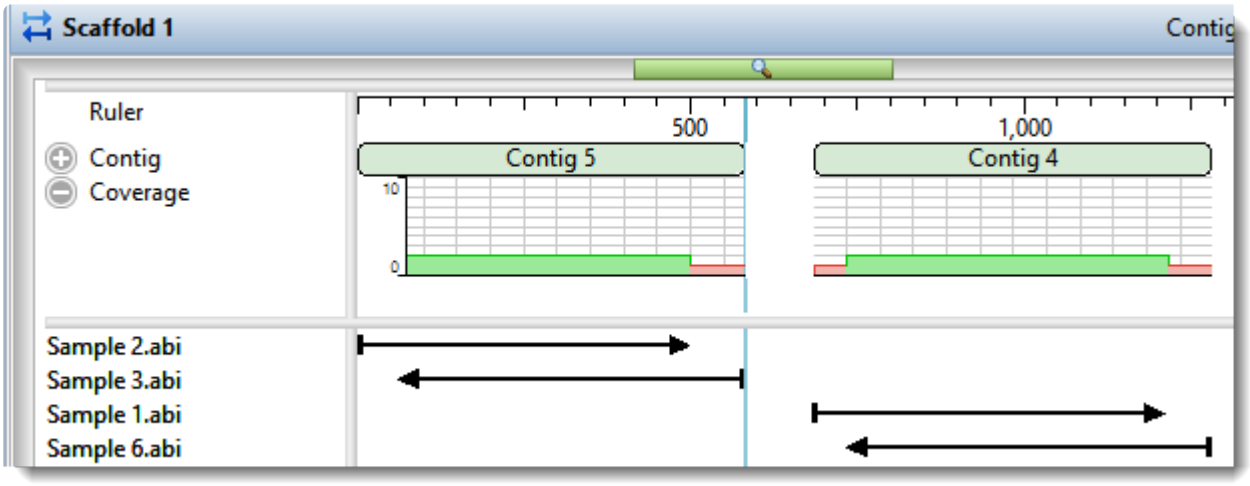
Before you begin, note that this method works only for *.sqd* files (not *.assembly* files) that contain mate-pair data.

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.

Copyright DNASTAR, Inc. 2022

- 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.

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



4. 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.

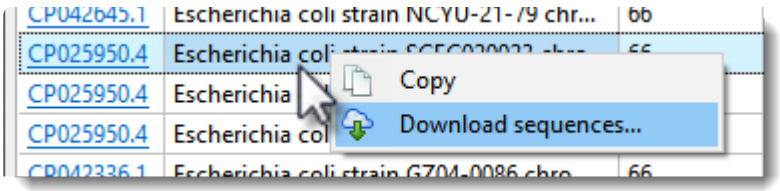


5. When you find a gap you would like to close, use your mouse to select a range of sequence near the end of either contig.

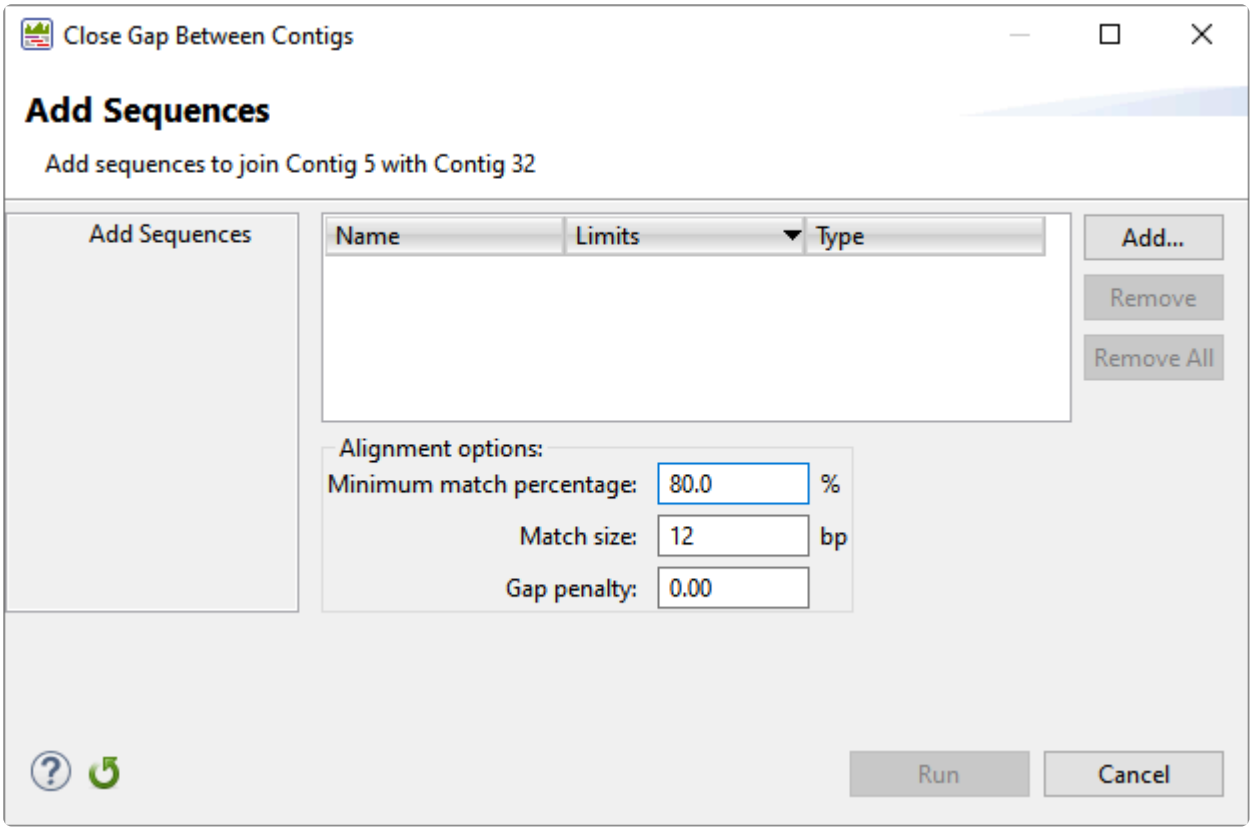
6. Choose **Search > Search** to initiate a BLAST search of that contig end. Choose the desired settings and press **Run Now** or **Run**.

7. Once the job finishes, click the link in the **Status** column of the Jobs panel to open the search results.

8. Click on the best match, then right-click and choose **Download Sequences**.



9. In the Explorer panel, select the two contigs on either side of the gap and choose **Contig > Add Sequences to Close Gap**.



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.

Copyright DNASTAR, Inc. 2022

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.

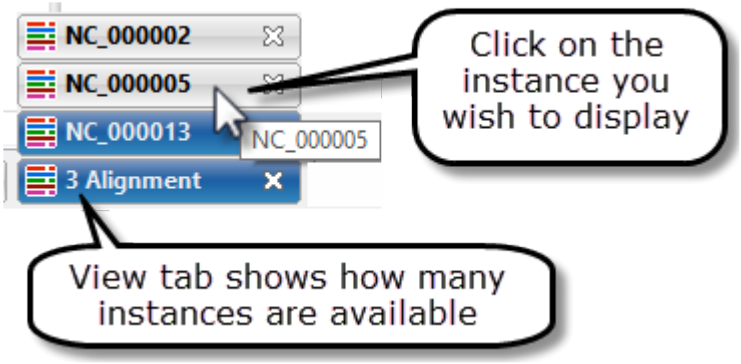
These views can be accessed using commands in the **View** menu, but the fastest way to access them is to use the toolbar to the right of the Explorer panel. Available tools vary depending on the type of assembly.

> **Note:** Once two views are open, subsequent views will open in separate moveable and resizable windows. To reattach the views, see Customize the look and layout.

Image	Tool Name	Shortcut to Menu Command	Description
	**Show alignment view of selected contigs**	**View > Alignment**	Causes one Alignment view to open for each selected contig.
	**Show strategy view of selected contigs or scaffolds**	**View > Strategy**	Causes one Strategy view to open for each selected contig or scaffold.
	**Show table of features in selected contigs**	**View > Features**	Causes one Features View 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 Variant View to open for each selected contig.
	**Show report view of selected contigs**	**View > Contig Report**	Causes one Contig Report View to open for each selected contig. This view comes in three versions: Contig Information, Contig Coverage, Contig Structural Variation. By default, the Contig Information 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.

Copyright DNASTAR, Inc. 2022



Copyright DNASTAR, Inc. 2022

Rename a contig

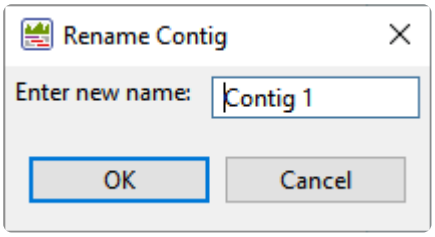
Contigs can only be renamed in *.sqd* projects. To rename a contig:

1. Do either of the following:

    - In the Explorer panel, click on the contig row you wish to rename. Then choose **Edit > Rename Contig**.

    - In the Explorer panel, right-click on the contig row and choose **Rename**.

2. Overwrite the existing name with the new one and press **OK**.



Copyright DNASTAR, Inc. 2022

Reassemble contigs

After you have performed a Sanger assembly and edited contigs —including removing sequences from them—you can elect to reassemble them. This reassembly respects any edits you have made to the contigs, but will update the alignment and gap positions.

To reassemble one or more contigs:

1. Perform the initial assembly and make any desired edits.
2. From the Explorer panel table, select the contig(s) you want to reassemble.
3. Choose one of the following commands.
    - **Contig > Reassemble contig 'n' individually** (where 'n' is the contig number) or **Reassemble selected contigs individually** are the most commonly-used reassembly commands. These commands rearrange sequences within their original contigs and can be used to fix a possible misalignment or false join in specified contig(s). These commands reassemble sequences only within their original contig. When you use these commands, sequences are not moved outside their original contig and contigs are not allowed to merge.
    - **Contig > Reassemble contig 'n' within the project** or **Contig > Reassemble selected contigs within the project** are less commonly used than the commands described above. These commands start by disassembling the reads within the selected contig(s). During the reassembly stage, the disassembled reads may be joined back together or may be incorporated into any of the other currently-selected contigs. These commands can be useful if you believe that sequences have been placed in the wrong contig, possibly due to an issue with repeats or with manual end-trimming. When you use these commands, contigs may merge, new contigs may be created, and/or sequences may be redistributed between the selected contigs. This is the type of reassembly done within SeqMan Ultra's predecessor, SeqMan Pro.

Notes:

* Reassembling a contig after using the **Banana** tool () in the Alignment view can result in unwanted mismatches.

* Reassembly acts only on the sequence data in the selected contig(s). It does not add any new data in the Unassembled view to the contig(s).

* If you want to perform the alignment using different parameters, we instead recommend starting a new SeqMan Ultra assembly from scratch.

Copyright DNASTAR, Inc. 2022

Features

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

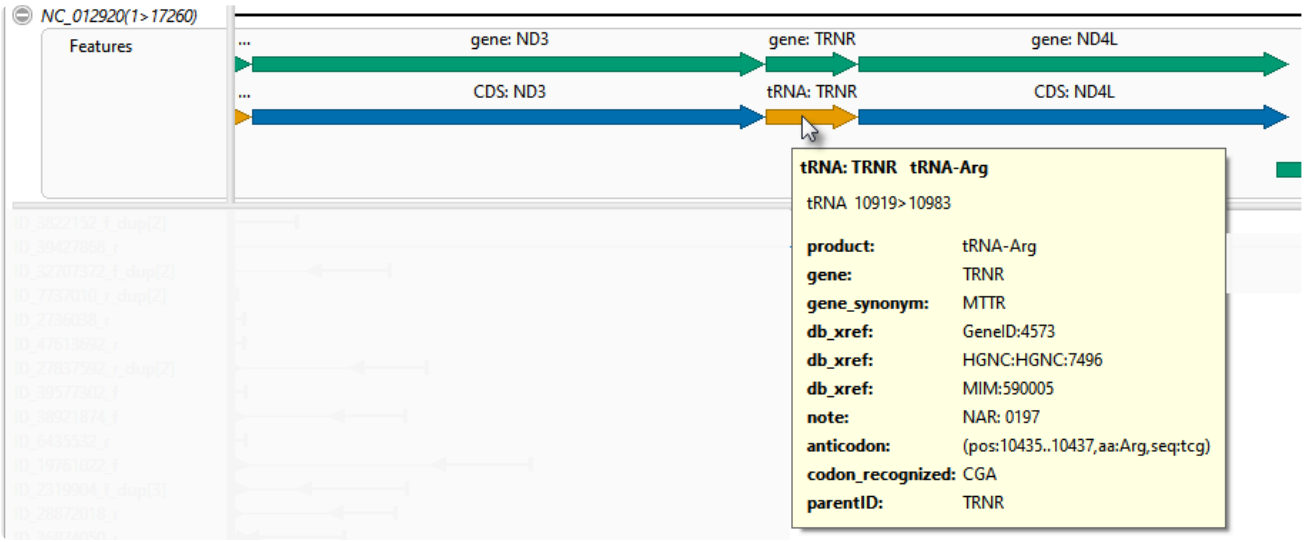
View a table with detailed feature information:

See Features view.

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.



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**.

Copyright DNASTAR, Inc. 2022

Tracks

“Tracks” are rows of information that can be displayed or hidden, as desired, in the Strategy and Alignment views. Tracks can pertain to a header, footer, or individual sequence row.

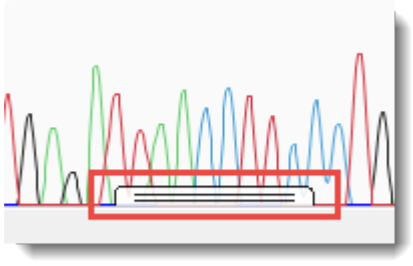
The following tracks can be applied, removed or (in most cases) edited:

- Ruler
- Contig
- Coverage
- Pair consistency
- Features
- Split reads
- Translation
- Traces
- Majority
- Sample-Alignment

The following table shows tasks that affect the display of tracks within SeqMan Ultra’s graphical views:

Task	How to…
Create tracks	Track data can be imported automatically with some types of sequence data (e.g., Features tracks), or can be calculated by the application itself (e.g., Ruler tracks).
Display tracks in the views	To display any type of track, you must check the box next to the name of the track in the Tracks panel. This is the only condition needed to view a Ruler track. For any type of track other than a Ruler, you must also expand the track as described below.
Expand/collapse tracks for selected sequences	Do any of the following: • Click the plus/minus sign to the left of the sequence name (/). • To expand tracks for selected sequences, choose **Sequence > Expand Sequence Rows** or **Overview > Expand Overview Rows**; or right-click on the selection and choose **Expand**. • To collapse tracks for selected sequences, choose **Sequence > Collapse Sequence Rows** or **Overview > Collapse Overview Rows**; or right-click on the selection and choose **Collapse**.
To expand/ collapse tracks for all sequences in a particular view	Ensure that no sequences are selected. • To expand tracks for all sequences, choose **Sequence > Expand All Rows** or **Overview > Expand All Rows**; or right-click in the view and choose **Expand**

Copyright DNASTAR, Inc. 2022

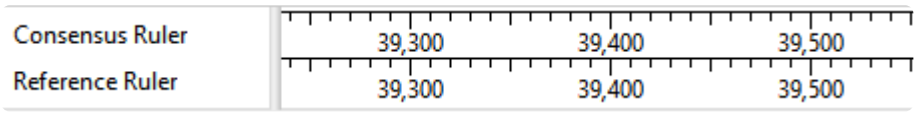
	All. • To collapse tracks for all sequences, choose **Sequence > Collapse All Rows** or **Overview > Collapse All Rows**; or right-click in the view and choose **Collapse All**.
Hide tracks	Do one of the following: • To hide a specific track from all views, uncheck the box next to the name of the track in the Tracks panel. • To hide a specific track in one view only, right-click on the track within that view and choose **Hide Track**. Alternatively, select a track and use **Edit > Delete** or press the **Delete** key. • To hide all tracks in one view only, click the minus sign ( ) to the left of the sequence name in that view, or use **Analysis > Collapse Sequence Rows** or **Overview > Collapse Sequence Rows**. • To hide all tracks in all views, do a combination of the above.
Access the options for a particular type of track	Do either of the following: • Right-click on a track in either view and choose **Show Track Options**. • Click **View > Tracks > Options** or open the Tracks panel and click the Options bar (Options), then do either of the following to specify the track of interest by clicking on its name in the Tracks panel.
Change the order of all tracks in a view	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.
Increase/decrease the vertical height of a visible track	Hover the mouse over the bottom of the track in the Overview or Sequences view, then grab the resize bar and drag it up or down. 
Rename a track	See Rename a track manually.

manually	
Rename tracks automatically using specified data fields	See Rename tracks automatically using specified data fields.

Copyright DNASTAR, Inc. 2022

Ruler track

Consensus and Reference ruler tracks can be displayed, if desired, in the Strategy and/or Alignment view headers.



The visibility of rulers is dependent on:

- Whether or not tracks are visible in a given view. To display tracks, click on the plus sign next to a sequence in the view of interest.
- Which boxes you check in the Tracks panel. There are four ruler boxes that can be checked and unchecked individually.
    - **Strategy header > Consensus Ruler** – A gap-aware sequence ruler.
    - **Strategy header > Reference > Reference Ruler**
    - **Alignment header > Consensus: Trace > Consensus Ruler**
    - **Alignment header > Reference > Reference Ruler**.

Note: Checking a **Reference Ruler** box has no effect if the project is a *de novo* assembly.

Note that templated assemblies have two coordinate systems: ungapped for the sequences and gapped for the consensus. SeqMan Ultra's "Reference ruler" track display the ungapped coordinate system. By contrast, the "Consensus ruler" tracks number and tick any gaps found in the consensus.

To change Ruler options:

Currently, the Ruler tracks do not have editable options. However, you can add white space above and/or below the ruler using sliders in the Layout section.

Copyright DNASTAR, Inc. 2022

Contig track

The Contig track can be displayed in the Strategy view header, if desired. To display the Contig track, you must check its box in the Tracks panel and also click the associated plus sign (if any) in the Strategy view header.

Applying this track only has an effect if you are displaying the Strategy view for a contig scaffold. If a scaffold Strategy view is open, the track shows bars denoting the placement and name of each contig.



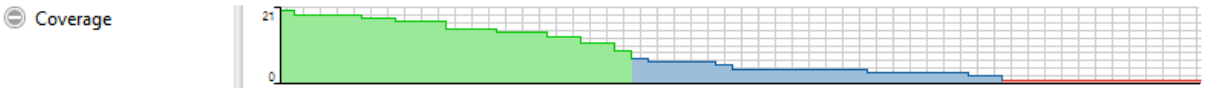
To change Contig track options:

Currently, the Contig track does not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

Copyright DNASTAR, Inc. 2022

Coverage track

The Coverage track can be displayed in the Strategy view header, if desired. To display the Coverage track, you must check its box in the Tracks panel and also click the associated plus sign (if any) in the Strategy view header.



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.

Based on the Strategy Viewing and Coverage parameters, the histogram will appear:

- thick green where coverage meets both the **Coverage threshold** and **Minimum number on each strand** criteria
- thin green where both strands are covered but **Coverage threshold** is not met
- thin blue where coverage is on one strand only
- red where coverage is by a single read only

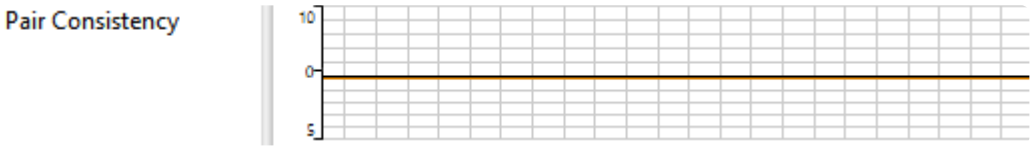
To change Coverage track options:

Currently, the Coverage track does not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

Copyright DNASTAR, Inc. 2022

Pair consistency track

The Pair Consistency track can be displayed in the Strategy view header, if desired. To display the Pair Consistency track, you must check its box in the Tracks panel and also click the associated plus sign (if any) in the Strategy view header.



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. The deeper the Pair Consistency histogram , the more recognized pair of reads there are.

- Green above the baseline indicates the presence of pairs that are consistent with the current assembly.
- Red below the baseline indicates the presence of pairs that are inconsistent with the current assembly due to relative orientation, distance apart, or both.
- 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.
- Pink below the baseline represents split reads in the same contig whose locations or orientation are inconsistent with pair specifier parameters.

Note that SeqMan Ultra cannot distinguish whether inconsistent pairs are due to naming errors or assembly errors.

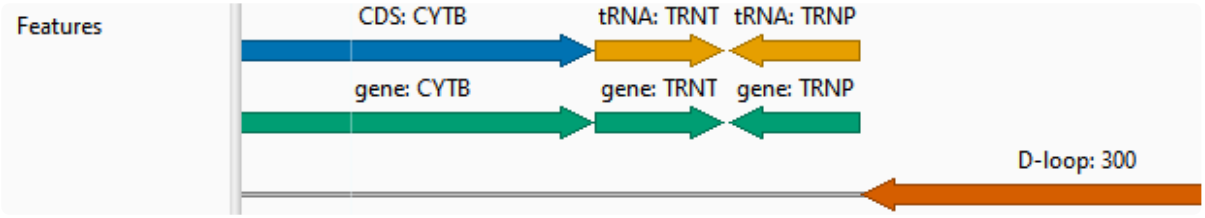
To change Pair Consistency track options:

Currently, the Pair Consistency track does not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

Copyright DNASTAR, Inc. 2022

Features track

Reference and/or alignment read features can be displayed as Feature tracks in the Strategy and Alignment views.



To display the desired Features track, you must check its box in the Tracks panel and also click the associated plus sign (if any) in the area of the view where the track is located. The Tracks panel lets you display features in three locations.

- **Strategy header > Reference > Features** – To display reference sequence annotations.
- **Alignment header > Reference > Features** – To display reference sequence annotations.
- **Alignment reads > Features** – To display annotations from the individual read files.

The following color scheme is used for feature display in SeqMan Ultra, MegAlign Pro and GenVision Pro.

Feature Type	Color
gene	green
CDS	blue
RNA	orange
repeat	yellow
transcription	red-purple
all others	orange-red

To change Features track options:

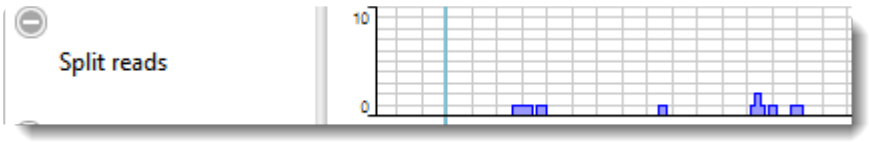
Currently, the Features tracks do not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

Copyright DNASTAR, Inc. 2022

Split reads track

The Split reads track shows a graph in the Strategy view header of all the reads that are split in two. These reads correspond with the **Split read** arrows (controlled via the Strategy section of the Style panel), and can be used to identify structural variations in genomic and exome assemblies and exon usage. This track can only be displayed for *.assemby* projects.

To display the Split reads track, you must check the **Strategy header > Split reads** box in the Tracks panel and also click the associated plus sign (if any) in the Strategy view header.



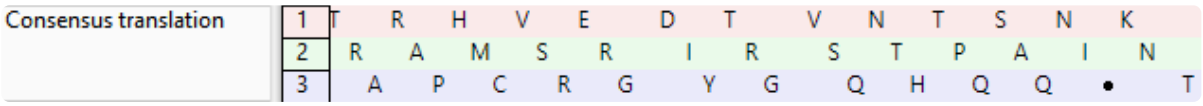
To change Split reads track options:

Currently, the Split reads track does not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

Copyright DNASTAR, Inc. 2022

Translation track

The Translation track can be displayed in the Alignment view header, if desired. To display the track, you must check the **Alignment header > Consensus translation** box in the Tracks panel and also click the associated plus sign (if any) in the Alignment view header.

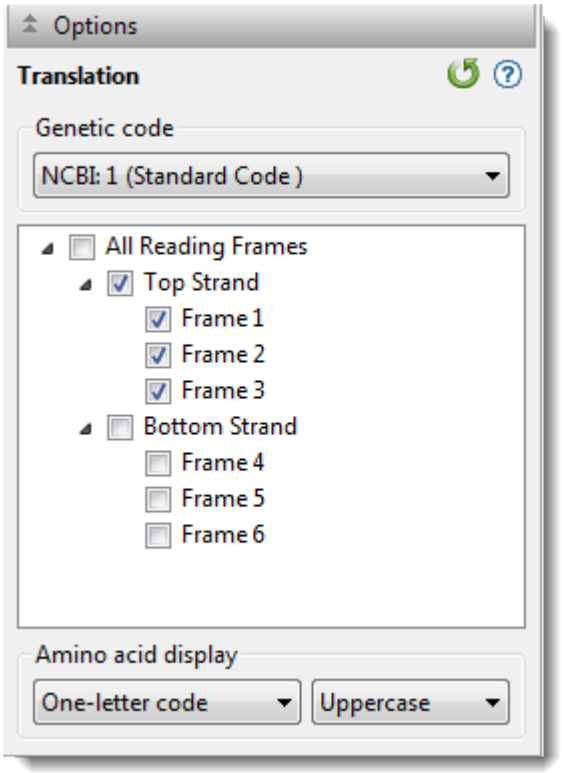


Consensus translation track conventions are as follows:

- Stop codons are represented by dots.
- Translations are not displayed for incomplete codons.
- Translation frames 4-6 are counted from the right end (the 3' end of the bottom strand).
- For purposes of translation, alignments are assumed to be linear.

To change Translation options:

To learn how to access the options section for this track, see Tracks.



- Use the upper drop-down menu to choose from a variety of NCBI and Lasergene translation codes.

Copyright DNASTAR, Inc. 2022

- 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.

Copyright DNASTAR, Inc. 2022

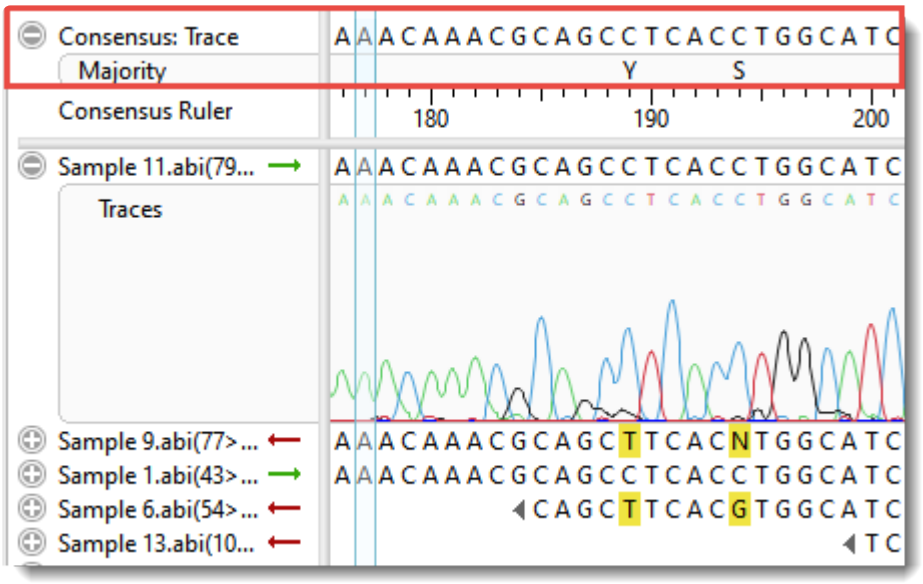
Trace track

The Trace track can be shown in two different views and shows the Trace Evidence consensus call, which is based on the trace data graph for all samples combined. This track is only available if the project is an *.sqd* assembly with Sanger/ABI data.

To display the track in the Alignment view header:

Check the **Alignment header > Consensus: Trace > Trace** box in the Tracks panel and also click the associated plus sign (if any) in the Alignment view header.

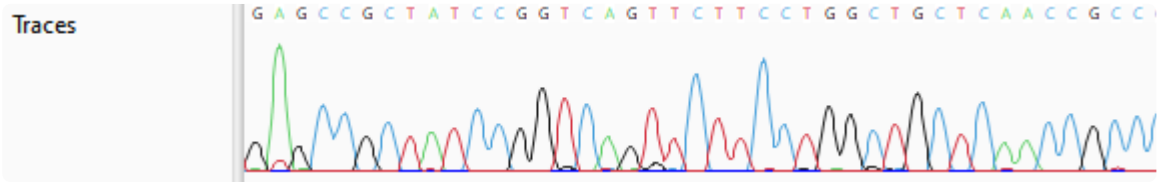
If the current Consensus parameters **Call by** drop-down menu is set to **Trace evidence**, the Trace evidence track will appear above the Majority track, if that track is also shown. The Trace track will contain all calls, while the Majority track shows only disagreements with the Trace track.



Currently, the Trace track does not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

To display the track in the lower portion of the Unassembled view :

Use **Alignment reads > Traces**. You can also use the **Quality scores** tool ( ) to display quality scores if desired.



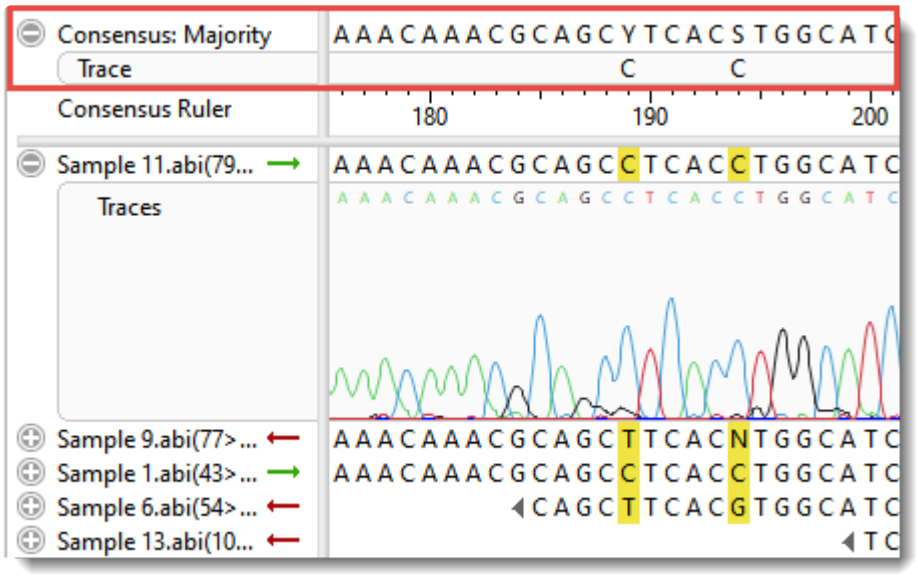
Copyright DNASTAR, Inc. 2022

Majority track

The Majority track can be displayed in the Alignment view header and shows the majority consensus call. This track is only available if the project is an *.sqd* assembly with Sanger/ABI data.

To display the track, check the **Alignment header > Consensus: Trace > Majority** box in the Tracks panel and also click the associated plus sign (if any) in the Alignment view header.

If the current Consensus parameters **Call by** drop-down menu is set to **Majority**, the Majority track will appear above the Trace (evidence) track, if that track is also shown. The Majority track will contain all calls, while the Trace track shows only disagreements with the Majority track.



To change Majority track options:

Currently, the Majority track does not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

Sample-alignment track

The Sample-alignment track can be displayed in the Alignment view and displays sample alignment under a consensus for a multi-sample assembly. To display the track, you must check the **Alignment reads >Sample-alignment** box in the Tracks panel and also click the associated plus sign (if any) in the Alignment view. Note that the track will not appear unless the project involves a multi-sample assembly.

To change Sample-alignment track options:

Currently, the Sample-alignment track does not have editable options. However, you can add white space above and/or below the track using sliders in the Layout section.

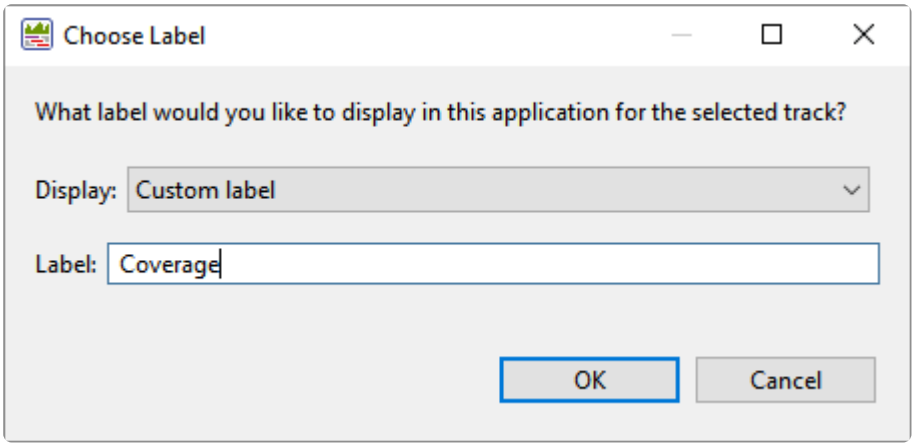
Copyright DNASTAR, Inc. 2022

Rename a track manually

To manually rename a selected track:

1. Select a single track name from any view, then right-click on it and choose **Rename Track**. Alternatively, select the track name and choose **Edit > Rename Track**.

   The Choose Label dialog opens with the **Display** drop-down menu set to **Custom label**.



}!

2. Type the desired name into the **Label** text box.

3. Click **OK** to save changes and exit from the dialog, or **Cancel** to leave the dialog without saving any changes.

Copyright DNASTAR, Inc. 2022

Rename tracks automatically using specified data fields

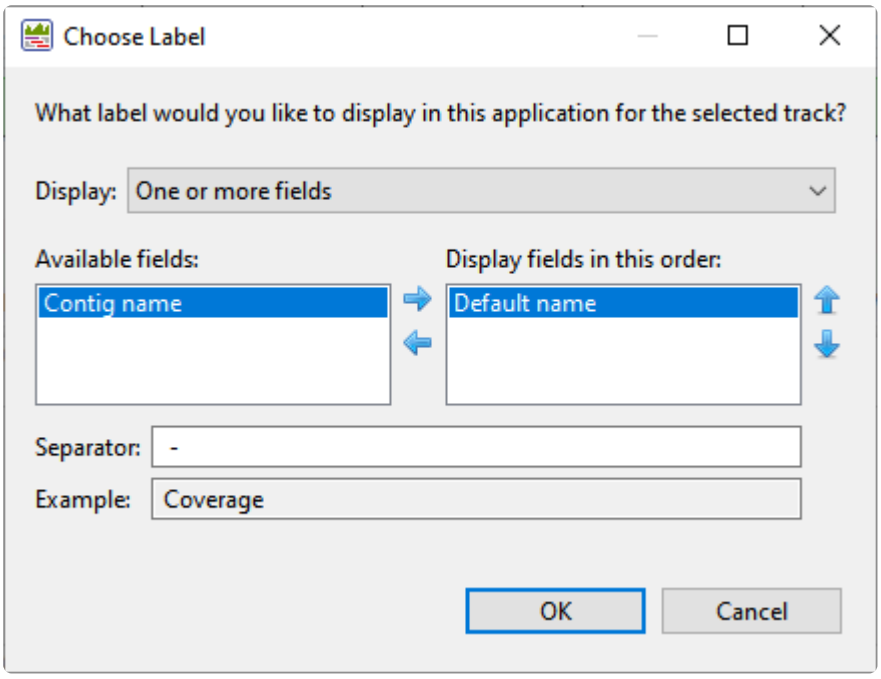
The following procedure lets you create labels for one or more selected tracks automatically, using specified fields such as the contig name or default track name.

To rename tracks using specified data fields:

1. Select one or more track names from any view.

2. Right-click on the selection and choose **Rename Track with Fields**, or instead choose **Edit > Rename Track with Fields** from the main menu. 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**.



   The left pane shows available fields, and the right pane shows fields that will be displayed as part of the track 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.

3. Select and organize the fields you wish to display as part of the track 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

Copyright DNASTAR, Inc. 2022

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.

4. When you are satisfied with the choices made in this dialog, click **OK** to save changes and exit from the dialog. Or click **Cancel** to leave the dialog without saving any changes.

Copyright DNASTAR, Inc. 2022

Customize the Look and Layout

The following topics describe the many ways you can customize the look and layout of SeqMan Ultra.

- Use the Style panel to change the look of a view (font styles and sizes, colors, number of decimal places, etc.)
- Customize the appearance and layout of views
- Customize tables in the views (descriptions of table columns)
- Customize the layout of other window components
- Apply information tracks to a view

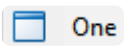
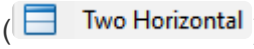
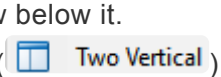
Once you have customized the look and layout as desired, you can:

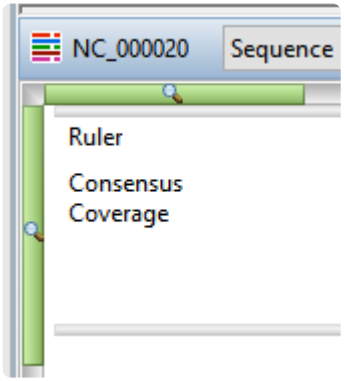
- Save Styles for individual views and apply them to other projects
- Save the entire layout and apply it to other projects

Copyright DNASTAR, Inc. 2022

Customize the appearance and layout of views

The following table lists ways in which you can customize the appearance and layout of views.

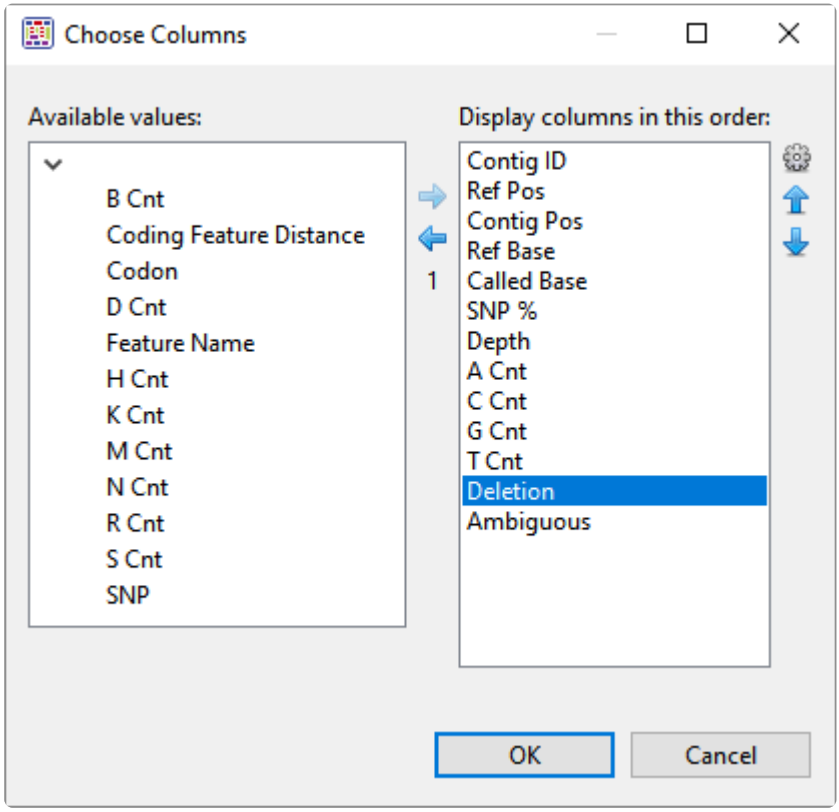
Task	How to…
Show a view	Use subcommands of **View > (View Name)**, which may include a contig number or the word **Show**. For instance, to see the Alignment view for Contig x, you would choose **View > Alignment > Contig x > Show**.
Change the style (fonts, colors, calculation metrics and more)	Use the Style panel and its subsections. Customization for font style and size applies to all text in a view. Individual rows cannot have different styles.
Detach a view	Only certain views can be detached. To do this, use subcommands of **View > (View Name)** that end in **Detach**, or right-click on the view tab and choose **Detach (View Name)**. Detaching a view turns it into a free-floating window.
Reattach all detached views to their original locations	Use **View > Reattach All Views**.
Show the active view in full-screen mode	Use the **View > Full Screen** command or the the **Full screen** tool  at the bottom right corner of the view.
Exit from full-screen mode	Click the **Esc** key, use the **Exit Full Screen** tool () at the bottom of the full-screen view, or press **Ctrl/Cmd+Shift+F**.
Configure the number and location of visible views.	Use the **View > Document Layout** command or the **Change layout** tool (image varies) in the bottom right corner of the view to access possible layouts. The following options are available: • **One** –  – To show only the current view. • **Two Horizontal** () – To show the current view on top and one additional view below it. • **Two Vertical** () – To show the current view on the left, and one additional view to the right. • **Three Horizontal** () – To show the current view on top and two additional views below it.

	• **Three – Left** (Three - Left) – To show the current view on the left and two additional views stacked one on top of the other on the right. • **Three – Right** (Three - Right) – To show the current view on the right and two additional views stacked one on top of the other on the left. • **Three – Top** (Three - Top) – To show the current view on the top and two additional views side-by-side on the bottom. • **Three – Bottom** (Three - Bottom) – To show the current view on the bottom and two additional views side-by-side on the top.
Zoom in or out on a view	Some 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. 
To view two or more documents simultaneously on a split screen	By default, only one document at a time is displayed, even if several documents are open. To view two or more documents simultaneously on a split screen, drag a project tab away from its current position. When you see a black arrow, release the mouse button. Documents can be re-docked in their original positions by reversing this process.
Return to the default window display	Choose **View > Reset Panel Layout**.

Copyright DNASTAR, Inc. 2022

Customize tables in the views

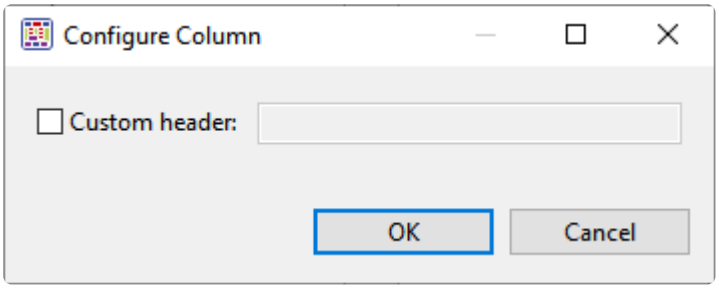
The Contig Report, Features, and Variants views allow you to add data columns to a table, rename the columns, change the order of the columns, or remove existing columns. This is done by clicking the **Change options in this view** ( ) tool in the top left of the header.



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**.

Copyright DNASTAR, Inc. 2022



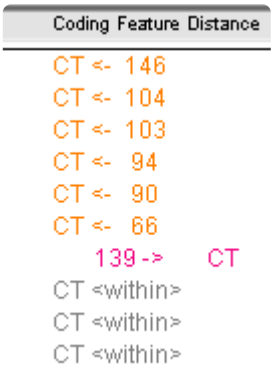
Copyright DNASTAR, Inc. 2022

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	This column is only available if **Show Codon Bases & Distance to feature** is unchecked. The column shows the change(s) in the amino acid sequence, using the nomenclature established by the Human Genome Variation Society and The Sequence Ontology Project. This includes: • 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). • 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). • Conservative in-frame deletions. Example: p.(C28_M30del) a deletion of three amino acids, from Cysteine-28 to Methionine-30. • 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).
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	Shows whether variants are within or near a named feature, and the distance from that feature. For *.assembly* files and certain *.sqd* files (e.g., from *de novo* or special templated workflows), the following color scheme may be used: • Gray + feature name + – Variant is within the named feature.

Copyright DNASTAR, Inc. 2022

	• Pink + arrow + feature name – Distance from the variant to the closest upstream coding feature. • Orange + arrow + feature name – Distance from the variant to the closest downstream coding feature. 
Codon	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 Features view.
Cons Pos	Consensus position that includes gaps.
Contig ID	The name of the reference sequence or chromosome.
Contig Pos	Consensus position that includes gaps
COSMIC	The Catalogue of Somatic Mutations in Cancer (COSMIC) ID for positions with known variants. Double-clicking on the entry opens the corresponding page at COSMIC. For human assemblies only.
Count	(isoforms only) The total number of aligned reads assigned to an isoform feature, after adjustments for repeat distribution.
Coverage Depth	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.
Coverage Strand Depth	(stranded RNA-seq only) The total number of sequenced bases within the

Copyright DNASTAR, Inc. 2022

	feature and on the correct strand, divided by the length of the feature.
dbSNP ID	The dbSNP rs ID, if available, for positions with known variants. Double-clicking on the entry opens the corresponding page at dbSNP.
Deletion	The number of deleted bases in the aligned column.
Depth	The number of reads overlapping the aligned column. Since this calculation disregards bases below the quality threshold, the Alignment view may show a greater number of sequences than the **Depth** 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.
DNA Change	Change(s) in the DNA sequence affecting either CDS features or splice sites are indicated using the nomenclature established by the Human Genome Variation Society. A “c.” prefix, followed by coordinates taken from the ORF, denotes a change in a CDS feature. For example: • Substitutions. Example: c.76A>C denotes that at nucleotide 76 an A is changed to a C. • Insertions within coding regions. Example: c.76_77insT denotes that a T is inserted between nucleotides 76 and 77. • Deletions within coding regions. Example: c.76_78delACT denotes an ACT deletion from nucleotides 76 to 78. A “g.” prefix followed by genomic coordinates denotes a change in the intronic region of a splice site. **Note:** 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.
Deletion	
Depth	
Feature Name	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 **Feature Name** to indicate which isoform from the Feature view table was used (e.g., TP53 [2]).

Copyright DNASTAR, Inc. 2022

	Note: 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.
Feature Type	For variants within a gene feature, the feature type is shown in the following order of precedence: **CDS**, **mRNA**, **Gene**. If **Show Codon Bases Distance to feature** 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 **CDS**, **mRNA**, etc. When **Show Codon Bases** is checked, the **Feature Type** 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: • **gene** – 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.) • **CDS** – Within an exon or splice site. • **mRNA** – In the 5’ and 3’ untranslated portions of an mRNA.
GERP	The Genomic Evolutionary Rate Profiling (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. 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.
Genotype	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

Copyright DNASTAR, Inc. 2022

	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. **Note:** 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.
Group	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 Alignment view to check whether there is a Sanger-called variant at the same position.
Homopolymer	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 homopolymeric indels. **Note:** When possible, insertions or deletions are placed at the 5’ end (top strand) of the run during alignment.
Impact	The impact of the variant or indel on the protein, displayed as: • **Synonymous** – No amino acid changes. • **Non Synonymous** – Amino acid substitution only. • **Nonsense** – Amino acid to translational stop. • **Frameshift** – An indel within a coding region and which is not a multiple of 3, thereby changing the reading frame. • **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.

Copyright DNASTAR, Inc. 2022

	• **Inframe Insertion** – An insertion within a coding region whose length is divisible by 3. The type is followed by the word Conservative if the insertion occurs between two codons, and Disruptive if it occurs with a codon. • **Inframe Deletion** – A deletion within a coding region whose length is divisible by 3. The type is followed by the word **Conservative** if the insertion occurs between two codons, and **Disruptive** if it occurs within two codons. If sorting by the **Impact** column, the column is ordered by severity. For example, a **Frameshift** is more severe than a **Nonsense** change.
Location	The range of sequence associated with the feature, including gaps.
MID	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.
Name	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.
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%.
PDB ID	Worldwide Protein Data Bank (PDB) ID number.
Q Call	The Phred-like quality score of the called genotype. It is a measure of the probability that the called genotype is correct.
Quality	
Raw_count	(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.
Raw_repeat_count	(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.
Ref	

Copyright DNASTAR, Inc. 2022

Ref Base	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.
Ref ID	Reference sequence or chromosome.
Ref Pos	Reference position that does not include gaps. Coordinates matching entries in the VCF Variant table are shown. For deletions, **Ref Pos** is the genomic coordinate of the first deleted base. For insertions, **Ref Pos** is the genomic coordinate of the base preceding the insertion.
Region Capture	Indicates whether the variant occurs within a region specified in the *.bed* or manifest file used. Values are Yes and No.
Repeat_distrib_count	(isoforms only) The proportional number of repeated reads assigned to this exon, gene or isoform.
Repeat_distrib_percent	(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.
Replicate set total count	(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.
Replicate set total RPK	(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.
Replicate set total RPKM	(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.
Replicate set total RPM	(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.
Sequence	The name of the sequence that contains the feature.
Sequence Pos	
Skew	When a maximum “strand bias” (also called “skew”) is set for a templated SeqMan NGen assembly (i.e., by setting the scripting parameter snp_maxStrandBias to ‘true’), SeqMan NGen calculates the strand bias for each variant. The results can be viewed in the Skew column of this table. 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

Copyright DNASTAR, Inc. 2022

	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. SeqMan NGen calculates strand bias using the following formula: Strand bias = (SNP%f – SNP%r)/(SNP%) … where: • SNP%f= Strand-specific SNP percentage for the forward strand • SNP%r= Strand-specific SNP percentage for the reverse strand • SNP%= Overall SNP percentage Below are interpretations of several **Skew** values: • 0 = unbiased • 2 = all SNPs on one strand, where strands are equally abundant • > 2 = SNPs are present in large numbers on a strand that is, itself, rare. 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. 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.
SNP	Shows the symbol for whether this is a confirmed (checkmark), putative (question mark), rejected (‘x’) or mixed (dot) variant.
SNP %	The percentage of the single most prevalent non-reference base in the aligned column. As a very general rule, significant variants tend to occur at 25% and higher.
Splice	Variant is in or near an exon splice site. Splice site variations are changes to the

Copyright DNASTAR, Inc. 2022

	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.
Total_count	(RNA-seq with no normalization specified only) The sum of the counts for all of the isoform features within a gene.
Target Length	The target length of the fragment, equal to the end position minus the start position.
Total raw count	(genes only) The total number of reads which were initially, uniquely assigned to any isoform within the gene.
Total_repeat_distrib_count	(genes only) The total count of reads which were proportionally distributed to any isoform within the gene.
Total RPK	(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.
Total RPKM	(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.
Total RPM	(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.
Trace%	
Transcript ID	Transcript ID number from ENSEMBL.
Type	Specifies the variation type as SNP, Del (deletion) or Ins (insertion). For *.assembly* files and certain *.sqd* files (e.g., from *de novo* or special templated workflows), pink typeface may be used to indicate non-synonymous variants.
User ID	Positions corresponding to a custom VCF Variant Table are labeled with the ID from that set.
Variant Count	The number of called variants in the feature. Note that a gene feature in an exome project can have a very low **Coverage Depth** value and still have one or many variants.

Copyright DNASTAR, Inc. 2022

Visible	A checkmark in this column indicates that the feature is currently displayed in the Alignment view (for features on constituent sequences) or the Strategy view (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 layout of other window components

Controls for this application's view, panels and expand bars are described in the following table. For information that pertains only to views, also see Customize the appearance and layout of views.

Task	"How To"
Show or open a hidden item	Use the associated **View** menu command, click on a closed expand bar, click on a plus sign, or click the **Restore** icon ().
Show/Hide the side panels	Double-click on the project name to toggle between showing and hiding the side panels on the right of the window.
Display panel sub-sections	Click on a closed expand bar (medium gray bar) or use the associated **View** menu command. If you can't see all the information in an expanded section, close other expanded sections and/or increase the size of the whole panel as described elsewhere in this table.
Hide panel sub-sections	Click on an open expand bar.
Maximize an item	use the **Maximize** icon ( ).
Change the relative sizes of panes	Drag the divider bar between them to the left, right, up or down.
Hide or close a visible item	Use the **Close** icon ( ) or **Minimize** 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 **Close**, **Close Others**, **Close Tabs to Right**, or **Close All**. Some of these options are only available for certain tabs.
Move a view, panel or project to another location	Right-click on the view tab and choose **Move (View Name) to (Location)**. In some cases, you can use your mouse to grab a project, view or panel tab, then drag and drop it in the desired location.
Scroll vertically	Use the vertical scrollbar or the mouse wheel.
Scroll horizontally	Use the horizontal scrollbar, or hold down the **Alt** key (Win) or **Option** key (Mac) and drag the mouse.

Copyright DNASTAR, Inc. 2022

Return to the default side panel layout	Choose **View > Reset Side Panel Layout**. This command restores the default layout of panels, including opening any default panels that have been closed; and restoring panels that have been minimized or maximized. It will not reattach a detached view. In that case, use **View > Reattach All Views** instead.
Return to the default layout for the entire window	Choose **View > Document Layout > Restore DNASTAR Default Layout**.

Copyright DNASTAR, Inc. 2022

Save the layout and apply it to other projects

You can change the layout of SeqMan Ultra's views and panels as described in Customize the look and layout.

Once you have arrived at a layout you like, you can save it and then apply it to future projects, either by default or only when you specify that it be applied.

Task	How To
Save the current layout to a file that can be applied to future projects	Use **View > Document Layout > Save Layout**.
Save the current layout to a file that will be applied as the default layout for future projects	Use **View > Document Layout > Save Layout as Default**.
Apply a layout that was saved earlier to the current project	Use **View > Document Layout > Apply Layout**.
Restore the DNASTAR default layout to the current project	Use **View > Document Layout > Restore DNASTAR Default Layout**.

Copyright DNASTAR, Inc. 2022

Search Within the Project

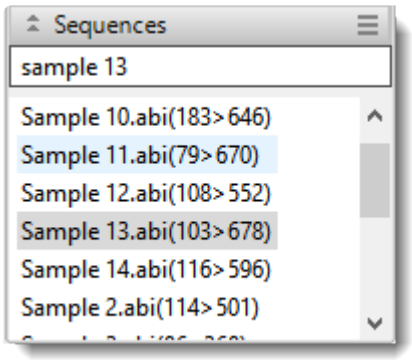
See the following topics for different ways to search within a project.

- Search for and select a read sequence by name
- Search for elements within a contig
- Search for a numerical position on a contig

Copyright DNASTAR, Inc. 2022

Search for and select a read sequence by name

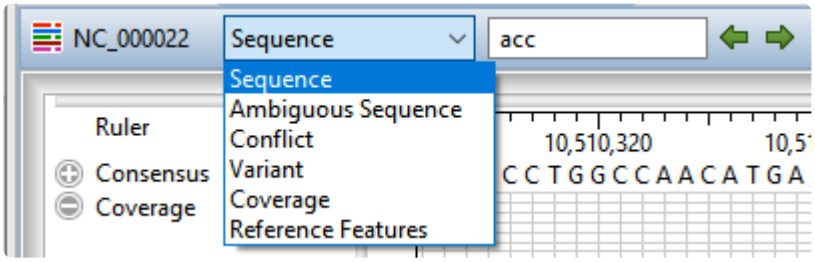
In the Explorer panel, open the Sequences section (if closed) by clicking on the **Sequences** bar. At the top, type the name of the sequence in the search bar (labeled “Type a read name”) until the sequence is highlighted. Alternatively, once the sequence comes into view, click on it to select it. The sequence will now be selected in all views.



Copyright DNASTAR, Inc. 2022

Search for elements within a contig

By default, the search tools are hidden. To display them, use **Edit > Find** or 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. Different views have different types of search targets. 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 **Find** drop-down menu to choose the item to search for. The subset of available options depends on the project type (*.sqd* vs. *.assembly*), assembly type (*de novo* vs. templated), and view in which the search is being done (Alignment vs. Strategy). The table below shows possible options in alphabetical order.

Search Target	Description
Conflict	Search for any conflicts between the read files and the consensus or reference sequence. This search type 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. **Conflict** is available for *de novo* assembly projects only.
Coverage	Search for the edge of a region that is below the minimum threshold coverage. The coverage threshold can be customized in the Strategy Viewing and Coverage parameters dialog.
Go To consensus position	Go to a specific numerical position on the consensus sequence.
Go to reference position	Go to a specific numerical position on the reference sequence.
In all sequences	Search for the query bases (text string) in all sequences: reads, reference (if any) and consensus.

Copyright DNASTAR, Inc. 2022

In consensus	Search for the query bases (text string) only in the consensus row.
In selected sequence	Search for the query bases (text string) only in the currently selected sequence.
Reference features	Searches for and selects the next or previous feature on the reference sequence. To see these features in the view, be sure to check the **Reference > Features** box(es) in the Tracks panel.
Sequence	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.
Variant	Locates bases that have been identified by SeqMan NGen's SNP-calling detector as variants. **Variant** is available for reference-guided assembly projects only.

2. If the search type requires it, enter the query bases (text string) in the text box before using the arrows. Otherwise, skip to the next step. In some cases, a box is provided to **Allow ambiguity**. Check the box if you want to include IUPAC ambiguity codes in your search.

   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.

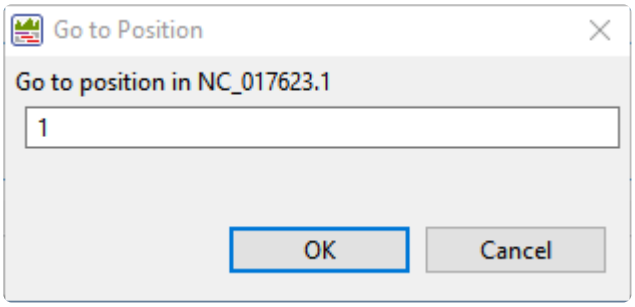
3. Use the right/left arrows () to jump to the next/previous find. After reaching either end of an alignment, SeqMan Ultra wraps around and continues searching from the other end.

Copyright DNASTAR, Inc. 2022

Search for a numerical position on a contig

To search for a numerical position on a contig:

1. Select the contig in the Explorer panel.
2. Open the Alignment or Strategy view for the contig.
3. Choose **Edit > Go to Position**.



4. In the dialog, type in the numerical coordinate and then press **OK**.

Copyright DNASTAR, Inc. 2022

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**.

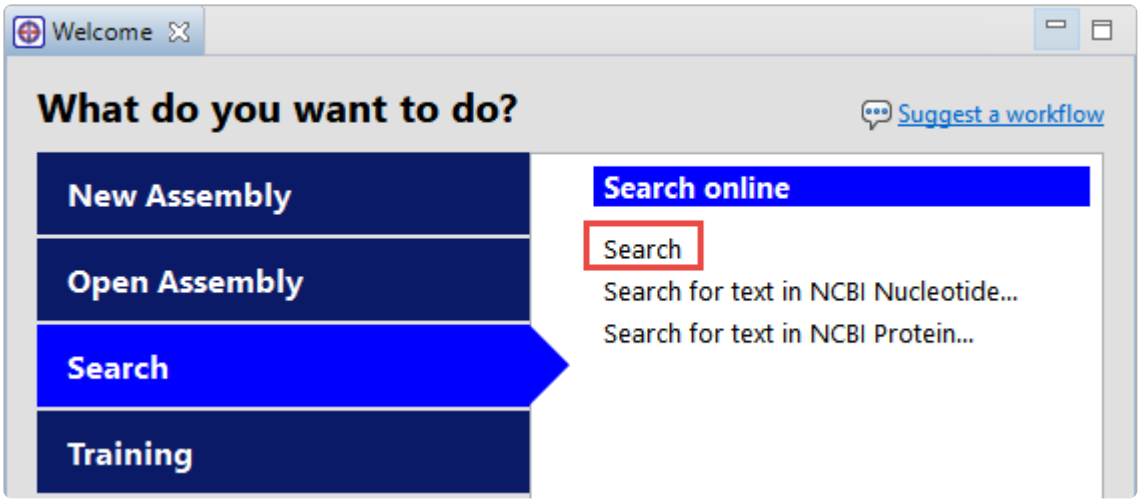
Copyright DNASTAR, Inc. 2022

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 **Search** on the left of the Welcome screen, then choose **Search** on the right 	Launches the wizard at the Query screen.
Search > Sequence Search	Launches the wizard at the Query screen.
Use the "Change search strategy" tool ( ) from the Text, Table or Pairwise views	
Search > Sequence Search with Selection	Launches the wizard at the Query screen, whose table is pre-populated with the currently-selected portion of sequence.
Press **Ctrl/Cmd+B**	

Search > Sequence Search with All Sequences	Launches the wizard at the Query screen, whose table is pre-populated with all sequences in the document.

Copyright DNASTAR, Inc. 2022

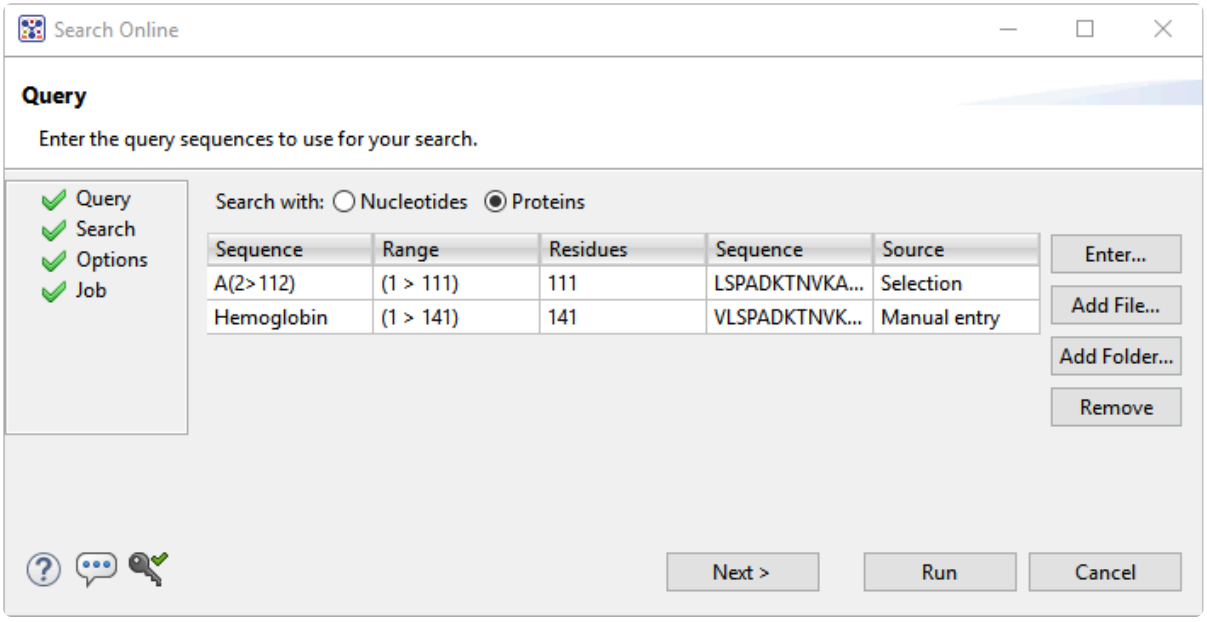
Search wizard: Query

The Query screen is the first of four screens in the BLAST Search wizard. The Query screen is where you add the query sequence, i.e., the sequence for which you wish to locate matches in the NCBI BLAST database.

At the top of the screen, next to “Search with,” select the **Nucleotides** or **Proteins** button.

To add one or more query sequence files or a folder of sequences:

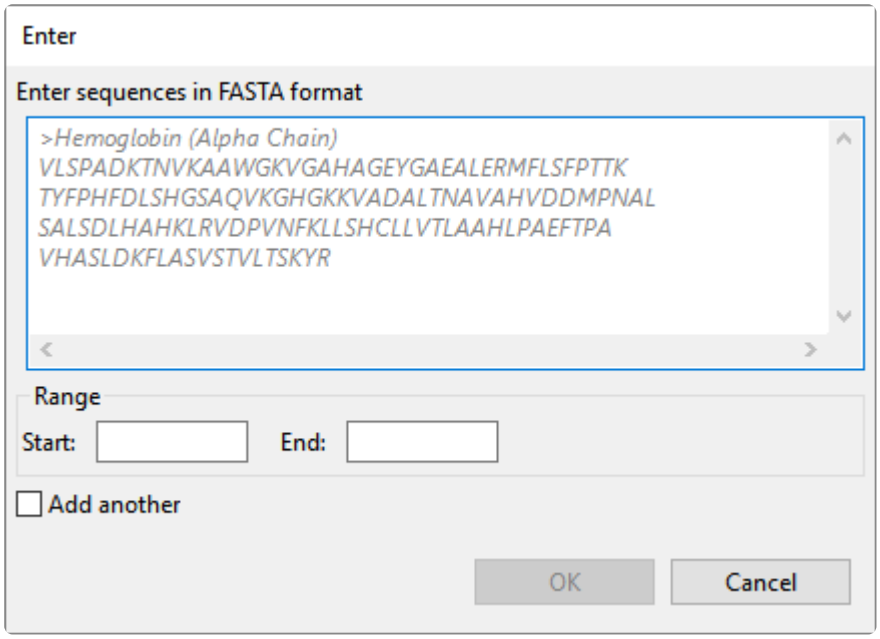
Press the **Add** or **Add Folder** buttons. In the file explorer, navigate to and select the desired file(s)/folder(s), and then click **Open**. If you add a multi-file sequence or multiple sequences, each sequence will appear on a separate row in the table.



To paste or type in the query sequence:

Press **Enter**.

Copyright DNASTAR, Inc. 2022



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**.

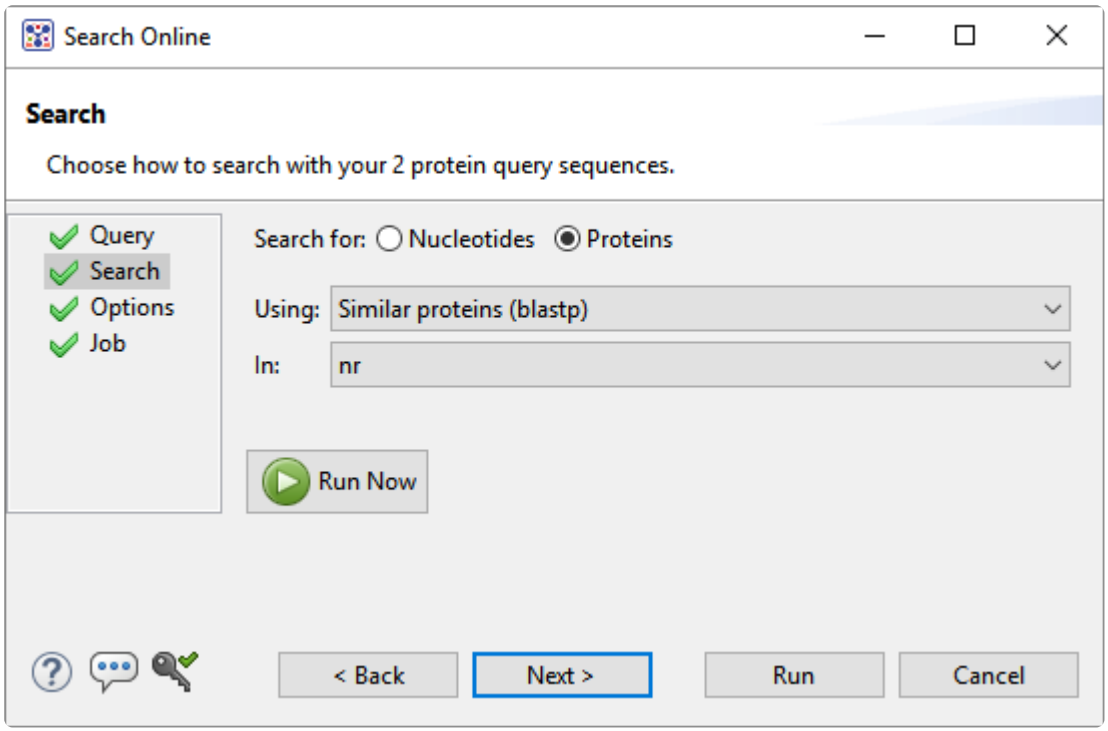
Copyright DNASTAR, Inc. 2022

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

Copyright DNASTAR, Inc. 2022

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.

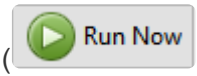


- Specify whether you wish to search for **Nucleotides** or **Proteins**.
- Select the desired BLAST program from the **Using** drop-down menu. For information on the different BLAST programs, see the BLAST Program Selection Guide PDF. Available options depend on the combination of the selections you 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.

Copyright DNASTAR, Inc. 2022

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:

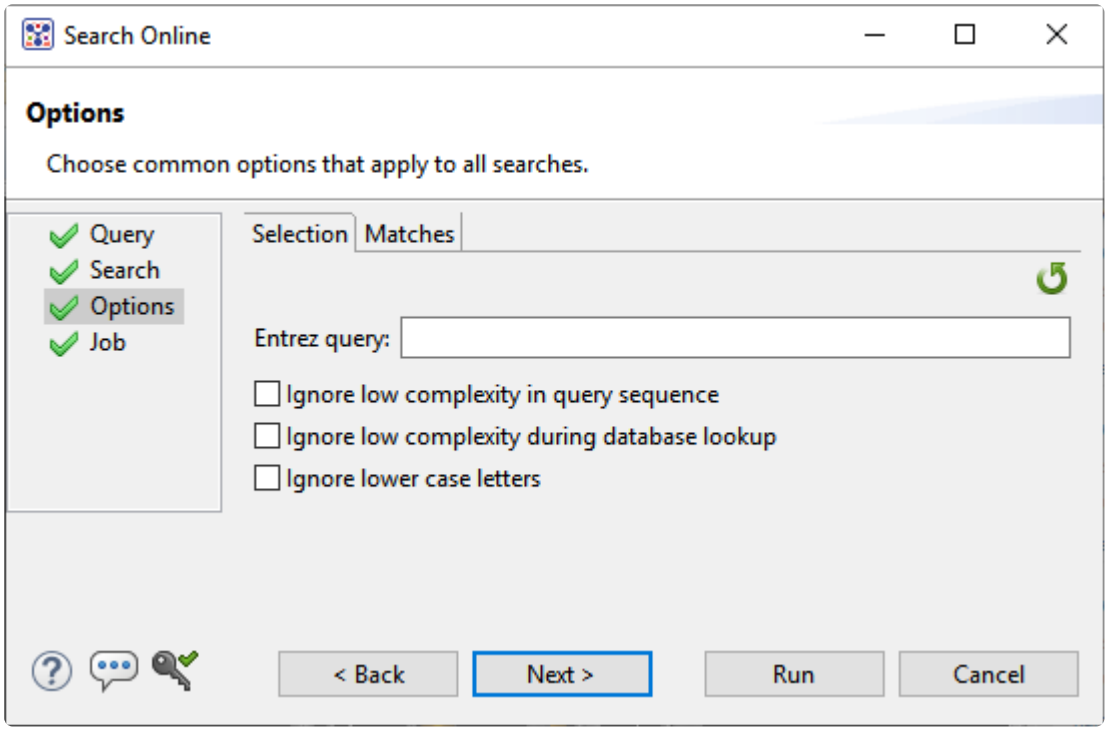
- To go back to the Query screen, press **Back**.
- To add additional parameters to the search, press **Next** to continue to the Options screen.
- To start the search immediately, press the **Run Now** button (

) 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.

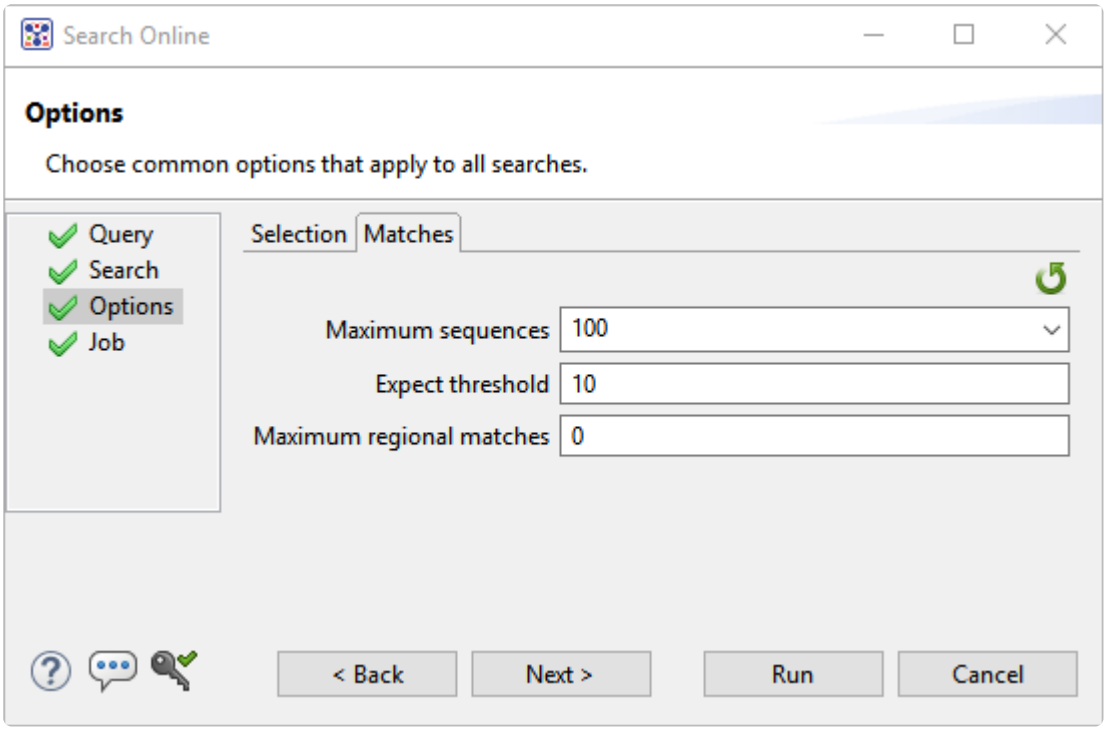


In the **Selection** tab:

  - 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.

Copyright DNASTAR, Inc. 2022

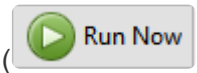


In the **Matches** tab:

- 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.

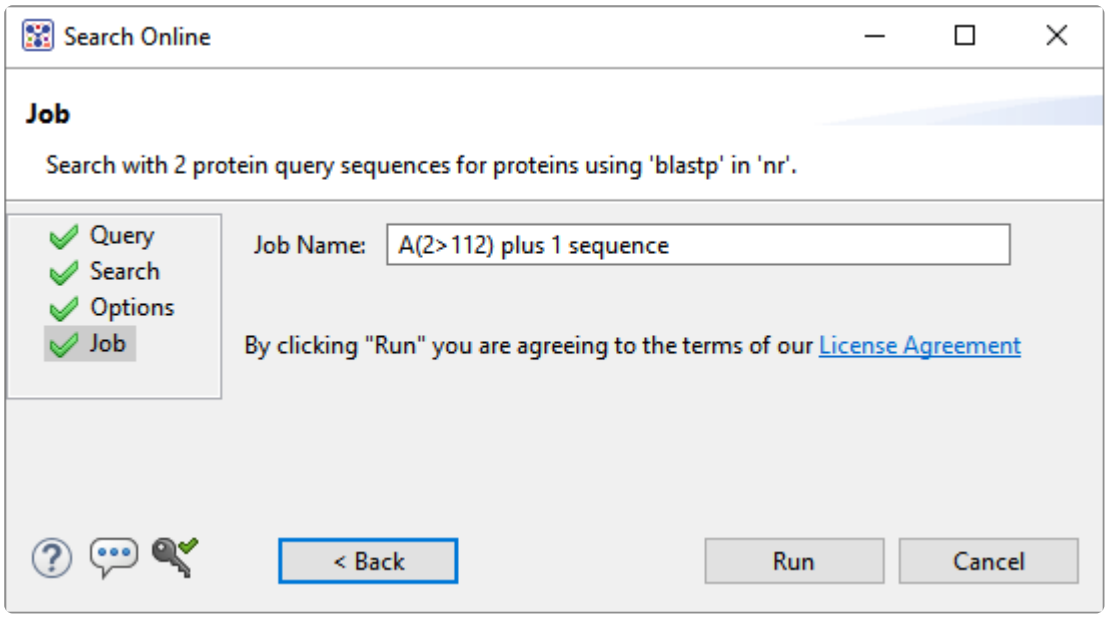
Copyright DNASTAR, Inc. 2022

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

Copyright DNASTAR, Inc. 2022

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.



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**.

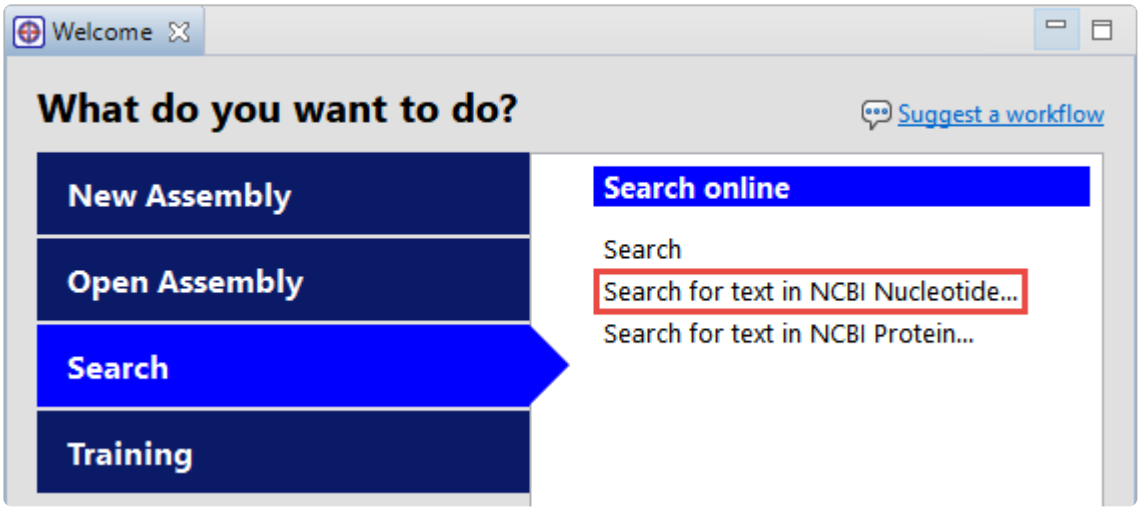
Copyright DNASTAR, Inc. 2022

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.



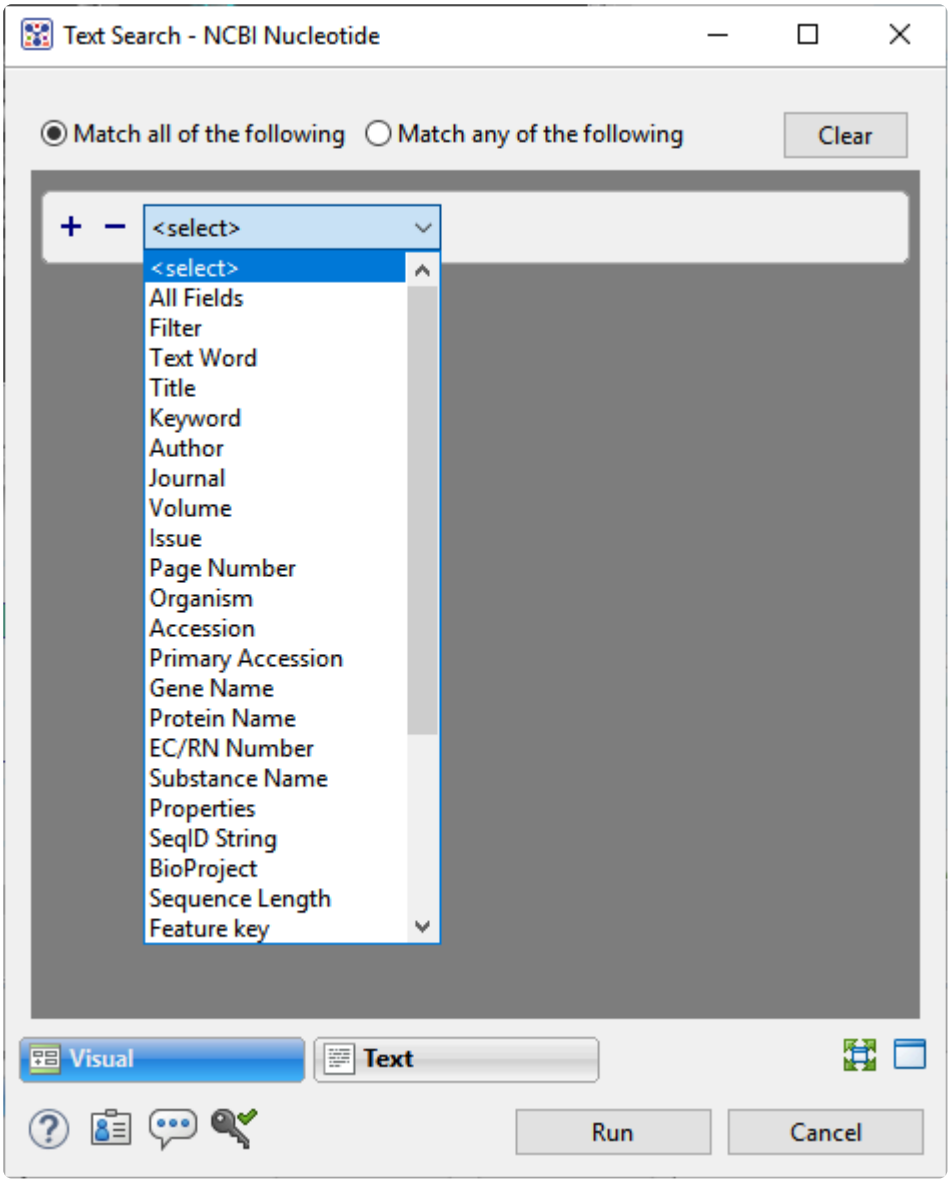
- To initiate a search of the protein database, use **Search > Text Search in > NCBI Protein** or click on the **Welcome Tab**, choose **Search** on the left and then **Search for Text in NCBI Protein** on the right.

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.

1. 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.

Copyright DNASTAR, Inc. 2022

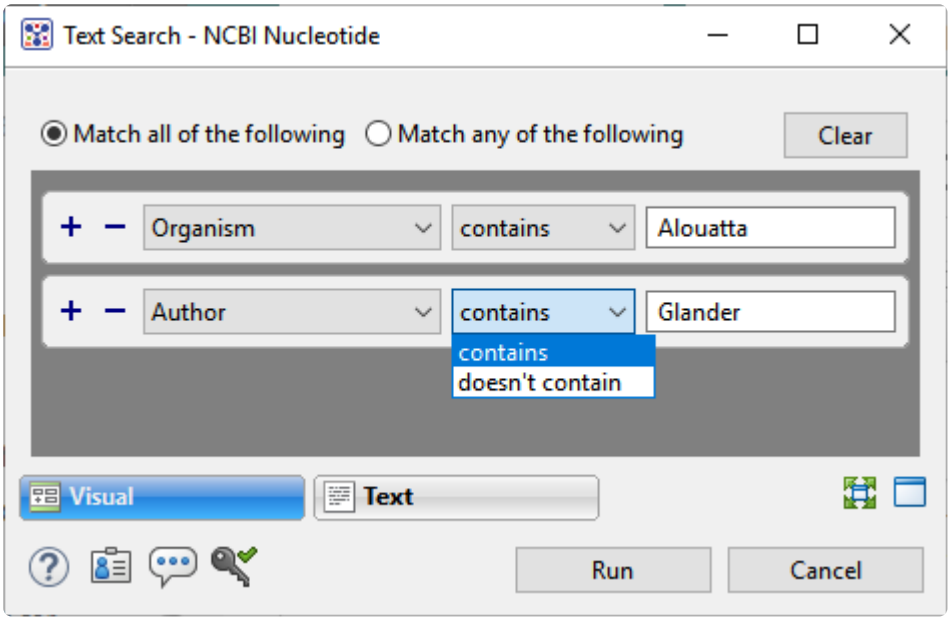


Once you make a selection, a second drop-down menu and an empty text box appear to the right.

- 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 (−).

Copyright DNASTAR, Inc. 2022



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.

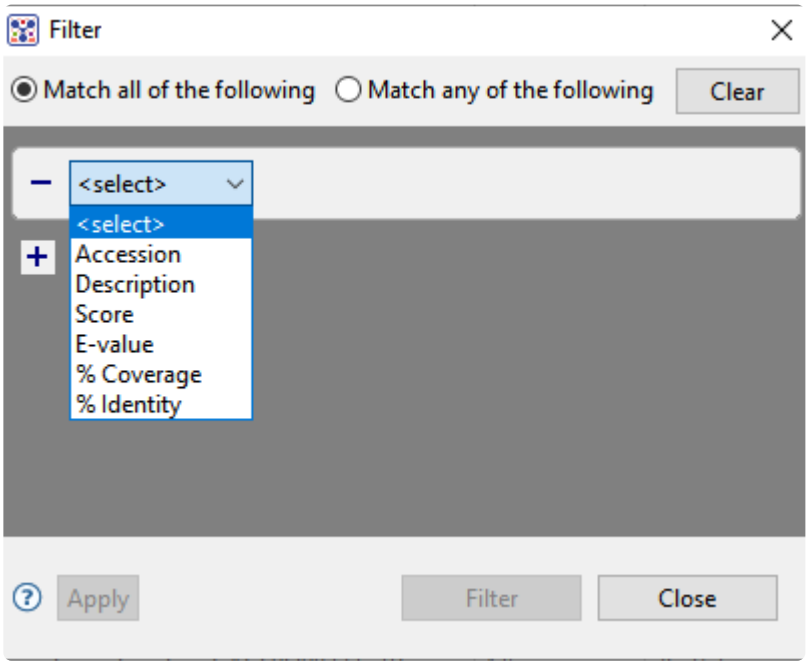
5. Click **Run** to send the query to the database. To instead exit the dialog without searching, press **Cancel**.

During the search, a progress dialog is displayed. Once the search is completed, any matches are displayed in the Jobs panel.

Copyright DNASTAR, Inc. 2022

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
Accession	The Accession number.	
Description	The NCBI “Definition” field text.	
Score	The similarity score for the match. (See the NCBI definition for **raw score**). In general, a higher score denotes a better match.	
E-value	The chance that the match is expected to occur coincidentally. (See the NCBI definition for **E value**). In general, a lower score denotes a better match.	
% Coverage	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. **%Coverage** = 565/596 = 95%.	
% Identity	The percentage of bases that match when the query and result sequence are aligned. (NCBI definition). Example: Query has 565 bases that align to the match and 3 bases that don’t	

Copyright DNASTAR, Inc. 2022

	match. **%Identity** = 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.

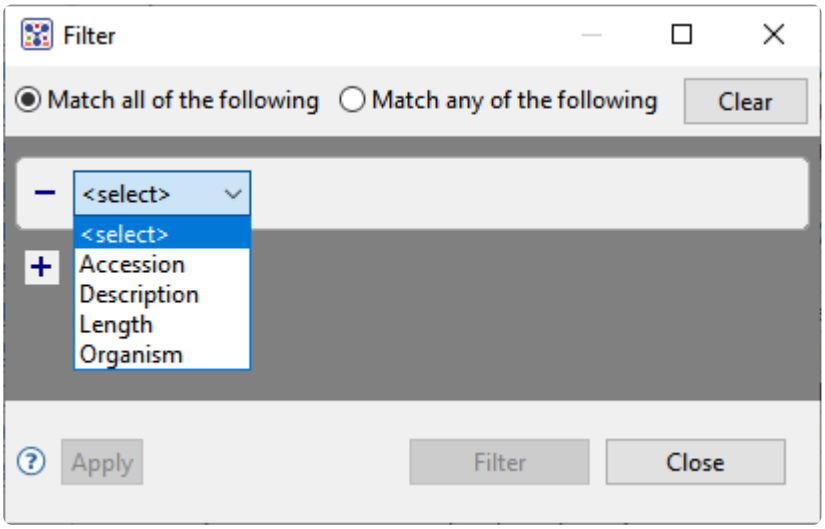
Once you are finished setting up filters, do any of the following:

- 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**.

Copyright DNASTAR, Inc. 2022

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
Accession	The Accession number.	
Description	The NCBI “Definition” field text.	
Length	The length of the match sequence in bases.	
Organism	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**.

Copyright DNASTAR, Inc. 2022

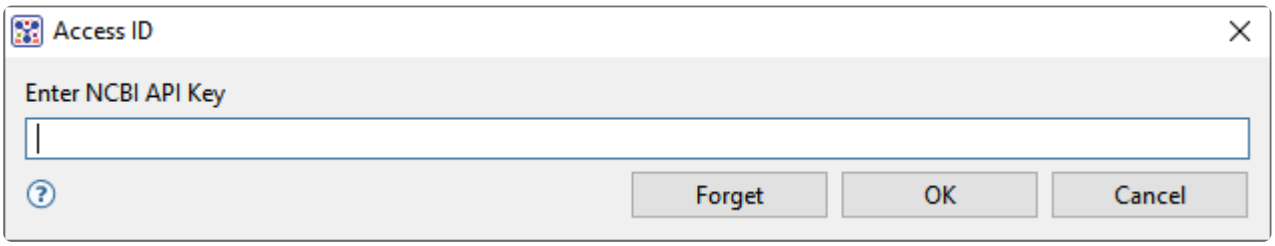
Access ID dialog

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.



- 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**.

Copyright DNASTAR, Inc. 2022

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 Explorer panel, click on the name of the contig.
Multiple read sequences	In the Alignment or Strategy view, use **Ctrl/Cmd+click** or **Shift+click** to select multiple read sequences.
One entire sequence	In the Alignment or Strategy view, click on the sequence name. To search for a sequence by name, open the Sequences section of the Explorer panel and use the search bar at the top.
A portion of a constituent or consensus sequence	In the Alignment or Strategy views: • 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 **Ctrl/Cmd+drag**. • Click on the left-most residue you wish to highlight. Hold down the **Shift** key and click on the right-most residue. • To select an entire sequence, insert the cursor anywhere along the sequence, then right-click and choose **Select All**.
Everything that can be selected	Choose **Edit > Select All**.
Nothing (remove all selections)	Choose **Edit > Select None**.
Everything but the currently selected range of the consensus	In the Alignment or Strategy views, drag the mouse across the consensus sequence to select the range you do *not* want. Then choose **Edit > Invert Selection** to unselect that portion and select the other part(s) of the consensus.
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 feature to select it. To add additional features to the selection, use **Ctrl/Cmd+click**.

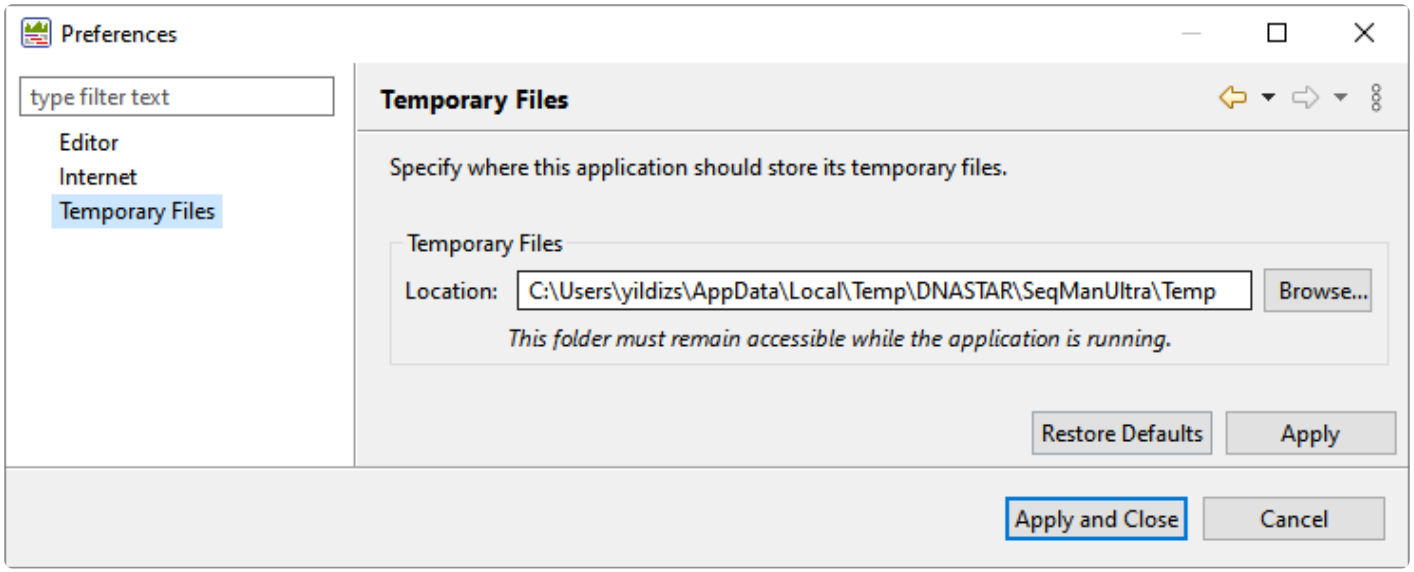
Copyright DNASTAR, Inc. 2022

Specify Parameters and Preferences

The following dialogs are accessed through the **Project > Project Parameters** command and are used to customize assembly and other parameters in SeqMan Ultra.

- Consensus parameters
- Strategy Viewing and Coverage parameters
- Pair Specifier parameters
- Assembly parameters

To set preferences, use **Edit > Preferences** (Windows) or **SeqMan Ultra > Preferences** (Macintosh) to open the three-tabbed Preferences dialog.



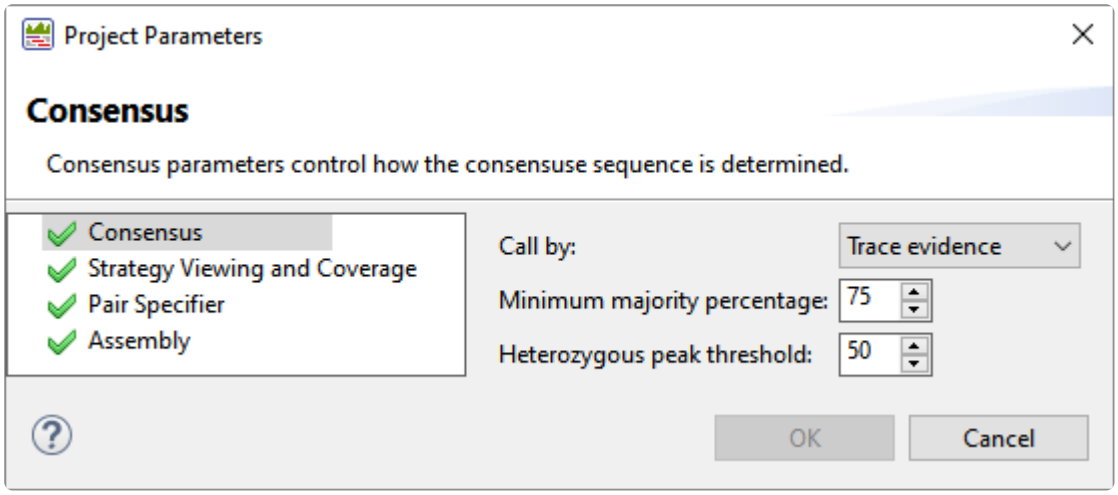
- **Editor** tab – To choose whether or not to open alignments in a detached window.
- **Internet** tab – To display or hide update notifications.
- **Temporary Files** tab – To specify where SeqMan Ultra should store temporary files.

Copyright DNASTAR, Inc. 2022

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**.
    - **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.

Copyright DNASTAR, Inc. 2022

Click **OK**.

Copyright DNASTAR, Inc. 2022

Consensus calling algorithms

In 1999, DNASTAR software developer Carolyn Allex published a doctoral thesis (Allex CF, 1999), in which she compared several algorithmic methods for consensus calling, including the “Majority” method and her own “Trace Evidence” method (Allex CF et al. 1997). The latter method was a novel approach for generating quality scores and consensus calling based on geometry and quality of peaks in the trace data.

Allex’s analysis indicated that Trace Evidence had significantly better consensus calling accuracy than Majority, even when many of the individual bases had been called incorrectly. Trace Evidence was also more likely than Majority to make the correct call when the base of the well-defined (true) peak was hidden below a high-intensity valley. By contrast, Majority methods often incorrectly called the base that was associated with the valley.

DNASTAR implemented the Trace Evidence algorithm into SeqMan Pro and, later, SeqMan Ultra. It remains the preferred consensus calling method today. The Majority method is now recommended only when data consist of text sequences rather than fluorescence trace data.

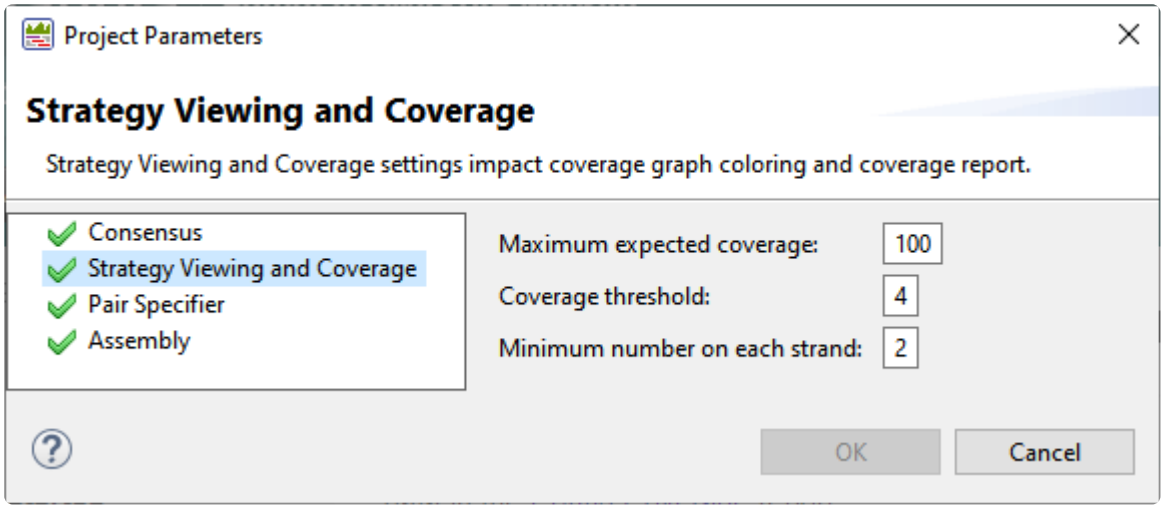
In addition, the quality scoring algorithm that was developed for use with the Trace Evidence method is now also used in SeqMan Ultra for variant calling and quality-based sequence end trimming.

Copyright DNASTAR, Inc. 2022

Strategy Viewing and 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.

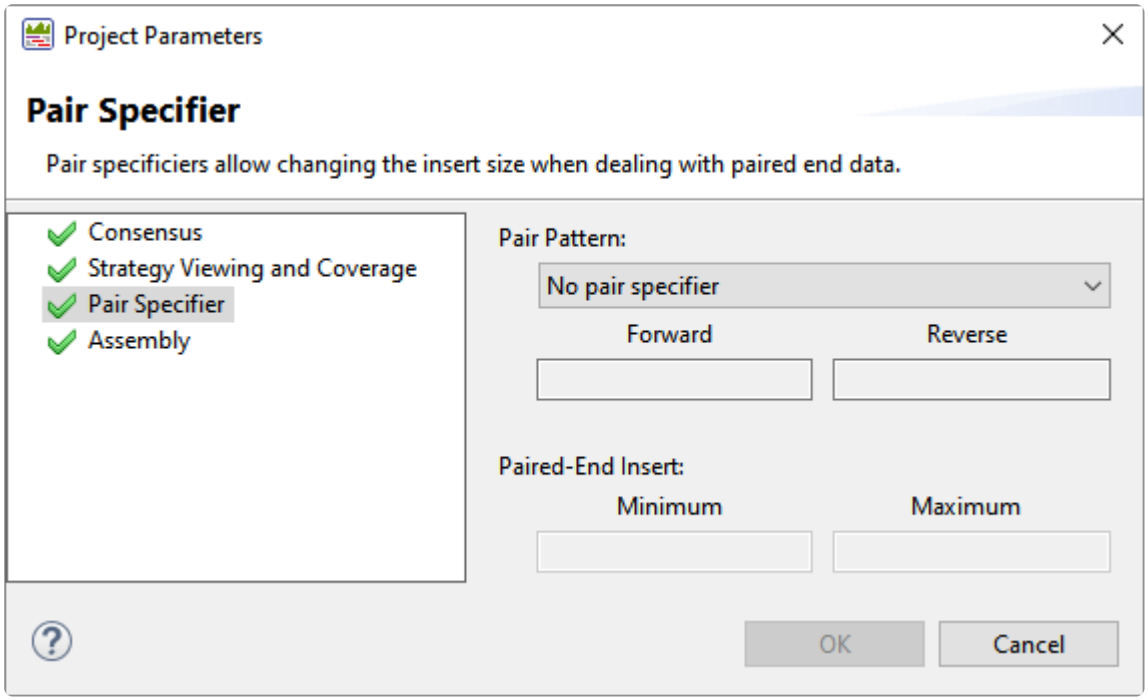
- **Minimum number on each strand** – The minimal number of reads on each strand required to be considered sufficient. Positions/ranges where either strand is below this value will be indicated as “Only one strand” in the Contig Coverage report. The default value is 2.

Copyright DNASTAR, Inc. 2022

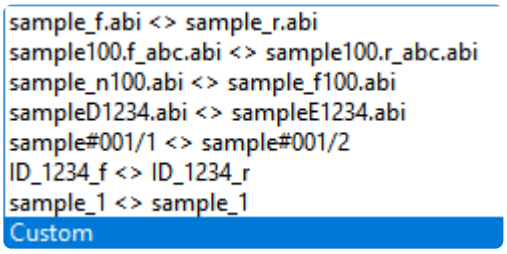
Pair Specifier parameters

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.



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



  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.

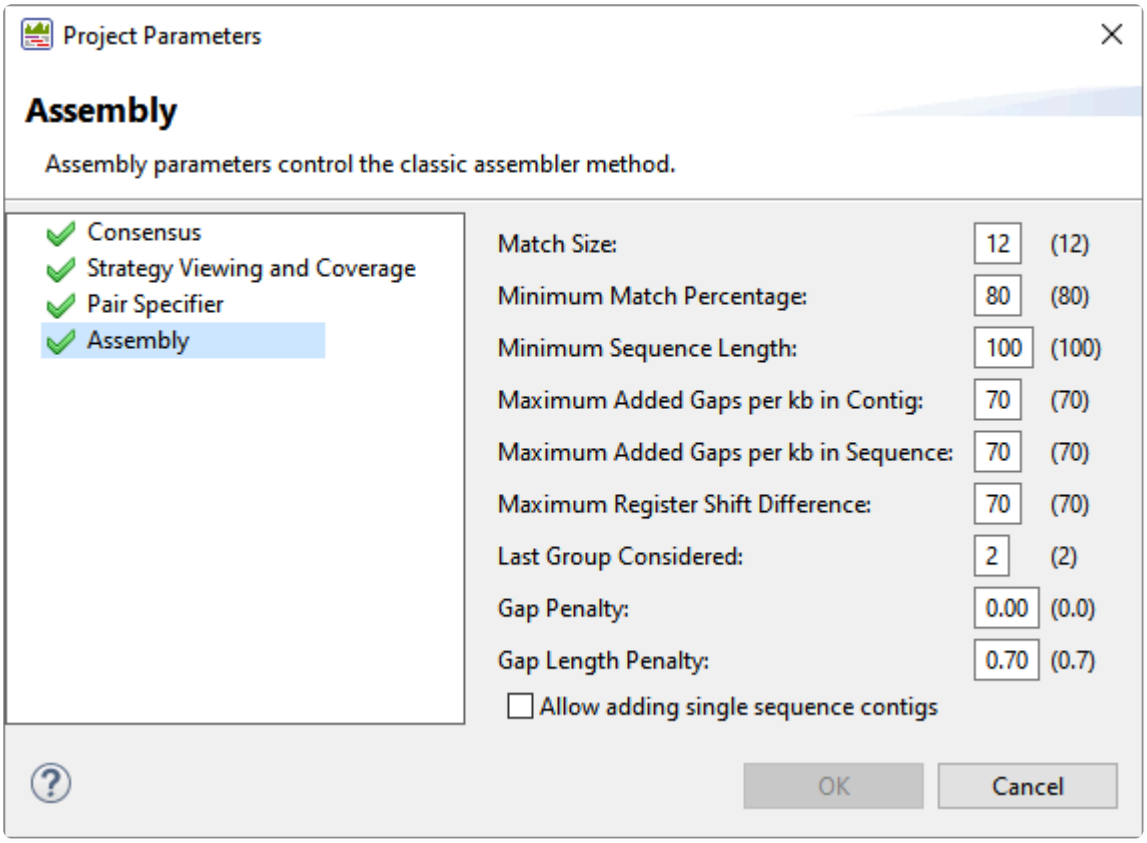
Click **OK**.

Copyright DNASTAR, Inc. 2022

Assembly parameters

Assembling parameters determine how read sequences are combined to form contigs during assembly in SeqMan Ultra.

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



Default values are shown to the right of the text boxes where you can enter custom values.

- **Match Size** – the smallest number of matching consecutive bases required to extend comparisons between the new sequence and a contig or between two sequences.
- **Minimum Match Percentage** – the minimum % of matches in an overlap required to join two sequences in the same contig. (Default is 80%)
- **Minimum Sequence Length** – the threshold length of sequence that qualifies for addition to the assembly. If sequences are first trimmed for vector and/or quality, **Minimum Sequence Length** refers to the remaining length of the trimmed sequences. If you want to include in the assembly sequences smaller than the default value, simply enter a smaller number here. However, note that this value should be no less than the **Match Size** value specified above.
- **Maximum Added Gaps per kb in Contig** – the maximum number of gaps that can be added to a

Copyright DNASTAR, Inc. 2022

contig while merging with a new sequence. If a new sequence requires that a longer gap be inserted in the contig, the new sequence is not allowed to merge. It will be checked against any other contigs before being considered for a new contig.

- **Maximum Added Gaps per kb in Sequence** – the maximum number of gaps that can be added to a new sequence while merging with an existing contig. If a sequence requires a longer gap in order to be inserted in the contig, the new sequence is not allowed to merge. It will be checked against any other contigs before being considered for a new contig.

- **Maximum Register Shift Difference** – the maximum separation, in bases, between nearby matches.

- **Lastgroup Considered** – the maximum number of alignment groups SeqMan Pro investigates to see where a new sequence aligns with a contig. This must be set to one or higher. Alignment groups are groups of identical matches likely derived from the same region of a sequence.

- **Gap Penalty** – the penalty to be deducted from the pairwise score for each alignment for every gap introduced into either strand. Each gap incurs the same penalty regardless of its length. A high gap penalty suppresses gapping while a low value promotes gapping.

- **Gap Length Penalty** – the penalty assessed to gaps in the pairwise alignment step. This value deters long gaps by making them more costly than shorter gaps, proportional to their length. If it doesn't matter to you how long gaps are, this value can be set to zero.

- Decide whether or not to check the box to **Allow adding single-sequence contigs**.

Copyright DNASTAR, Inc. 2022

Export

The following topics discuss how to:

- Export an image of the view
- Export data to a file
- Export data from the Text, Table or Pairwise views

Copyright DNASTAR, Inc. 2022

Export an image of a view

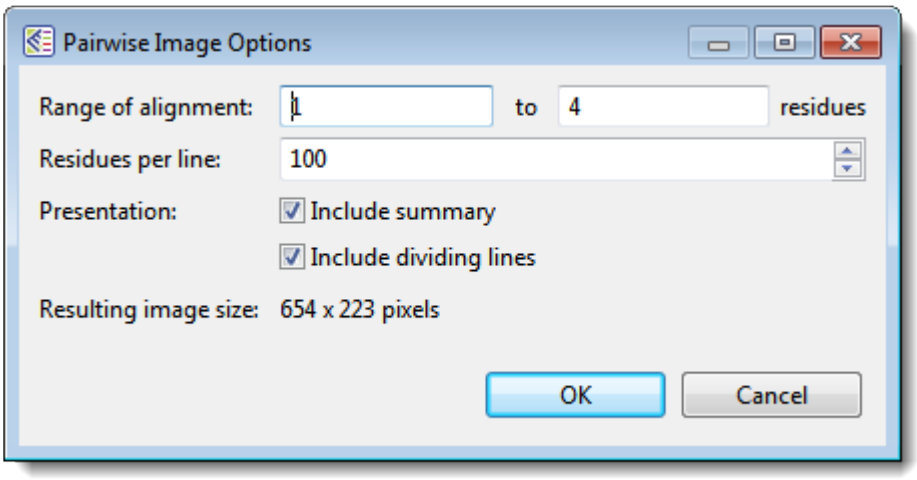
To export an image of a view:

1. Choose **File > Export Image > (View Name)** or click the **Export image** tool ( ) at the upper right of a view that allows exporting.

2. Specify the file name and choose the export format: Adobe PDF (*.pdf*), Microsoft PowerPoint-optimized (*.pptx*), PNG image (*.png*) or JPG image (*.jpg*, *.jpeg*). We recommend choosing PDF for the highest quality output.

> **Notes:** If you choose Adobe PDF format and a particular font is not supported there, a similar font will be substituted for the original one automatically. If you choose the format that is optimized for use with Microsoft PowerPoint, the MegAlign Pro image is converted into an editable set of objects that can be grouped, ungrouped and moved within the target application. To constrain an object (e.g., a MegAlign Pro track) within Microsoft PowerPoint to move only vertically, rather than laterally, hold down the **Shift** key.

The following dialog (or an abbreviated version) will open.



3. Choose export options.

   - Use the **Range of view** or **Range of alignment** boxes to specify the range of residues that should appear in the image.

   - Type in the number of **Residues per line**. By default, MegAlign Pro places the entire range on a single line. If you set the number of residues per line smaller than the range, the image will be rastered to improve image quality. If you set the range to be the same as the number of residues per line, a single raster will result.

Copyright DNASTAR, Inc. 2022

- Check the **include summary** box if you want to include the summary (above the sequences area) in the image.

- Check the **Include dividing lines** box if you wish to include separators between the header, sequences, and footer.

The **Resulting image size** updates automatically based on the options you choose.

4. Click **OK** to export the image using the specified options or **Cancel** to close the dialog without exporting.

The following video shows how to export an image from MegAlign Pro and edit it in PowerPoint. The procedure is similar in SeqMan Ultra.

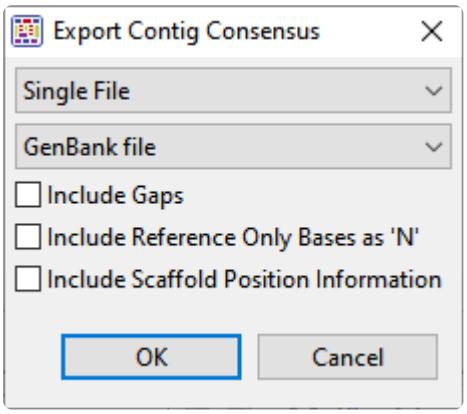
Copyright DNASTAR, Inc. 2022

Export data to a file

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

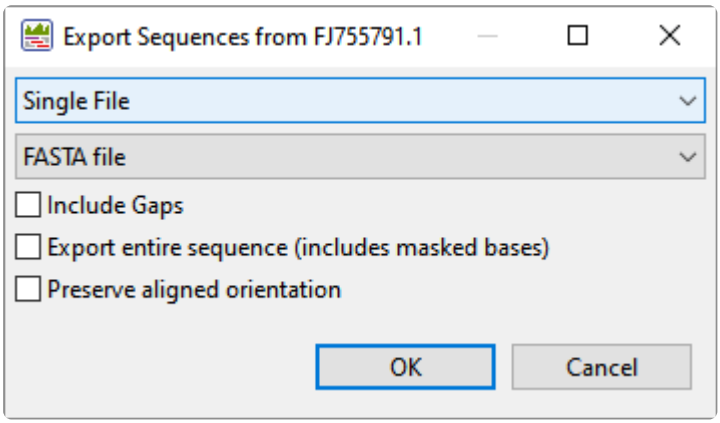
Item to Export	How To
Any version of the Project Report, Contig Report or Variants views	With the view open, click on the **Export data** tool () at the right of the header.
Features view	With the Features view open, click on the **Export data** tool () at the right of the header or choose the **File > Export Data > Features > (Contig Name)** command. The tabular data can be saved in either comma- or tab-separated value (*.csv* or *.tab*) formats.
Data or sequences from the Text, Table or Pairwise views	Click on the **Export data** tool () at the right of the Table view header. Select from the following: • **Matches without Gaps** – To export an ungapped FASTA (*.fasta*) file consisting of matching regions of displayed matches, but not the query sequence itself. • **Aligned Matches** – To export a gapped FASTA (*.fasta*) file with displayed matches aligned against the query. • **Table** – To save Table view contents in either Comma- or tab-separated value (*.csv* or *.tab*) formats. • **Text** – To save the Text view contents in Text (*.txt*) format.
Consensus sequences	Select one or more contigs in the Explorer panel, then choose **File > Export Data > Consensus > (Contig Name)**, **File > Export Data > Consensus > Selected Contigs** or **Contig > Export Consensus of (Contig Name)**. Alternatively, from the Alignment view, use the **Export data** tool (). In each case, the following dialog opens:

Copyright DNASTAR, Inc. 2022



- 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.

```
413511  NNNN
```



- 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 based 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**.

Copyright DNASTAR, Inc. 2022

Export data from the Variants view

To export data from the Variants view in *.csv* or *.tab* format:

1. Select a contig in the Explorer panel and open the Variants view.
2. Use this tool () to filter the table to include the variants of interest.
3. Choose or rearrange columns using the gear tool ().
4. Click the **Export data** tool ().

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%**

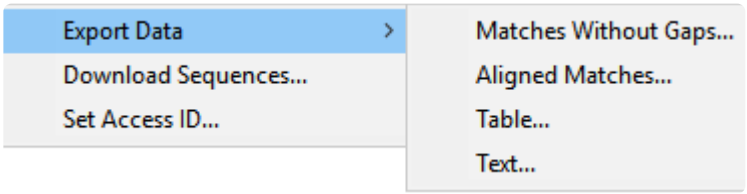
If you wish to determine coding vs. non-coding regions or synonymous vs. non-synonymous changes in ArrayStar, you must also export the following columns:

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

Copyright DNASTAR, Inc. 2022

Export data from the Text, Table or Pairwise views

The **Export data** tool ( ), located in the header above the Text, Table or Pairwise views, lets you export data from those views. This tool provides a choice of three options, one of which has four sub-options.



Options include:

- **Export Data >**
    - **Matches without Gaps** – To export an ungapped FASTA (*.fasta*) file consisting of matching regions of displayed matches, but not the query sequence itself.
    - **Aligned Matches** – To export a gapped FASTA (*.fasta*) file with displayed matches aligned against the query.
    - **Table** – To save Table view contents in either Comma- or tab-separated value (*.csv* or *.tab*) formats.
    - **Text** – To save the Text view contents in Text (*.txt*) format.
- **Download Sequences** – To download the selected match sequence in GenBank (*.gbk*) format.

The commands above all launch a Save As dialog. Choose a file name, location and file type (fixed, in some cases) then press **Save**.

Copyright DNASTAR, Inc. 2022

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 **File > Save All** or press **Ctrl/Cmd+Shift+S**.
To save the current project	Choose **File > Save**, select the **Save** tool (), or use **Ctrl/Cmd+S**. 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 **File > Save As** instead. By default, SeqMan Ultra saves to the desktop unless directed to save elsewhere.
To close an active project	Choose **File > Close Editor** or **Ctrl/Cmd+W**. 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 **File > Exit.** On Macintosh, use **SeqMan Ultra > Quit SeqMan Ultra**. The shortcut **Cmd+Q** 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.

Copyright DNASTAR, Inc. 2022

Appendix

The Appendix contains the following topics:

- Troubleshoot the error
- IUPAC codes
- Supported file types
- Installed Lasergene file locations
- Research references

Copyright DNASTAR, Inc. 2022

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.

Copyright DNASTAR, Inc. 2022

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

Copyright DNASTAR, Inc. 2022

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.

Copyright DNASTAR, Inc. 2022

Supported file types

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

Copyright DNASTAR, Inc. 2022

Installed Lasergene file locations

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

File Category	Application	Path
Application[2]	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[2] (*DMx, DMx.exe)*	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[2]	All	Windows: C:\Program Files (x86)\DNASTAR\License Manager Macintosh: /Applications/ DNASTAR/License Manager
Server License File (*lservrc*), Server License Manager [2], *Server Executables (_lserv, lservnt.exe*)	All	Windows: C:\Program Files (x86)\DNASTAR-LicenseServer\Server

Copyright DNASTAR, Inc. 2022

		Macintosh: ~/Library/DNASTAR-LicenseServer
Standalone & Trial Licenses (*.license), License Server Client License (*.lshost), Key Server Client License (*.keyhost)	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**.

Copyright DNASTAR, Inc. 2022

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]

Copyright DNASTAR, Inc. 2022

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]

Copyright DNASTAR, Inc. 2022