Tutorial

De Novo Assembly of Paired Data

September 15, 2016

Sample to Insight

CLC bio, a QIAGEN Company · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark
Telephone: +45 70 22 32 44 · www.clcbio.com · support-clcbio@qiagen.com

De Novo Assembly of Paired Data

A de novo assembly involves taking many short sequences and trying to assemble them into longer, contiguous sequences. We recommend that you read about how the de novo assembly tool works before running this tool on your own data. You can read more in our manual http://clcsupport.com/clcgenomicsworkbench/current/index.php?manual=De_novo_sequencing.html or in the **White Paper**: http://www.clcbio.com/files/whitepapers/whitepaper-denovo-assembly-4.pdf.

This tutorial takes you through a typical *de novo* sequencing work flow with paired sequence data. We will

- Import sequence read data.
- Run a quality check on the read data.
- Trim the reads based on quality.
- Run a de novo assembly, including scaffolding.

We also include optional sections on finding broken pair mates in mapping results and exporting contigs from a set of mappings.

Import the data into the Workbench

We will use two Illumina read sets from *Pseudomonas aeruginosa*. One set is from a short insert (paired-end) library and the other from a jumping (mate-pair) library.

The data is available from the Short Read Archive with accession numbers SRR396636 and SRR396637. We have also made a copy of these datasets in fastq format available from the CLC bio website for the purposes of this tutorial.

To download the data:

1. Open up a web browser and go to the URL:

   http://download.clcbio.com/testdata/paeruginosa-reads.zip

2. Unzip this file, which is about 650Mb when compressed, and close to 2Gb when uncompressed, to a convenient area of your machine.

You should now have four fastq data files downloaded to your machine. We will first import the mate-pair data, and then the paired-end data.

Import the mate-pair data

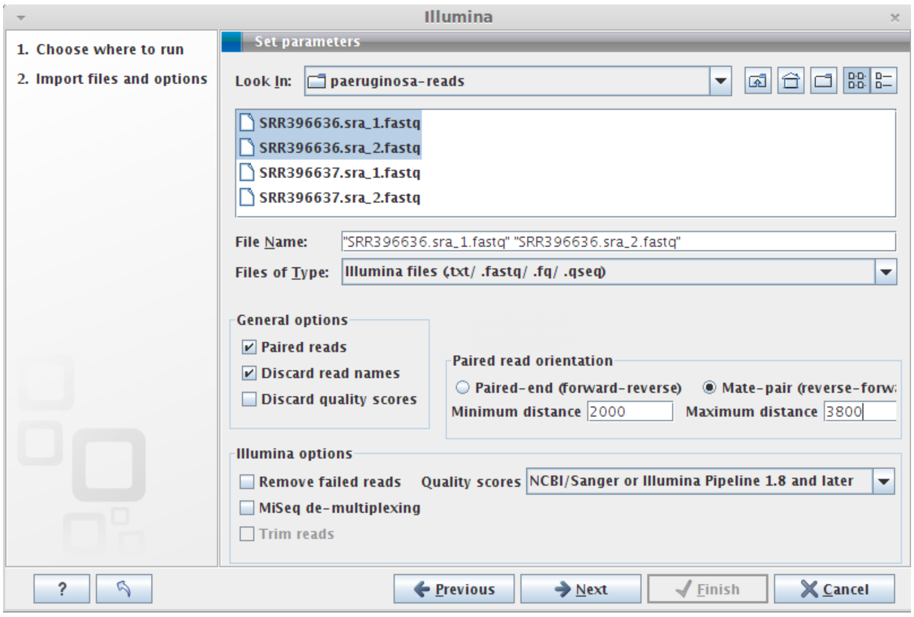
1. Start the *CLC Genomics Workbench* if you have not already, and go to the menu option:

   File | **Import** | **Illumina**

   This brings up a wizard where you specify information about your data, such as whether it is paired, whether quality values should be imported, and what Illumina pipeline was used when the data was generated. It is important to set the Illumina pipeline version correctly, as it affects how your quality scores are interpreted.

2. Set the options as shown in figure 1. That is:
   - Select the mate-pair data files: SRR396636.sra_1.fastq and SRR396636.sra_2.fastq
   - Under "General Options" section, ensure the **Paired reads** and **Discard read names** checkboxes are checked.
   - In the "Paired read orientation" section, ensure the **Mate-pair (reverse-forward)** option is checked.
   - Set the **Minimum distance** to 2000 and the **Maximum distance** to 3800.
   - Under **Quality scores** choose the *NCBI/Sanger or Illumina pipeline 1.8 or later* option.
3. Click on the button labeled **Next**.
4. Click on the button labeled **Save** in the wizard page that appears.
5. Click on the folder you wish to save to.

   You may wish to create a folder for saving the data and results of this tutorial. You can do this by clicking on the icon to add a data folder (📁) near the top of the Wizard window where you choose where to save your data.
6. Click on the button labeled **Finish**.



Look In: paeruginosa-reads

SRR396636.sra_1.fastq
SRR396636.sra_2.fastq
SRR396637.sra_1.fastq
SRR396637.sra_2.fastq

File Name: "SRR396636.sra_1.fastq" "SRR396636.sra_2.fastq"

Files of Type: Illumina files (.txt/ .fastq/ .fq/ .qseq)

General options
☑ Paired reads
☑ Discard read names
□ Discard quality scores

Paired read orientation
○ Paired-end (forward-reverse) ● Mate-pair (reverse-forw
Minimum distance 2000 Maximum distance 3800

Illumina options
□ Remove failed reads Quality scores NCBI/Sanger or Illumina Pipeline 1.8 and later
□ MiSeq de-multiplexing
□ Trim reads

Figure 1: *Importing mate-pair Illumina data.*

Import the paired-end data

1. Go to the menu option:

   File | **Import** | **Illumina**
2. Set the options as shown in figure 2. That is:
   - Select the paired-end data files: SRR396637.sra_1.fastq and SRR396637.sra_2.fastq

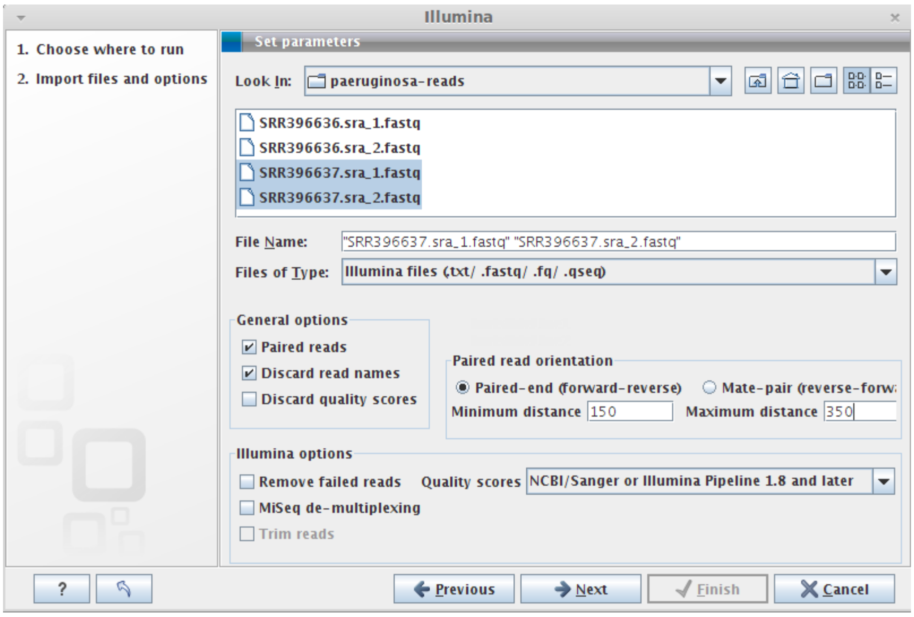
- Under "General Options" section, ensure the **Paired reads** and **Discard read names** checkboxes are checked.
- In the "Paired read orientation" section, ensure the **Paired-end (forward-reverse)** option is checked.
- Set the **Minimum distance** to 150 and the **Maximum distance** to 350.
- Under **Quality scores** choose the *NCBI/Sanger or Illumina pipeline 1.8 or later* option.

3. Click on the button labeled **Next**.

4. Click on the button labeled **Save** in the wizard page that appears.

5. Click on the folder you wish to save to.

6. Click on the button labeled **Finish**.



Look In: paeruginosa-reads

SRR396636.sra_1.fastq
SRR396636.sra_2.fastq
SRR396637.sra_1.fastq
SRR396637.sra_2.fastq

File Name: "SRR396637.sra_1.fastq" "SRR396637.sra_2.fastq"
Files of Type: Illumina files (.txt/ .fastq/ .fq/ .qseq)

General options
☑ Paired reads
☑ Discard read names
□ Discard quality scores

Paired read orientation
● Paired-end (forward-reverse) ○ Mate-pair (reverse-forw
Minimum distance 150 Maximum distance 350

Illumina options
□ Remove failed reads Quality scores NCBI/Sanger or Illumina Pipeline 1.8 and later
□ MiSeq de-multiplexing
□ Trim reads

Figure 2: *Importing paired-end Illumina data.*

Note that it is also possible to use the **Auto-detect paired distances** in the **De Novo Assembly** wizard step 3 to automatically determine the distances between paired-end or mate pair reads. This is particularly relevant in cases where you do not know the paired distances.

Tip: Importing large datasets can take some time. You can take a look to see if the import is still running by clicking on the tab called **Processes** in the bottom left side of the Workbench.

Once the import is completed, you should see two new files visible in the **Navigation Area** of the Workbench. By default, these will be called and SRR396636.sra_1 (paired) and SRR396637.sra_1 (paired).

Renaming the data objects

To help remember which dataset is which, please rename these data objects. In this case, we suggest you rename them to reflect the type of data is in each object.

1. Click on the name of SRR396636.sra_1 (paired) in the **Navigation Area** so it is highlighted.

2. Click again on the name. This should put it in a mode where you can edit it. (If not, try pressing the F2 key.)

3. Edit the name, changing it to *Mate-pair*.

4. Do the same steps for SRR396637.sra_1 (paired), but change it to *Paired-end*.

Sequencing Quality Analysis

We wish to use only high quality data in the de novo assembly. A sequencing quality analysis helps assess the quality of the datasets we are about to use. If the data contains lower quality regions, it should be trimmed, and the trimmed sequences used as input to the de novo assembly. If there are over-represented sequence motifs in the data, then it is worth checking if any adapters remain. If so, it is very important that these be trimmed away before using the data for de novo assembly.

Running a quality analysis

Here, we run the Sequence QC Report tool on both our sequence data objects. We will use the batch functionality, allowing us to simultaneously launch the analysis of multiple sets of data.

1. Go to:

   Toolbox | **NGS Core Tools ( )** | **Create Sequencing QC Report**

2. Check the box labeled **Batch** at the bottom of the Wizard window. Note that if the box labeled **Batch** is not checked, you will only be able to move data objects to the Selected Elements pane on the right, not folders, and you will generate only one report instead of one report per element.

3. Click on the folder containing your sequence data objects and then click the arrow pointing right. This will move the folder into the right hand pane as shown in figure 3. Click on the button labeled **Next**.

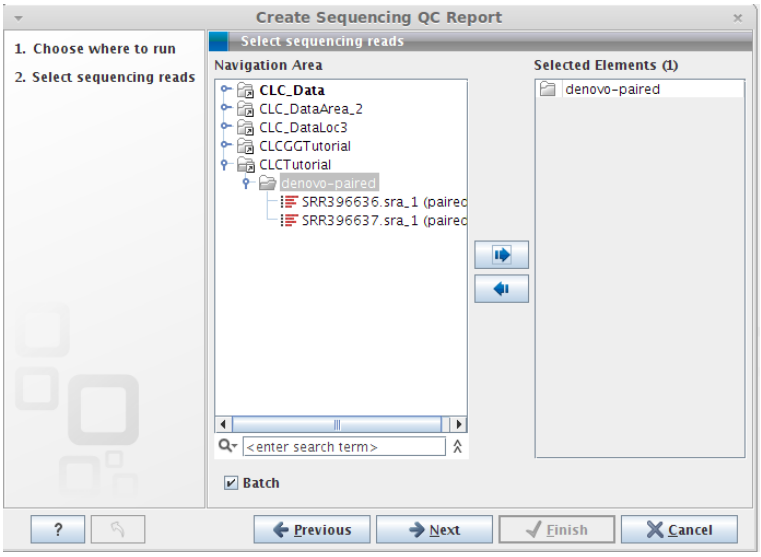


Figure 3: *Set up a batch job for running a quality check analysis.*

4. The next window allows you to fine tune which data objects the analysis should run on. If you click on either of the object names in the left hand side, the data object(s) to be worked on will be listed on the right hand side. Here, each folder contains only the data that we wish to work with, so we can just proceed to the next step.
5. Ensure that the **Quality analysis** and **Over-representation analysis** boxes are checked. Click on the button labeled **Next**.
6. Check the boxes labeled **Create graphical report** and **Create supplementary report**. Uncheck the box labeled **Create duplicated sequence list**. Check the box labeled **Save** to save the results before clicking on the button labeled **Finish**.

When both the quality check analyses are finished, you should see graphical and summary reports for each of your data sets listed in the **Navigation Area**.

Investigating the Sequencing Quality Check Reports

Here, we wish to look at the quality distribution for the reads and the sequence duplication information.

1. Open the file called **Mate-pair - graphical QC report** and take a look at the content.
2. Look at the **Quality distribution** plots in section 2.4 and 3.5 as shown in figure 4.

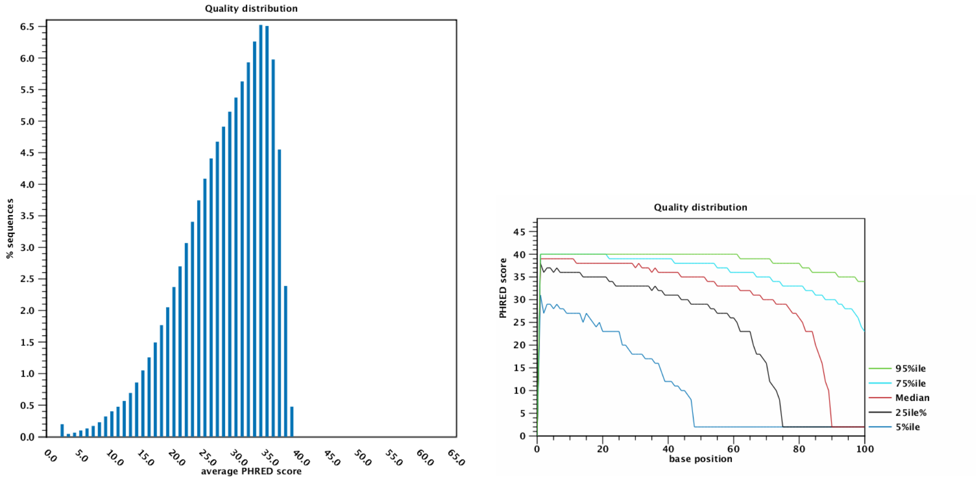


Figure 4: *Quality distributation for the mate-pair dataset.*

The overall distribution suggests that there is some lower quality data present, while the per-base plot shows that the data quality drops near the end of the sequences.

3. Look at the **Sequence duplication levels** plot in section 4.2. That plot does not suggest that sequence duplication is an issue here.

   There is also further information in the supplementary QC report file in the table called **Duplicated sequences** in section 4.3, if you suspect that sequence duplication might be an issue in your data set.
4. Open the file called **Paired-end - graphical QC report** and take a look at the plots relating to quality. The plots will be similar to those for the mate-pair data.

If you look at the information in the supplementary reports, you will see that there are some sequences that contain only ambiguous nucleotides. Such sequences will be removed during the trimming process, which we run next.

Trimming the data

Based on what we know of the data, we will trim low quality data away as well as removing sequences that contain too many ambiguous bases.

As we will be running the same trimming task on both our sets of data, we can again take advantage of the batch functionality of the Workbench.

1. Go to:

   Toolbox | **NGS Core Tools ()** | **Trim Sequences ()**

2. Check the box labeled **Batch** at the bottom of the Wizard window. Click on the folder containing your sequence data objects and then click the arrow pointing right. Click on the button labeled **Next**.

3. Here, we still have only our original sequence data objects that the trimming tool can work on. So if the trim adapter list box is empty, then you can just proceed by clicking on the button labeled **Next**. If there is an adapter list present from an earlier trimming job, please click on the reset button (↶) in the bottom left hand corner of the Wizard to remove it before clicking on **Next**.

4. Use the default settings, which set a quality score of 0.05 and a maximum number of ambiguous nucleotides of 2. Click on the button labeled **Next**.

5. The next step allows trimming of adapters. This is not necessary for this data, so click on the button labeled **Next**.

   If you have run an adapter trimming previously, and adapters appears in this Wizard step, please click on the small arrow icon (↶) in the bottom left of the Wizard window to clear the previous settings.

6. Click in the box labeled **Discard reads below length...** and set the number 15. Click on the button labeled **Next**.

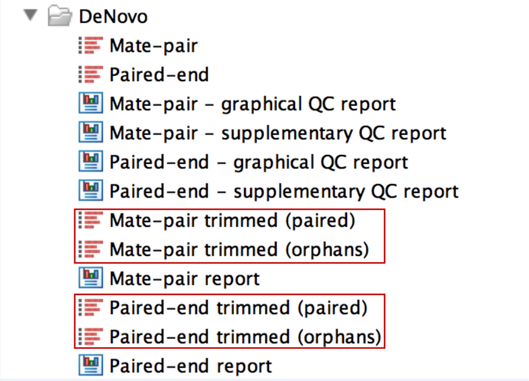
7. In the Output options section, choose to **Save broken pairs** and to **Create report**. **Save** the results and click on the button labeled **Finished**.

You should see four sequence lists resulting from the trimming process. These are the ones highlighted in figure 5. There are also two reports generated. These are called Mate-paired report and Paired-end report. We will now use the four trimmed sequence lists as input to a de novo assembly.

De novo assembly

Here we run a de novo assembly of the trimmed reads from the previous section.

There are two general types of output you can generate from the de novo assembly tool in the *CLC Genomics Workbench*:



DeNovo
- Mate-pair
- Paired-end
- Mate-pair – graphical QC report
- Mate-pair – supplementary QC report
- Paired-end – graphical QC report
- Paired-end – supplementary QC report
- Mate-pair trimmed (paired)
- Mate-pair trimmed (orphans)
- Mate-pair report
- Paired-end trimmed (paired)
- Paired-end trimmed (orphans)
- Paired-end report

Figure 5: *Four sequence list files are generated by the trimming tool, each with a name that includes the word: trimmed. The sequence lists with (paired) in their names contain paired sequences. The other trimmed lists contain single reads, where the mates of these reads were removed during the trimming process.*

- **Simple contigs** Here, the output is a sequence list of the contigs generated.
- **Stand-alone mappings** Here, a read mapping is carried out after the de novo assembly, where the sequence reads used for the assembly are mapped to the contigs that were assembled.

In this tutorial, we will choose to run a mapping of the reads to the assembled contigs when we set up the de novo assembly.

1. Go to:

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

   In the Wizard that starts up:

2. Select the four sequence lists that were generated by the trimming tool: **Mate-pair trimmed (paired)**, **Mate-pair trimmed (orphans)**, **Paired-end trimmed (paired)** and **Paired-end trimmed (orphans)** and click the arrow pointing right so they appear in the Selected Elements pane on the right. Click on the button labeled **Next**.

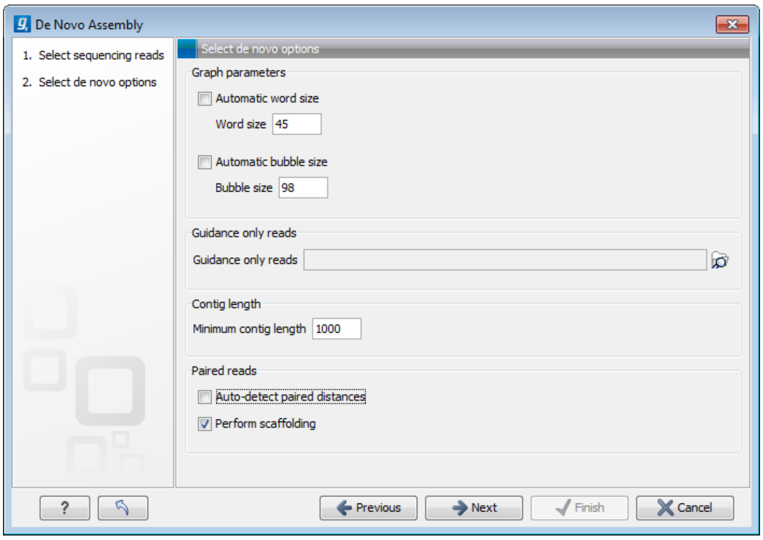
3. Set the parameters to use as shown in figure 6[1]. That is:

   - Uncheck the Automatic word size box and enter a value of 45.
   - Set the bubble size to 98.
   - Enter a Minimum contig length of 1000.
   - Leave the *Auto-detect paired distances* unchecked as we have already specified the known paired distances during import of the paired data.
   - Check the box labeled *Perform scaffolding*.

   Click on the button labeled **Next**.

4. In this next step, you choose the type of output to produce. Here we choose to generate mapping objects, i.e., we are choosing to map the read data back to the contigs produced by the de novo assembly.

[1] We have run some tests on this data, and think that the word size and bubble size chosen give reasonable results.



De Novo Assembly

1. Select sequencing reads
2. Select de novo options

Select de novo options

Graph parameters

□ Automatic word size

Word size 45

□ Automatic bubble size

Bubble size 98

Guidance only reads

Guidance only reads

Contig length

Minimum contig length 1000

Paired reads

□ Auto-detect paired distances

☑ Perform scaffolding

Figure 6: *Configure the parameters to be used for the de novo assembly.*

- Click on the radio button next to **Map reads back to contigs (slow)**. There are a number of parameter settings for the mapping of your reads back to the assembled contigs. Leave these set to the defaults.
- The checkboxes labeled **Global alignment**, **Update contigs**, and **Create list of unmapped reads** should all be unchecked for this tutorial. If you check the box labeled **Update contigs**, then areas of your contigs where no reads map will be cut out of the final contigs. The idea there is that if no reads map back to a region of a contig, then there is no evidence in the data that such a region exists. We choose not to do this in this tutorial.

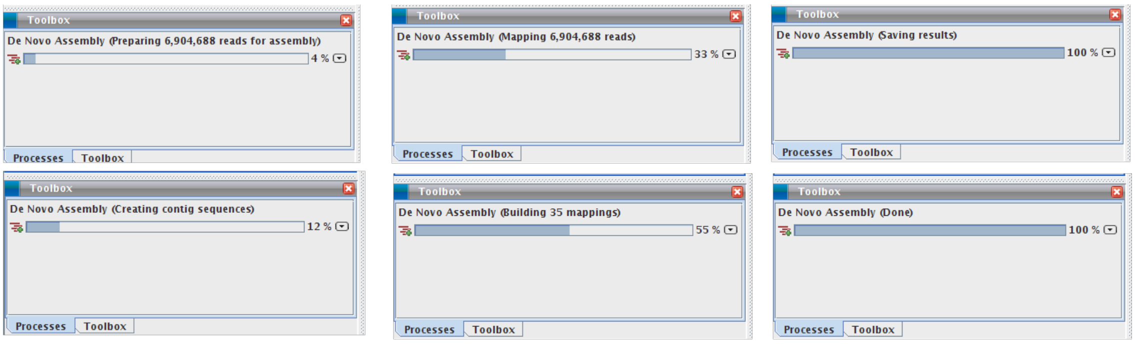
Click on the button labeled **Next**.

5. Choose to **Create report** and to **Save** the output. Click on the button labeled **Next**.

6. Click on the folder you wish to save the assembly outputs to before clicking on **Finish**.

Two analyses, which will run consecutively, have just been launched: a de novo assembly and a read mapping. Note that both can take some time. The length of time this assembly will take depends on the specifications of your machine. This task took between 4 and 5 minutes using the Genomics Workbench 6.5 on a machine with 8GB RAM, 4GB of which were allocated to the Workbench java process, and 4 cores, when no other memory hungry programs are running. You can monitor the progress of the analysis within the Processes tab in the bottom left hand side of the Workbench. The progress in percentage points will generally be quite uneven, as the progress bar provides information on the stage the task is on, rather than being a good indicator of the relative time taken or remaining for a task.

For multi-stage jobs such as de novo analysis and mappings, the text above the progress bar is a useful indicator of what stage the task has progressed to. Figure 7 shows what you might typically see when running a de novo analysis followed by a mapping of the reads back to the contigs.



De Novo Assembly (Preparing 6,904,688 reads for assembly) 4 %

De Novo Assembly (Mapping 6,904,688 reads) 33 %

De Novo Assembly (Saving results) 100 %

De Novo Assembly (Creating contig sequences) 12 %

De Novo Assembly (Building 35 mappings) 55 %

De Novo Assembly (Done) 100 %

Figure 7: *Phases reported in the progress tab for a de novo assembly where reads are also mapped back to the assembled contigs.*

Investigating the de novo assembly results

As you chose to map the reads back to the assembled contigs, the main output generated will be a summary read mapping object, which contains read mappings against each of the scaffolds that were generated during the de novo assembly stage of this analysis. This will be called something like **Mate-pair trimmed (paired) de novo assembly** with an icon (▤) beside it in the **Navigation Area**. You should also see a corresponding report file with a name like **Mate-pair trimmed (paired) summary report**.

1. Double click on the summary report object to open it.

   In the report is summary information such as the nucleotide distribution, information on contig length, as well as the N25, N50 and N75 values. You should see a table for the contig lengths with scaffolding and another table lower down for contigs without scaffolding.

2. Double click on the assembly, which has an icon beside it that looks like: (▤).

   This opens a table view in the Navigation pane of the Workbench that lists information about the mappings that were done.

3. Sort the table in descending order of consensus length by clicking on the column heading **Consensus Length** twice.

4. Double click on the row in the table for the longest contig, which is the top row now.

   This opens a mapping object, where the reference is longest contig from the de novo assembly.

5. Double click on the name of the mapping object within the tab in the viewing area.

   This expands the mapping to take up all the available viewing space. (To view the full Workbench again, double click on the name in the tab again.)

   We will not spend time focussing on the details of viewing mappings in this tutorial. Details about this topic can be found in our manual starting at:

   http://clcsupport.com/clcgenomicsworkbench/current/index.php?manual=View_settings_in_Side_Panel.html

6. Click on the **Show Annotation Table** icon (▤) at the bottom of the viewing area.

   The annotations tell you about certain actions carried out when determining the output of the de novo assembly. If you do not see any annotations, then please check that the options to show the annotation columns are checked in the right hand pane in the section labeled Shown annotation types.

There are three annotation types that can be reported by the de novo assembly tool:

- **Alternatives Excluded:** More than one path through the graph was possible in this region but evidence from paired data suggested the exclusion of one or more alternative routes in favour of the route chosen.
- **Contigs Joined:** More than one route was possible through the graph such that an unambiguous choice of how to traverse the graph cannot by made. However evidence from paired data supports one of these routes and on this basis, this route is followed to the exclusion of the other(s).
- **Scaffold:** The route through the graph is not clear but evidence from paired data supports the connection of two contigs. A single contig is then reported with N characters between the two connected regions. This entity is also known as a scaffold. The number of N characters represents the expected distance between the regions, based on the evidence the paired data.

Quite a number of scaffolding events took place when building the longest contig in this assembly. You can also see these annotations via the graphical view of the mapping.

7. Click on the **Show Read Mapping** icon (☰) at the bottom of the viewing area.

8. Click on the **Zoom Fit** button in the right lower corner to zoom out completely and see your whole read mapping within the viewing area.

9. In the right hand pane called **Read Mapping Settings**, expand the section called **Annotation types**. You can now check (or uncheck) the annotations you wish to see on the reference sequence. If you select all those available, you can see the same annotations you saw in the table view earlier.

   If you are interested in investigating certain regions containing particular annotations in more detail, you can zoom in by using the **Zoom in** and **Zoom out** buttons in the bottom toolbar.

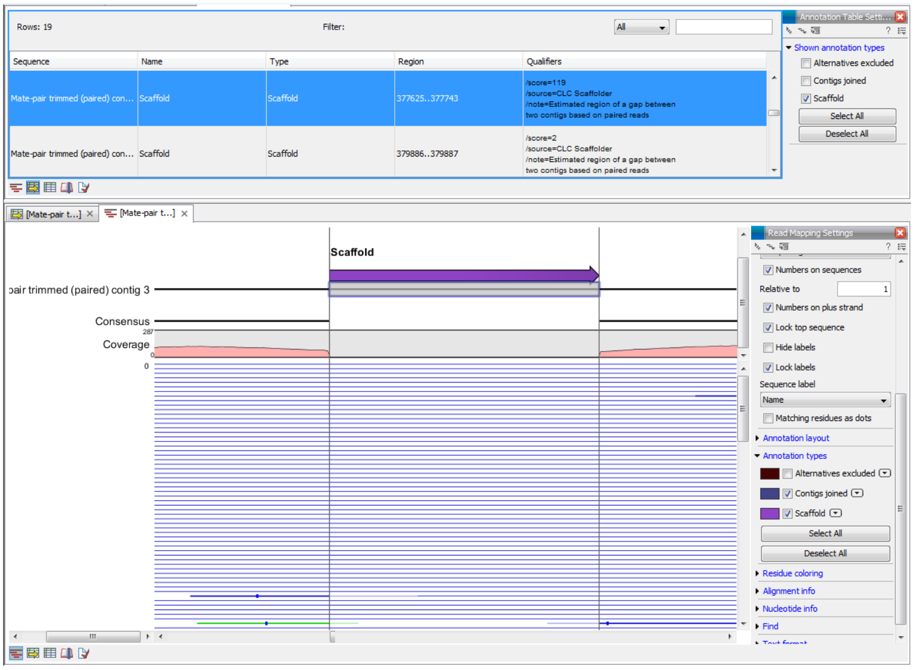
   Opening the table view of annotations as a linked view is often the easiest way to work with standard mapping objects and their annotations. When you click on a row in a table linked to a graphical view, the cursor moves to the relevant position within the mapping view. To open a linked table, you just need to depress the **Ctrl** key on your keyboard (**cmd** on Mac), and simultaneously use the mouse to click on the **Show Annotation Table** icon (▦) at the bottom of the viewing area.

   Now you should have the annotation table open as a linked view for the mapping object.

10. To view your reference sequence in detail, click on the mapping view, so it is the selected view, and then click on the "Zoom to base level" icon (1:1) in the bottom toolbar.

11. Now, in the annotation table view, double click on a row.

    You should see that the cursor jumps to the section of the mapping the row refers to. Figure 8 shows something similar to what you should be seeing, where the table above is linked to the mapping below.



Rows: 19 Filter: All

Sequence	Name	Type	Region	Qualifiers
Mate-pair trimmed (paired) con...	Scaffold	Scaffold	377625..377743	/score=119 /source=CLC Scaffolder /note=Estimated region of a gap between two contigs based on paired reads
Mate-pair trimmed (paired) con...	Scaffold	Scaffold	379886..379887	/score=2 /source=CLC Scaffolder /note=Estimated region of a gap between two contigs based on paired reads

Annotation Table Setti...

Shown annotation types

- ☐ Alternatives excluded
- ☐ Contigs joined
- ☑ Scaffold

Read Mapping Settings

- ☑ Numbers on sequences
- Relative to: 1
- ☑ Numbers on plus strand
- ☑ Lock top sequence
- ☐ Hide labels
- ☑ Lock labels
- Sequence label: Name
- ☐ Matching residues as dots

Annotation types

- ☐ Alternatives excluded
- ☑ Contigs joined
- ☑ Scaffold

Figure 8: *Double clicking on the row shown highlighted in the table above jumps the cursor to the point in the mapping view shown below. If you wish to, you can change the colors of the annotations displayed by just clicking in the coloured box next to each annotation type, and selecting a new color.*

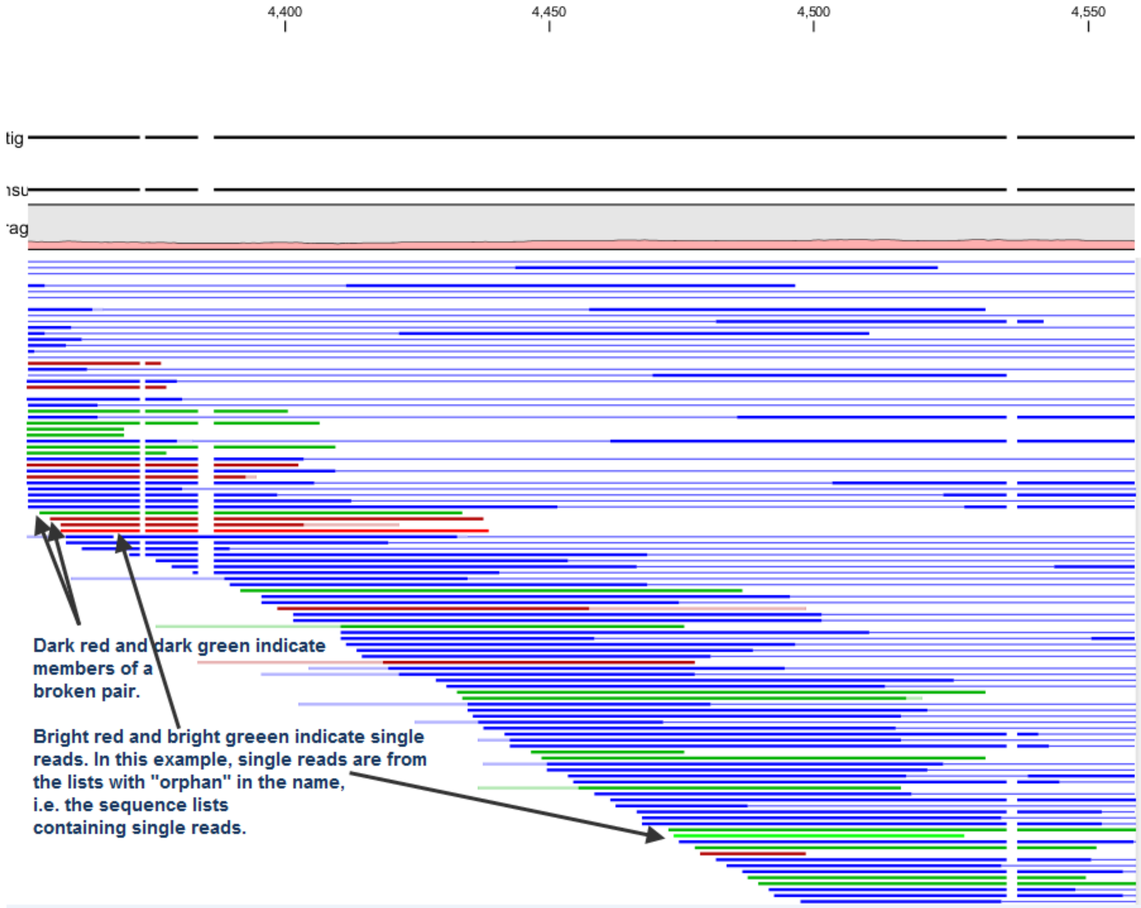
Note that the results of analyses carried out in the Workbench include history information. This can be useful if you need to recall what parameters you used or what version of the Workbench your analysis was run using. To access the history information, just click on the Show History (🕓) button at the bottom of the viewing area.

Finding broken pair mates

Reads mapped back to the contigs where both partners of a pair map in the correct relative orientation and within the expected distance range are coloured blue in the mapping and are called an intact pair. Those where only one member of the pair mapped, or those where both partners mapped, but in the wrong relative orientation or outside the expected distance range are considered members of a broken pair. Members of a broken pair that map to a unique location against the reference are colored green or red, according to whether they mapped in the forward or reverse orientation respectively.

To investigate where mapped mates of a broken pair are in the assembly:

1. Highlight a region of interest. For example, find a scaffold annotation where you see green and red coloured lines on either side of the gap, and highlight the region around this, as shown in figure 10.
2. Right click the mouse cursor over the highlighted region (underneath the annotation). Choose the option **Find Broken Pair Mates** as shown in figure 10.
3. Check the **Show overlapping annotations** box and select both annotation types present after clicking on the yellow arrow icon in the Wizard (figure 11). Note that a summary of the



Dark red and dark green indicate members of a broken pair.

Bright red and bright greeen indicate single reads. In this example, single reads are from the lists with "orphan" in the name, i.e. the sequence lists containing single reads.

Figure 9: *Read colors reveal information. Solid dark blue lines represent members of an intact pair. Light blue lines represent the connection between intact pair members. Green solid lines and red solid lines indicate single reads mapped in the forward direction and reverse direction relative to the reference respectively; dark green and dark red represent members of a broken pair, while bright red and bright green represent single reads. Pale green and pale red represent portions of a read that do not match the reference. Not shown here are reads mapping in yellow, meaning the read could map to multiple locations equally well.*

number of broken pairs in the region selected is also given in this step of the Wizard. Click on the button labeled **Next**.

4. Choose to **Open** the results and click on the button labeled **Finish**.

The table generated is linked to your mapping object. Open both in a linked view by going to the menu option:

View | **Split Horizonally** (⊟)

Alternatively, when the tab is the active tab, just press Ctrl-T on your keyboard. At this point, you can double click on rows in the linked table and jump to those locations where the mates of the broken pair reads from the selected region are mapped.

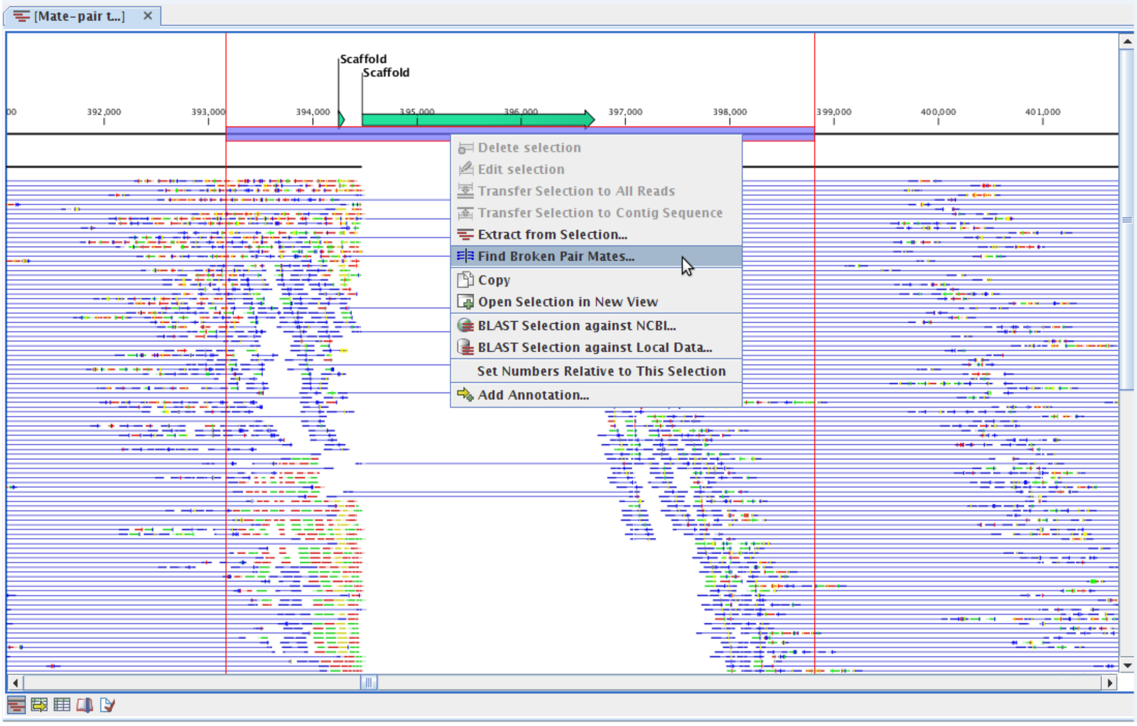
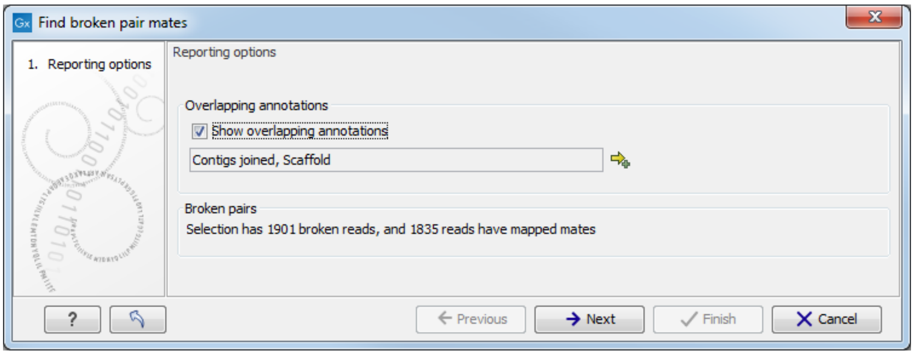

[Mate-pair t...]
Scaffold
Scaffold
392,000
393,000
394,000
395,000
396,000
397,000
398,000
399,000
400,000
401,000
Delete selection
Edit selection
Transfer Selection to All Reads
Transfer Selection to Contig Sequence
Extract from Selection...
Find Broken Pair Mates...
Copy
Open Selection in New View
BLAST Selection against NCBI...
BLAST Selection against Local Data...
Set Numbers Relative to This Selection
Add Annotation...

Figure 10: *Finding the pair mates for broken pair reads in a particular region.*



Overlapping annotations

☑ Show overlapping annotations

Contigs joined, Scaffold

Broken pairs

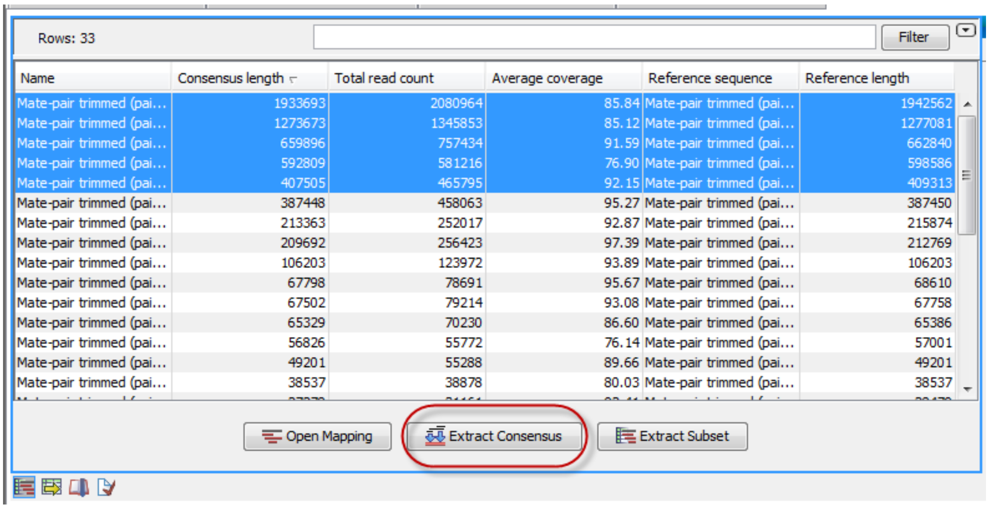
Selection has 1901 broken reads, and 1835 reads have mapped mates

Figure 11: *Choosing to include overlapping annotations.*

Exporting contigs from mapping results

You can extract the output from the de novo assembly stage as a sequence list if you wish. This can be useful when working with downstream programs that expect sequences as input, e.g., running a BLAST search, or if you wish to export just the contig sequences from the Workbench. Here, we go through the steps to export contig sequences from the mappings were generated.

1. Highlight the rows of the summary mapping table relating to the contigs you wish to extract. In figure 12, the 5 longest contigs have been selected.
2. Click on the button labeled **Extract Consensus** (figure 12).
3. Choose to save the output.



Rows: 33

Name	Consensus length	Total read count	Average coverage	Reference sequence	Reference length
Mate-pair trimmed (pai...	1933693	2080964	85.84	Mate-pair trimmed (pai...	1942562
Mate-pair trimmed (pai...	1273673	1345853	85.12	Mate-pair trimmed (pai...	1277081
Mate-pair trimmed (pai...	659896	757434	91.59	Mate-pair trimmed (pai...	662840
Mate-pair trimmed (pai...	592809	581216	76.90	Mate-pair trimmed (pai...	598586
Mate-pair trimmed (pai...	407505	465795	92.15	Mate-pair trimmed (pai...	409313
Mate-pair trimmed (pai...	387448	458063	95.27	Mate-pair trimmed (pai...	387450
Mate-pair trimmed (pai...	213363	252017	92.87	Mate-pair trimmed (pai...	215874
Mate-pair trimmed (pai...	209692	256423	97.39	Mate-pair trimmed (pai...	212769
Mate-pair trimmed (pai...	106203	123972	93.89	Mate-pair trimmed (pai...	106203
Mate-pair trimmed (pai...	67798	78691	95.67	Mate-pair trimmed (pai...	68610
Mate-pair trimmed (pai...	67502	79214	93.08	Mate-pair trimmed (pai...	67758
Mate-pair trimmed (pai...	65329	70230	86.60	Mate-pair trimmed (pai...	65386
Mate-pair trimmed (pai...	56826	55772	76.14	Mate-pair trimmed (pai...	57001
Mate-pair trimmed (pai...	49201	55288	89.66	Mate-pair trimmed (pai...	49201
Mate-pair trimmed (pai...	38537	38878	80.03	Mate-pair trimmed (pai...	38537

Figure 12: *Select the 5 longest contigs in the summary mapping table and then cick on the button labeled Extract Consensus.*