QIAGEN

Tutorial

De Novo Assembly and BLAST

June 27, 2019

Sample to Insight

QIAGEN Aarhus · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark
Telephone: +45 70 22 32 44 · www.qiagenbioinformatics.com · ts-bioinformatics@qiagen.com

De Novo Assembly and BLAST

This tutorial takes you through some of the tools for a typical *de novo* sequencing workflow with a data set from a high-throughput sequencing machine. Here, we de novo assemble some reads, and then search a database at the NCBI with some of the contigs produced. Note that because of the way the De Novo Assembly and the Map Reads to Contigs work (see `http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=_CLC_de_novo_assembly_algorithm.html`), the results you will generate may be slightly different than the ones seen in the figures of this tutorial.

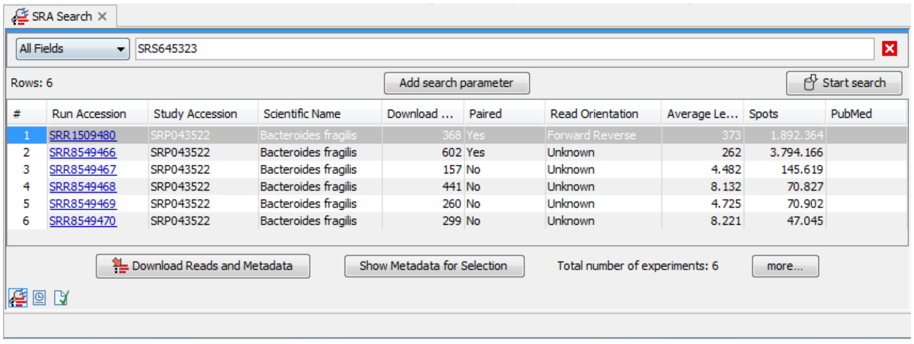
For this tutorial, we use part of an *Bacteroides fragilis* data set from Sydenham et al., 2015.

Importing the data

1. Download the read files using the tool:

   Download | Search for Reads in SRA ( )

2. This will bring up the dialog shown in figure 1. Search for one of the run from the study with the following accession number SRS645323.



Rows: 6

#	Run Accession	Study Accession	Scientific Name	Download ...	Paired	Read Orientation	Average Le...	Spots	PubMed
1	SRR1509480	SRP043522	Bacteroides fragilis	368	Yes	Forward Reverse	373	1.892.364	
2	SRR8549466	SRP043522	Bacteroides fragilis	602	Yes	Unknown	262	3.794.166	
3	SRR8549467	SRP043522	Bacteroides fragilis	157	No	Unknown	4.482	145.619	
4	SRR8549468	SRP043522	Bacteroides fragilis	441	No	Unknown	8.132	70.827	
5	SRR8549469	SRP043522	Bacteroides fragilis	260	No	Unknown	4.725	70.902	
6	SRR8549470	SRP043522	Bacteroides fragilis	299	No	Unknown	8.221	47.045	

Total number of experiments: 6

Figure 1: *Search for a particular run using the Search for Reads in SRA tool.*

3. Select the run with a "Forward reverse" read orientation and click "Download reads and Metadata". In the download dialog that pops up, choose to discard read names and quality scores.

4. In the next dialog, the paired end settings are already estimated. Note that you can always edit them if necessary when using your own data.

5. Save the data in the folder you created for this tutorial.

After a short while, the reads will be imported.

Before we start analyzing the data, it is always good practice to run the Trim Reads tool. The tool automatically detect and trim read-through adapters from the reads, thus preventing potentially left-over adapters from compromising subsequent analyses such as de novo assemblies.

To start trimming:

Toolbox | Prepare Sequencing Data | Trim Reads ( )

1. In the first wizard, select the SRR1509480 sequence list you just saved to the Navigation Area, and click **Next**.

2. Leave all settings as they are set by defaults by using the Reset button in ALL the following dialogs: Quality trimming, Adapter trimming and Sequence filtering. When in doubt, use the Reset button to remove any changes (or a Trim adapter list that could be pre-selected in the Adapter trimming wizard) from a previous run of the tool with different data.
3. Choose to Create a report and Save the results in the appropriate folder of the Navigation Area.

A quick glance at the Trim report (called SRR1509480 report) will inform you on how many adapters where trimmed, based on what criteria. We can now proceed to the rest of the tutorial using the trimmed reads.

Run a De novo assembly

The reads we are using in this tutorial are shorter than 251 bp. By doing a *de novo* assembly of the reads, we are trying to create longer, contiguous sequences from these relatively short sequences. We recommend that you refer to the CLC Genomics Workbench manual for information on how the de novo assembly tool works, and the meaning of the parameters you have control over.

There are two possible types of output you can generate from the de novo assembly tool in *CLC Genomics Workbench*: simple contigs, which is a sequence list of the contigs generated, and mapping objects, where a read mapping is carried out after the initial assembly, to map your reads back to the contigs created.

If you plan to do variant detection later, then generating mapping output can makes sense. When setting up your assembly, you are presented with an option allowing you to update your contigs based on the mapping. If you enable this option, the areas of your assembled contigs to which no reads mapped back will be cut out of the final contigs. Indeed, no reads mapping back to a region of a contig means that there is no evidence in the data that this region exists. Note that it is always possible to retrieve contig sequences from mapping objects afterwards.

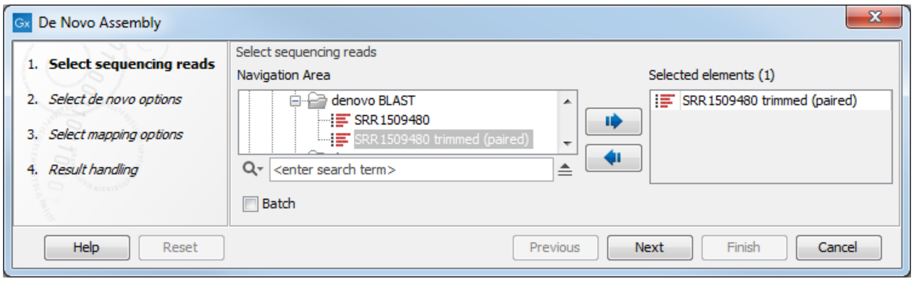
Sequence lists, like a list of contigs, can be used for different types of tasks, such as searching for patterns, or BLAST searching. In this tutorial, we will first generate a sequence list of contig sequences that we will use for BLAST. In the last and optional section of the tutorial, we will generate a mapping output instead, from which we will extract contig sequences for downstream use.

1. Start up a de novo assembly analysis by going to:

   Toolbox | **De Novo Sequencing ( )** | **De Novo Assembly ( )**

   This shows the dialog in figure 2.
2. Select the `SRR1509480 trimmed (paired)` ( ) sequence list and add it to the panel to the right. Click **Next**.
3. You now have the opportunity to set some parameters for your assembly. Here, we will accept the defaults, as shown in figure 3, by clicking **Next**.

   You can directly access the manual to find out more about these parameters by clicking on the **Help** button of the wizard window.



De Novo Assembly

1. **Select sequencing reads**
2. *Select de novo options*
3. *Select mapping options*
4. *Result handling*

Select sequencing reads

Navigation Area

- denovo BLAST
  - SRR1509480
  - SRR1509480 trimmed (paired)

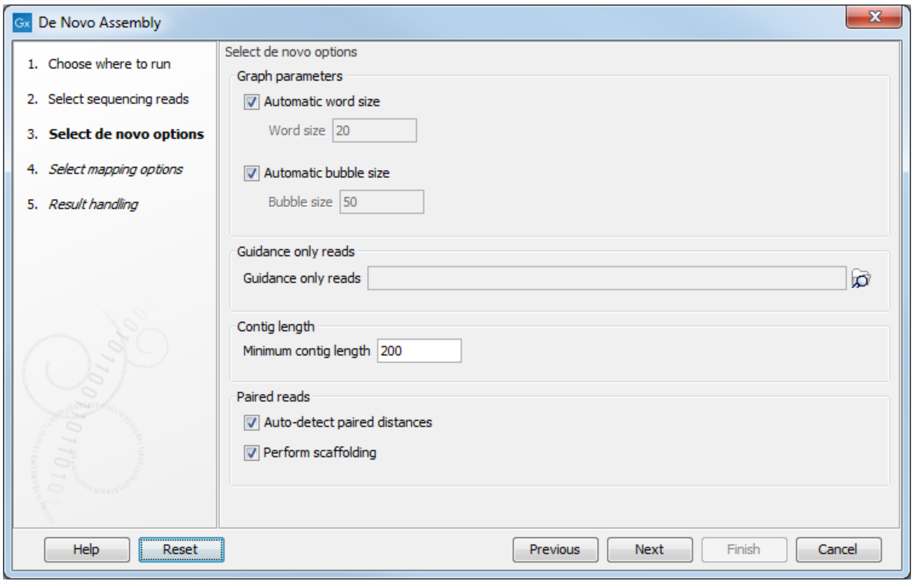
<enter search term>

□ Batch

Selected elements (1)

SRR1509480 trimmed (paired)

Figure 2: *Select the sequence list containing the reads.*



De Novo Assembly

1. Choose where to run
2. Select sequencing reads
3. **Select de novo options**
4. *Select mapping options*
5. *Result handling*

Select de novo options

Graph parameters

☑ Automatic word size

Word size 20

☑ Automatic bubble size

Bubble size 50

Guidance only reads

Guidance only reads

Contig length

Minimum contig length 200

Paired reads

☑ Auto-detect paired distances

☑ Perform scaffolding

Figure 3: *De novo assembly parameter options.*

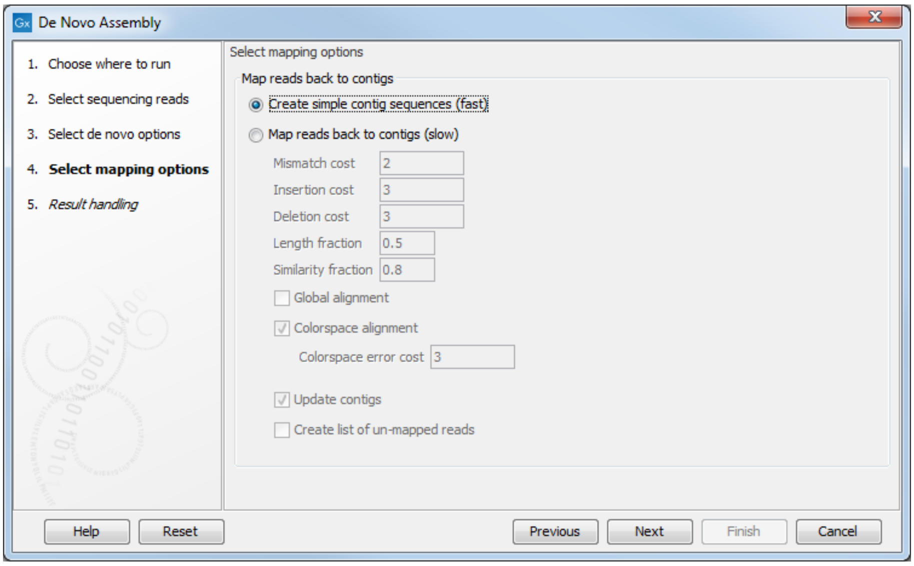
4. On this next page of the wizard, you get to choose the type of output you will generate. As mentioned earlier, we will choose to create a **simple contig output**, i.e., a sequence list of the contig sequences assembled during this job. Check the option **Create simple contig sequences (fast)**, as shown in figure 4. Note that this causes all other options on this wizard page to be grayed out as they pertain to mapping reads back to the contigs. Click **Next**.

5. Choose to create a report and to Save the results. Select the location where you want to save the assembly output (your tutorial folder for example) and click on **Finish**.

The assembly should be quite fast, but this depends on your computer. You can follow the progress in the Processes panel of the workbench. Note that the progress bar does not progress smoothly. It gives an indicator of the stage of the assembly more than an idea of the time left to finish. For large assemblies, you may notice that the progress bar may spend a lot of time at a certain percentage. This means it is working on a certain stage of the assembly. You can see a brief text description of the stage it is working on written in brackets above the progress bar (see figure 5).

Investigate the results

The outputs of the assembly are a sequence list and a report. If you have not changed the name of the sample data, the contig list will be called **SRR1509480 trimmed (paired) contig list**. The



De Novo Assembly

1. Choose where to run
2. Select sequencing reads
3. Select de novo options
4. **Select mapping options**
5. *Result handling*

Select mapping options

Map reads back to contigs

● Create simple contig sequences (fast)

○ Map reads back to contigs (slow)

Mismatch cost 2

Insertion cost 3

Deletion cost 3

Length fraction 0.5

Similarity fraction 0.8

□ Global alignment

☑ Colorspace alignment

Colorspace error cost 3

☑ Update contigs

□ Create list of un-mapped reads

Figure 4: *De novo assembly output types. Here we choose to create simple contigs. Note that if you choose to map reads back to the contigs, a read mapping job will be launched directly after the assembly task itself is done, and the outputs generated would be mapping objects.*

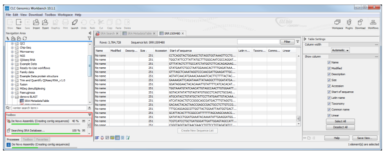


Figure 5: *You can monitor the progress of an assembly using the Processes tab of the Toolbox in the workbench.*

report will be called **SRR1509480 trimmed (paired) assembly summary report**.

1. Double click on the report object to open it. You will find general summary information in it such as the nucleotide distribution, information on contig lengths, as well as the N25, N50 and N75 values.

2. Double click on the contig list the tool has generated. This opens it in the View Area.

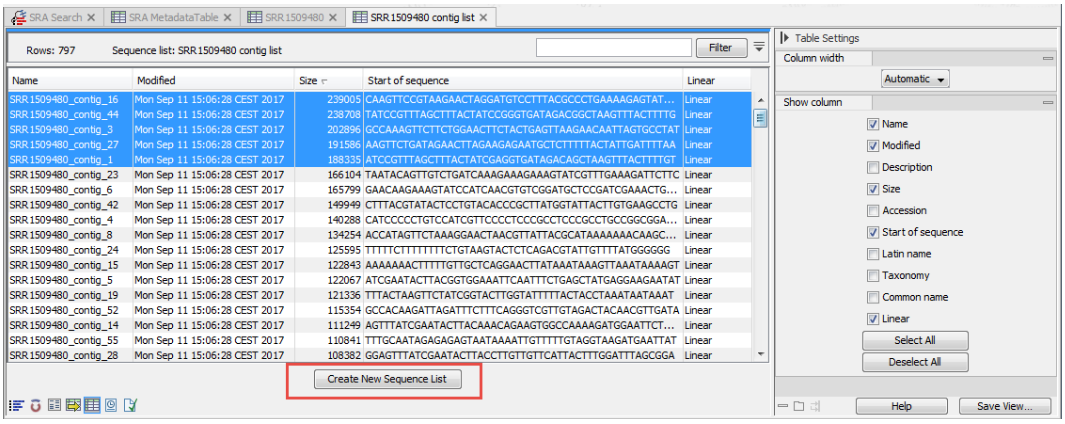
3. Click on the **table view icon** (▦) at the bottom of the workspace. You can see how many contigs were generated by how many rows are reported in the table.

4. Click on the **Size** column heading twice to sort the table according to the length of the contigs. Highlight the top five rows - here, the longest contigs - as shown in figure 6.

5. Click **Create a New Sequence List** just below the table. This new sequence list opens in a different tab in the View Area of the workbench.

6. Save this new sequence list. You can do this in a number of ways:

   - Right click on the tab at the top of the new view, and choose **Save As...** from the menu that appears, or
   - Drag the tab over into a folder in the **Navigation Area** of the Workbench. The data will then be saved into that folder.



Rows: 797 Sequence list: SRR1509480 contig list

Name	Modified	Size	Start of sequence	Linear
SRR1509480_contig_16	Mon Sep 11 15:06:28 CEST 2017	239005	CAAGTTCCGTAAGAACTAGGATGTCCTTTACGCCCTGAAAAGAGTAT...	Linear
SRR1509480_contig_44	Mon Sep 11 15:06:28 CEST 2017	238708	TATCCGTTTAGCTTTACTATCCGGGTGATAGACGGCTAAGTTTACTTTTG	Linear
SRR1509480_contig_3	Mon Sep 11 15:06:28 CEST 2017	202896	GCCAAAGTTCTTCTGGAACTTCTACTGAGTTAAGAACAATTAGTGCCTAT	Linear
SRR1509480_contig_27	Mon Sep 11 15:06:28 CEST 2017	191586	AAGTTCTGATAGAACTTAGAAGAGAATGCTCTTTTTACTATTGATTTTAA	Linear
SRR1509480_contig_1	Mon Sep 11 15:06:28 CEST 2017	188335	ATCCGTTTAGCTTTACTATCGAGGTGATAGACAGCTAAGTTTACTTTTGT	Linear
SRR1509480_contig_23	Mon Sep 11 15:06:28 CEST 2017	166104	TAATACAGTTGTCTGATCAAAGAAAGAAAGTATCGTTTGAAAGATTCTTC	Linear
SRR1509480_contig_6	Mon Sep 11 15:06:28 CEST 2017	165799	GAACAAGAAAGTATCCATCAACGTGTCGGATGCTCCGATCGAAACTG...	Linear
SRR1509480_contig_42	Mon Sep 11 15:06:28 CEST 2017	149949	CTTTACGTATACTCCTGTACACCCGCTTATGGTATTACTTGTGAAGCCTG	Linear
SRR1509480_contig_4	Mon Sep 11 15:06:28 CEST 2017	140288	CATCCCCCTGTCCATCGTTCCCCTCCCGCCTCCCGCCTGCCGGCGGA...	Linear
SRR1509480_contig_8	Mon Sep 11 15:06:28 CEST 2017	134254	ACCATAGTTCTAAAGGAACTAACGTTATTACGCATAAAAAAACAAGC...	Linear
SRR1509480_contig_24	Mon Sep 11 15:06:28 CEST 2017	125595	TTTTTCTTTTTTTTCTGTAAGTACTCTCAGACGTATTGTTTTATGGGGGG	Linear
SRR1509480_contig_15	Mon Sep 11 15:06:28 CEST 2017	122843	AAAAAAACTTTTTGTTGCTCAGGAACTTATAAATAAAGTTAAATAAAAGT	Linear
SRR1509480_contig_5	Mon Sep 11 15:06:28 CEST 2017	122067	ATCGAATACTTACGGTGGAAATTCAATTTCTGAGCTATGAGGAAGAATAT	Linear
SRR1509480_contig_19	Mon Sep 11 15:06:28 CEST 2017	121336	TTTACTAAGTTCTATCGGTACTTGGTATTTTTACTACCTAAATAATAAAT	Linear
SRR1509480_contig_52	Mon Sep 11 15:06:28 CEST 2017	115354	GCCACAAGATTAGATTTCTTTCAGGGTCGTTGTAGACTACAACGTTGATA	Linear
SRR1509480_contig_14	Mon Sep 11 15:06:28 CEST 2017	111249	AGTTTATCGAATACTTACAAACAGAAGTGGCCAAAAGATGGAATTCT...	Linear
SRR1509480_contig_55	Mon Sep 11 15:06:28 CEST 2017	110841	TTTGCAATAGAGAGAGTAATAAAATTGTTTTTGTAGGTAAGATGAATTAT	Linear
SRR1509480_contig_28	Mon Sep 11 15:06:28 CEST 2017	108382	GGAGTTTATCGAATACTTACCTTGTTGTTCATTACTTTGGATTTAGCGGA	Linear

Table Settings

Column width: Automatic

Show column

- ☑ Name
- ☑ Modified
- □ Description
- ☑ Size
- □ Accession
- ☑ Start of sequence
- □ Latin name
- □ Taxonomy
- □ Common name
- ☑ Linear

Select All

Deselect All

Figure 6: *The table view of the sequence list, with the top five (longest) contigs highlighted.*

Note: If you wish to shorten the time it will take to BLAST the new sequence list, choose less and/or smaller contigs to create the new sequence list.

BLAST contigs against NCBI

One way to investigate the contigs would be to see if there are any similar sequences in public databases. This can be done using the BLAST program suite. In this tutorial we will run a BLAST search of the five contigs we just saved into a new sequence list. Note that you need to be connected to the internet to do this part of the tutorial.

1. Start up the BLAST tool in the Workbench by going to:

   Toolbox | **BLAST** ( ) | **BLAST at NCBI** ( )

2. Select the sequence list subset with the 5 contig sequences that you just saved and click **Next**.

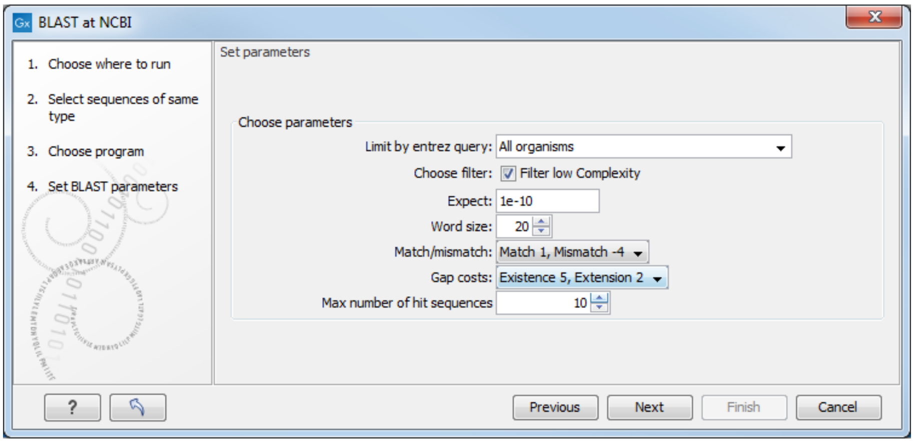
3. Choose **blastn** as the program type to run, and the **Nucleotide collection (nr)** as the database you wish to search. Click **Next**.

4. On this next page of the wizard, you can select the parameters you wish the blast search to be run with. Access a detailed explanation for the parameters by clicking on the **Help** button.

5. Change the settings of the parameters so they match those in figure 7. Here, we set a longer word size, and a much smaller e-value than the default. We have also limited the maximum number of matches returned to us to 10. The sequences we have sent to the NCBI are quite long, so the search will take a little while.

6. Choose to **Save** your results, click on **Next**, choose a place to save your results to and then click on the button **Finish**.

When the Blast search is finished, open the resulting data object. It should be called **Multi BLAST (5 sequences)**. The table contains a brief overview of the results for each of these query sequences. Look to the right of the table, in the View settings panel. You have many choices as to which columns you wish to see in this table. In figure 8, the "Description" column for the hit with the lowest e-value has been selected for display, in addition to the options checked by



Limit by entrez query: All organisms

Choose filter: ☑ Filter low Complexity

Expect: 1e-10

Word size: 20

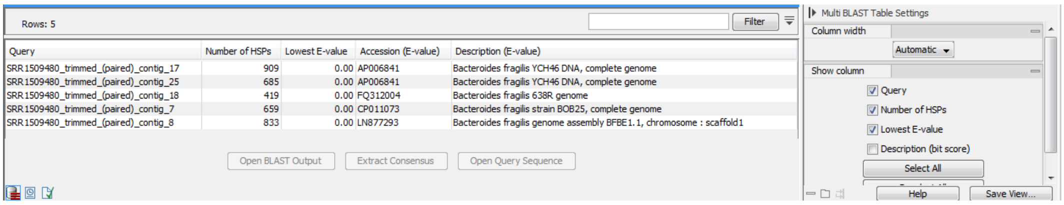
Match/mismatch: Match 1, Mismatch -4

Gap costs: Existence 5, Extension 2

Max number of hit sequences 10

Figure 7: *Setting parameters for a BLAST search.*

default. Note that you can choose to display your top hit according to "Description (e-value)" and/or according to "Description (identity)".



Rows: 5

Query	Number of HSPs	Lowest E-value	Accession (E-value)	Description (E-value)
SRR1509480_trimmed_(paired)_contig_17	909	0.00	AP006841	Bacteroides fragilis YCH46 DNA, complete genome
SRR1509480_trimmed_(paired)_contig_25	685	0.00	AP006841	Bacteroides fragilis YCH46 DNA, complete genome
SRR1509480_trimmed_(paired)_contig_18	419	0.00	FQ312004	Bacteroides fragilis 638R genome
SRR1509480_trimmed_(paired)_contig_7	659	0.00	CP011073	Bacteroides fragilis strain BOB25, complete genome
SRR1509480_trimmed_(paired)_contig_8	833	0.00	LN877293	Bacteroides fragilis genome assembly BFBE1.1, chromosome : scaffold1

Column width: Automatic

Show column:
☑ Query
☑ Number of HSPs
☑ Lowest E-value
☐ Description (bit score)

Figure 8: *The overview BLAST result table.*

For each contig, you can see the description of the top hit which is *Bacteroides fragilis* in each case, although of different strains.

You can, of course, export this summary table if you wish. Popular export formats for this sort of data would be Excel, or a comma separated values (csv) file.

Now double click on one of the rows of the summary blast table to open an individual BLAST result so you can investigate it in more detail.

By default, a visual depiction of the BLAST results will be shown. Click on the **table view icon** (▦) at the bottom of the workspace: with the table, one can filter and sort on different values. Like any table, you can export the data as Excel or comma separated values (csv). If you prefer to use the standard text blast format, click on the **Text Contents** (☰) icon at the bottom of the workspace.

BLAST contigs against local database

It is also possible to BLAST sequence list against a local database that only includes genes or elements of interest to your study. For this tutorial, we will use two resistance genes bexA and bexB that we will search for on NCBI:

1. Open the tool **Download** | **Search for sequences at NCBI...**

   And enter two search fields as shown on figure 9 (you can add the second field by clicking on the "Add search parameters" button).

2. Double click on the search result to open the sequence in split view.

3. In the side panel of the sequence view, make sure the annotation types "Genes" is checked.

4. Use the "Find" section of the side panel to look for the annotation of interest (the "Annotation" field must be checked). Click the "Find" button to highlight the annotation in the sequence view.

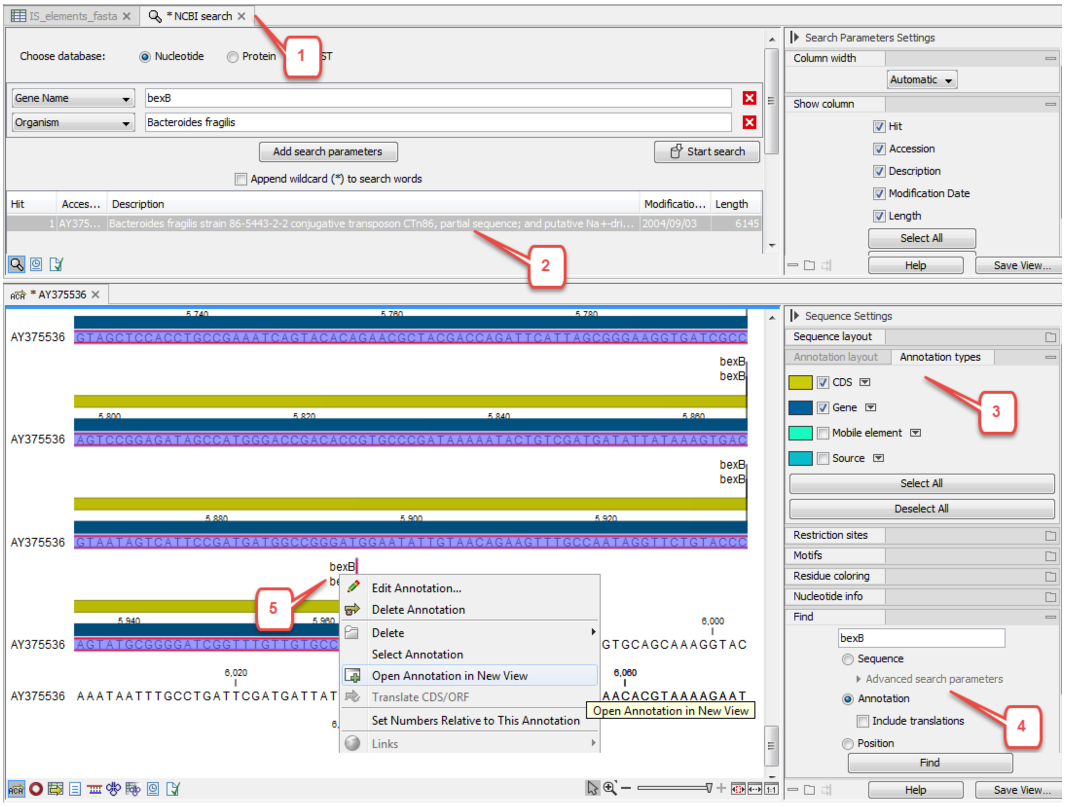


Figure 9: *Search at sequences at NCBI.*

5. Right-click the name of the annotation and choose to "Open Annotation in New View". You can then drag and drop the tab of the newly opened annotation to the tutorial folder in the Navigation Area to save the annotation.

Repeat the steps above for bexA and bexB. Once the two gene sequences are saved in the Navigation, go to:

Toolbox | **BLAST** () | **Create BLAST Database** ()

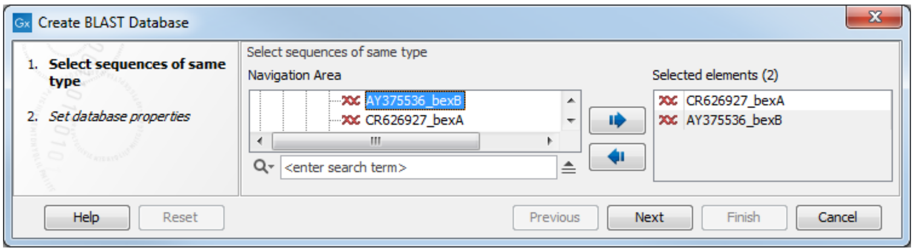
1. In the first dialog, select the two sequences you just saved from your NCBI search (figure 10) and click **Next**.

2. Set the database properties as illustrated in figure 11.

3. Click **Finish**.

4. Now launch the BLAST tool found here: **Toolbox** | **BLAST** () | **BLAST** ()

5. Select the sequence list subset with the 5 contig sequences and click **Next**.



Select sequences of same type

Navigation Area

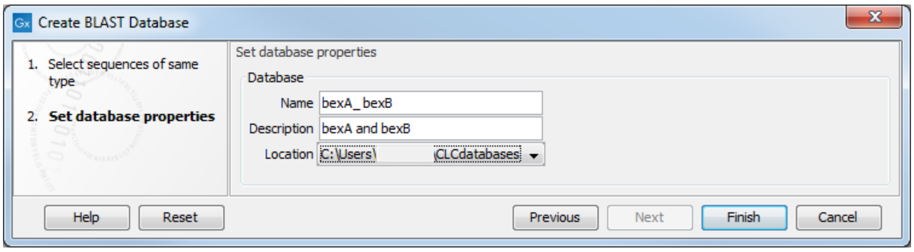
AY375536_bexB
CR626927_bexA

<enter search term>

Selected elements (2)

CR626927_bexA
AY375536_bexB

Figure 10: *Create a local BLAST database.*



Set database properties

Database

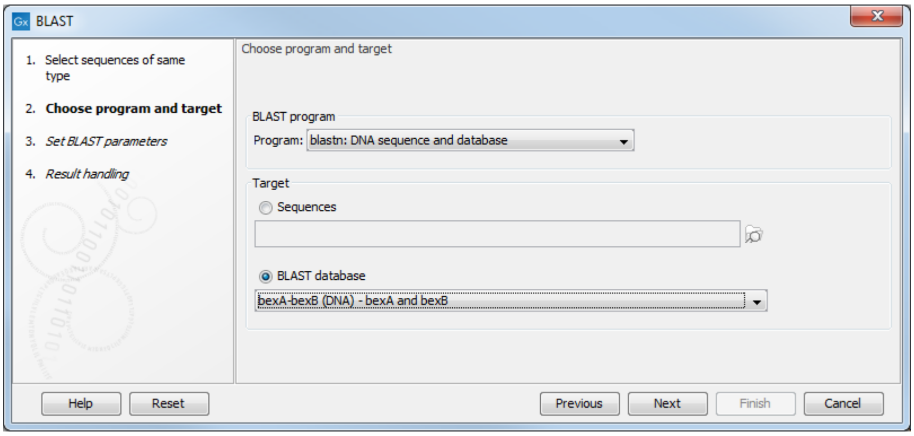
Name: bexA_ bexB

Description: bexA and bexB

Location: C:\Users\ CLCdatabases

Figure 11: *Set the local BLAST database properties.*

6. Select the BLAST program **bastn: DNA sequence and database** from the drop-down menu to be able to specify the local database you just created in the field below (figure 12).



Choose program and target

BLAST program

Program: blastn: DNA sequence and database

Target

○ Sequences

● BLAST database

bexA-bexB (DNA) - bexA and bexB

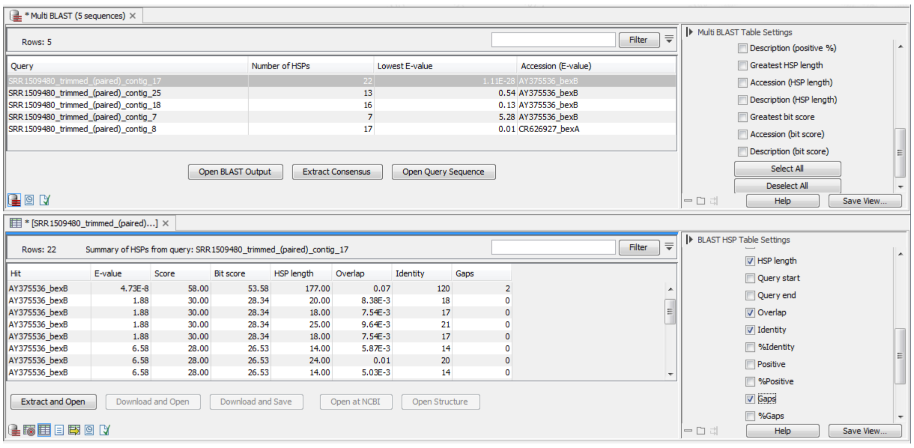
Figure 12: *Multi BLAST result table.*

7. Leave the BLAST parameters as they are set by default (you can always click on the **Reset** button to get back to default values).

8. Click **Finish** to start the tool.

The result is a Multi BLAST table. Double clicking on one of the row will open the BLAST output as seen in split view in figure 13.

Note that if you are interested in identifying antimicrobial resistance genes in a contig or a genome, you can use the Download Resistance Database followed by the Find Resistance with Nucleotide DB tool from the Microbial Genomics Module to obtain easily a resistance table. BLAST against a local database can however still be used to include genes or elements of interest that are not necessarily included in the database used by the module.



* Multi BLAST (5 sequences)

Rows: 5

Query	Number of HSPs	Lowest E-value	Accession (E-value)
SRR1509480_trimmed_(paired)_contig_17	22	1.11E-28	AY375536_bexB
SRR1509480_trimmed_(paired)_contig_25	13	0.54	AY375536_bexB
SRR1509480_trimmed_(paired)_contig_18	16	0.13	AY375536_bexB
SRR1509480_trimmed_(paired)_contig_7	7	5.28	AY375536_bexB
SRR1509480_trimmed_(paired)_contig_8	17	0.01	CR626927_bexA

Multi BLAST Table Settings

- □ Description (positive %)
- □ Greatest HSP length
- □ Accession (HSP length)
- □ Description (HSP length)
- □ Greatest bit score
- □ Accession (bit score)
- □ Description (bit score)

* [SRR1509480_trimmed_(paired)...]

Rows: 22 Summary of HSPs from query: SRR1509480_trimmed_(paired)_contig_17

Hit	E-value	Score	Bit score	HSP length	Overlap	Identity	Gaps
AY375536_bexB	4.73E-8	58.00	53.58	177.00	0.07	120	2
AY375536_bexB	1.88	30.00	28.34	20.00	8.38E-3	18	0
AY375536_bexB	1.88	30.00	28.34	18.00	7.54E-3	17	0
AY375536_bexB	1.88	30.00	28.34	25.00	9.64E-3	21	0
AY375536_bexB	1.88	30.00	28.34	18.00	7.54E-3	17	0
AY375536_bexB	6.58	28.00	26.53	14.00	5.87E-3	14	0
AY375536_bexB	6.58	28.00	26.53	24.00	0.01	20	0
AY375536_bexB	6.58	28.00	26.53	14.00	5.03E-3	14	0

BLAST HSP Table Settings

- ☑ HSP length
- □ Query start
- □ Query end
- ☑ Overlap
- ☑ Identity
- □ %Identity
- □ Positive
- □ %Positive
- ☑ Gaps
- □ %Gaps

Figure 13: *Multi BLAST result table.*

Choosing mapping as output type from an assembly

This is an optional part of the tutorial. Here, we re-run the de novo assembly, but this time choosing to map the reads back to the contigs, and to update the reads based on the mappings.

1. Start up a de novo assembly analysis by going to:

   Toolbox | **De Novo Sequencing** ( ) | **De Novo Assembly** ( )

2. Select the `SRR1509480 trimmed (paired)` ( ) sequence list as input by adding it to the panel to the right and click on **Next**.

3. You now have the opportunity to set some parameters for your assembly: accept the defaults, as shown in figure 3 and click on **Next**.

4. On this next page of the wizard, you get to choose the type of output you will generate. Here, choose **Map reads back to contigs (slow)** and leave all the mapping options as the defaults. Please refer to the manual by clicking on the **Help** button to find out more about these options. Also add a check in the box beside **Update contigs** before clicking on **Next**.

5. Choose to **Create a report** and to **Save** the results in your tutorial folder.

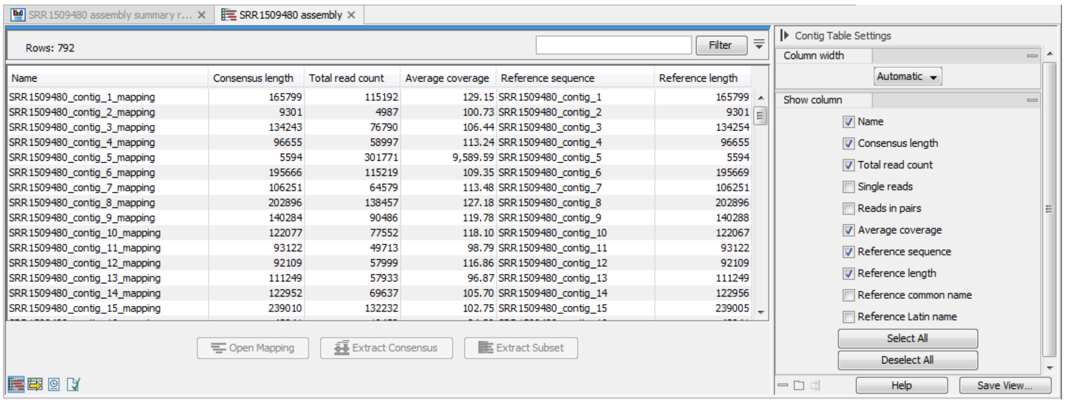
The outputs of the assembly are a summary table of the mappings (called **assembly**) and a report (called **assembly summary report-1**). There are a couple of extra graphs in this report compared to the one you generated earlier, where you requested only the contigs themselves as output.

1. Open the mapping output by double clicking on it in the **Navigation Area**. You should see a view like that shown in figure 14.

   Each row in the table represents a contig. Double-clicking a row will open the contig with the mapped reads.

2. Sort the mapping table on Consensus length and highlight the rows containing information about the five longest contigs.

3. Click on the button marked **Extract Contig** to get a sequence list with just these contigs. Choose to **Save** these.



Rows: 792

Name	Consensus length	Total read count	Average coverage	Reference sequence	Reference length
SRR1509480_contig_1_mapping	165799	115192	129.15	SRR1509480_contig_1	165799
SRR1509480_contig_2_mapping	9301	4987	100.73	SRR1509480_contig_2	9301
SRR1509480_contig_3_mapping	134243	76790	106.44	SRR1509480_contig_3	134254
SRR1509480_contig_4_mapping	96655	58997	113.24	SRR1509480_contig_4	96655
SRR1509480_contig_5_mapping	5594	301771	9,589.59	SRR1509480_contig_5	5594
SRR1509480_contig_6_mapping	195666	115219	109.35	SRR1509480_contig_6	195669
SRR1509480_contig_7_mapping	106251	64579	113.48	SRR1509480_contig_7	106251
SRR1509480_contig_8_mapping	202896	138457	127.18	SRR1509480_contig_8	202896
SRR1509480_contig_9_mapping	140284	90486	119.78	SRR1509480_contig_9	140288
SRR1509480_contig_10_mapping	122077	77552	118.10	SRR1509480_contig_10	122067
SRR1509480_contig_11_mapping	93122	49713	98.79	SRR1509480_contig_11	93122
SRR1509480_contig_12_mapping	92109	57999	116.86	SRR1509480_contig_12	92109
SRR1509480_contig_13_mapping	111249	57933	96.87	SRR1509480_contig_13	111249
SRR1509480_contig_14_mapping	122952	69637	105.70	SRR1509480_contig_14	122956
SRR1509480_contig_15_mapping	239010	132232	102.75	SRR1509480_contig_15	239005

Contig Table Settings

Column width: Automatic

Show column

- ☑ Name
- ☑ Consensus length
- ☑ Total read count
- □ Single reads
- □ Reads in pairs
- ☑ Average coverage
- ☑ Reference sequence
- ☑ Reference length
- □ Reference common name
- □ Reference Latin name

Figure 14: *A summary mapping table.*

Now you have a sequence list for your 5 longest contigs. They are similar to the contigs generated by the de novo assembly earlier, but have been updated by mapping the reads back to the contigs. From here, you could run a BLAST job, or any other task that requires a sequence list as input.

Bibliography

[Sydenham et al., 2015] Sydenham, T., Soki, J., Hasman, H., Wang, M., and Justesen, U. S. (2015). Identification of antimicrobial resistance genes in multidrug-resistant clinical bacteroides fragilis isolates by whole genome shotgun sequencing. *Anaerobe*.