BIONUMERICS Tutorial:

Performing a hybrid de novo assembly on the cloud calculation engine

1 Aim

In this tutorial, we will perform a *hybrid* de novo assembly based on both short and long reads on the Cloud Calculation Engine.

A hybrid de novo assembly can only be performed on the Cloud Calculation Engine and not on your own computer due to the large amount of required resources.

2 Preparing the demo database

A de novo assembly on the Cloud Calculation Engine can only be performed after installation of the *WGS tools plugin* in the BIONUMERICS database (***File*** > ***Install / remove plugins...*** ()).

During installation of the plugin, make sure to select the options ***Use default Cloud Calculation Engine*** and ***Enable running jobs on Cloud Calculation Engine*** to unlock the full potential of the default Cloud Calculation Engine. Note that this installation procedure requires a password and a project name, linked to a certain amount of credits. Please contact Applied Maths to obtain more information.

The **WGS demo database** for *Escherichia coli* already contains the installed *WGS tools plugin* (but without any credits). It also contains sequence read set data links for 60 samples, calculated (short reads only) de novo assemblies and wgMLST results (allele calls and quality information).

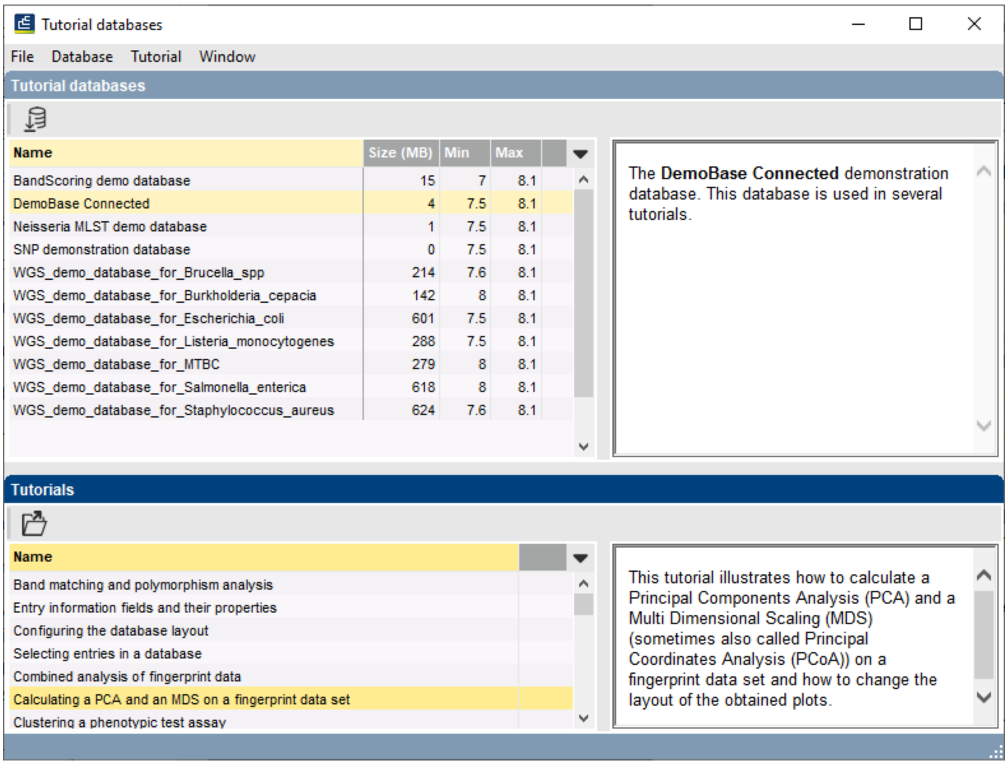
The short reads only *de novo* assembly, wgMLST workflow and results will not be discussed in this tutorial.

The **WGS_demo_database_for_Escherichia_coli** can be downloaded directly from the *BIONUMERICS Startup* window (see 2.1), or restored from the back-up file available on our website (see 2.2).

2.1 Option 1: Download demo database from the Startup Screen

1. Click the ↓ button, located in the toolbar in the *BIONUMERICS Startup* window.

This calls the *Tutorial databases* window (see Figure 1).



Name	Size (MB)	Min	Max
BandScoring demo database	15	7	8.1
DemoBase Connected	4	7.5	8.1
Neisseria MLST demo database	1	7.5	8.1
SNP demonstration database	0	7.5	8.1
WGS_demo_database_for_Brucella_spp	214	7.6	8.1
WGS_demo_database_for_Burkholderia_cepacia	142	8	8.1
WGS_demo_database_for_Escherichia_coli	601	7.5	8.1
WGS_demo_database_for_Listeria_monocytogenes	288	7.5	8.1
WGS_demo_database_for_MTBC	279	8	8.1
WGS_demo_database_for_Salmonella_enterica	618	8	8.1
WGS_demo_database_for_Staphylococcus_aureus	624	7.6	8.1

The **DemoBase Connected** demonstration database. This database is used in several tutorials.

Name
Band matching and polymorphism analysis
Entry information fields and their properties
Configuring the database layout
Selecting entries in a database
Combined analysis of fingerprint data
Calculating a PCA and an MDS on a fingerprint data set
Clustering a phenotypic test assay

This tutorial illustrates how to calculate a Principal Components Analysis (PCA) and a Multi Dimensional Scaling (MDS) (sometimes also called Principal Coordinates Analysis (PCoA)) on a fingerprint data set and how to change the layout of the obtained plots.

Figure 1: The *Tutorial databases* window, used to download the demonstration database.

2. Select **WGS_demo_database_for_Escherichia_coli** from the list and select ***Database*** > ***Download*** (🗄).

3. Confirm the installation of the database and press <***OK***> after successful installation of the database.

4. Close the *Tutorial databases* window with ***File*** > ***Exit***.

The **WGS_demo_database_for_Escherichia_coli** appears in the *BIONUMERICS Startup* window.

5. Double-click the **WGS_demo_database_for_Escherichia_coli** in the *BIONUMERICS Startup* window to open the database.

2.2 Option 2: Restore demo database from back-up file

A BIONUMERICS back-up file of the demo database for *Escherichia coli* is also available on our website. This backup can be restored to a functional database in BIONUMERICS.

6. Download the file `WGS_EC.bnbk` file from https://www.bionumerics.com/download/sample-data, under 'WGS_demo_database_for_Escherichia_coli'.

✎ In contrast to other browsers, some versions of Internet Explorer rename the `WGS_EC.bnbk` database backup file into `WGS_EC.zip`. If this happens, you should manually remove the `.zip` file extension and replace with `.bnbk`. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the `.zip` file extension if the option "Hide extensions for known file types" is checked in your Windows folder options.

7. In the *BIONUMERICS Startup* window, press the button. From the menu that appears, select ***Restore database...***.

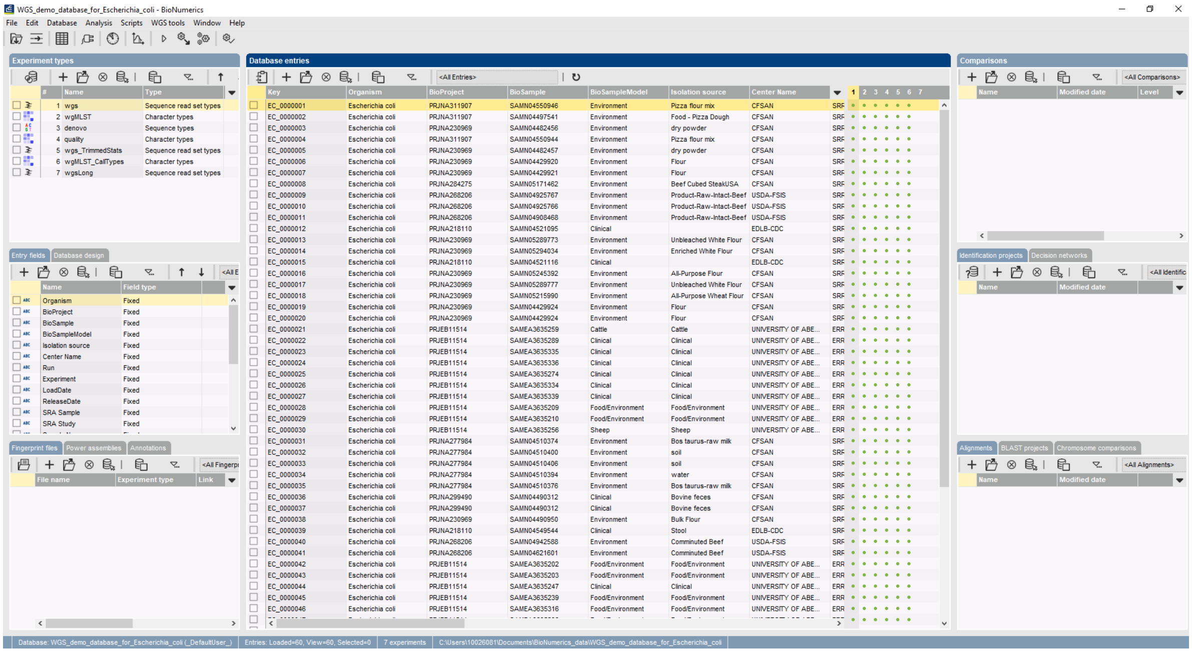
8. Browse for the downloaded file and select ***Create copy***. Note that, if ***Overwrite*** is selected, an existing database will be overwritten.

9. Specify a new name for this demonstration database, e.g. "WGS_Escherichia_coli_demobase".

10. Click <***OK***> to start restoring the database from the backup file.

11. Once the process is complete, click <***Yes***> to open the database.

The *Main* window is displayed (see Figure 2).



#	Name	Type
1	wgs	Sequence read set types
2	wgMLST	Character types
3	denovo	Sequence types
4	quality	Character types
5	wgs_TrimmedStats	Sequence read set types
6	wgMLST_CallTypes	Character types
7	wgsLong	Sequence read set types

Name	Field type
Organism	Fixed
BioProject	Fixed
BioSample	Fixed
BioSampleModel	Fixed
Isolation source	Fixed
Center Name	Fixed
Run	Fixed
Experiment	Fixed
LoadDate	Fixed
ReleaseDate	Fixed
SRA Sample	Fixed
SRA Study	Fixed

File name	Experiment type	Link

Key	Organism	BioProject	BioSample	BioSampleModel	Isolation source	Center Name		1	2	3	4	5	6	7
EC_0000001	Escherichia coli	PRJNA311907	SAMN04550946	Environment	Pizza flour mix	CFSAN	SRR	●	●	●	●	●	●	
EC_0000002	Escherichia coli	PRJNA311907	SAMN04497541	Environment	Food - Pizza Dough	CFSAN	SRR	●	●	●	●	●	●	
EC_0000003	Escherichia coli	PRJNA230969	SAMN04482456	Environment	dry powder	CFSAN	SRR	●	●	●	●	●	●	
EC_0000004	Escherichia coli	PRJNA311907	SAMN04550944	Environment	Pizza flour mix	CFSAN	SRR	●	●	●	●	●	●	
EC_0000005	Escherichia coli	PRJNA230969	SAMN04482457	Environment	dry powder	CFSAN	SRR	●	●	●	●	●	●	
EC_0000006	Escherichia coli	PRJNA230969	SAMN04429920	Environment	Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000007	Escherichia coli	PRJNA230969	SAMN04429921	Environment	Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000008	Escherichia coli	PRJNA284275	SAMN05171462	Environment	Beef Cubed SteakUSA	CFSAN	SRR	●	●	●	●	●	●	
EC_0000009	Escherichia coli	PRJNA268206	SAMN04925767	Environment	Product-Raw-Intact-Beef	USDA-FSIS	SRR	●	●	●	●	●	●	
EC_0000010	Escherichia coli	PRJNA268206	SAMN04925766	Environment	Product-Raw-Intact-Beef	USDA-FSIS	SRR	●	●	●	●	●	●	
EC_0000011	Escherichia coli	PRJNA268206	SAMN04908468	Environment	Product-Raw-Intact-Beef	USDA-FSIS	SRR	●	●	●	●	●	●	
EC_0000012	Escherichia coli	PRJNA218110	SAMN04521095	Clinical		EDLB-CDC	SRR	●	●	●	●	●	●	
EC_0000013	Escherichia coli	PRJNA230969	SAMN05289773	Environment	Unbleached White Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000014	Escherichia coli	PRJNA230969	SAMN05294034	Environment	Enriched White Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000015	Escherichia coli	PRJNA218110	SAMN04521116	Clinical		EDLB-CDC	SRR	●	●	●	●	●	●	
EC_0000016	Escherichia coli	PRJNA230969	SAMN05245392	Environment	All-Purpose Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000017	Escherichia coli	PRJNA230969	SAMN05289777	Environment	Unbleached White Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000018	Escherichia coli	PRJNA230969	SAMN05215990	Environment	All-Purpose Wheat Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000019	Escherichia coli	PRJNA230969	SAMN04429924	Environment	Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000020	Escherichia coli	PRJNA230969	SAMN04429924	Environment	Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000021	Escherichia coli	PRJEB11514	SAMEA3635259	Cattle	Cattle	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000022	Escherichia coli	PRJEB11514	SAMEA3635289	Clinical	Clinical	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000023	Escherichia coli	PRJEB11514	SAMEA3635335	Clinical	Clinical	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000024	Escherichia coli	PRJEB11514	SAMEA3635336	Clinical	Clinical	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000025	Escherichia coli	PRJEB11514	SAMEA3635274	Clinical	Clinical	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000026	Escherichia coli	PRJEB11514	SAMEA3635334	Clinical	Clinical	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000027	Escherichia coli	PRJEB11514	SAMEA3635339	Clinical	Clinical	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000028	Escherichia coli	PRJEB11514	SAMEA3635209	Food/Environment	Food/Environment	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000029	Escherichia coli	PRJEB11514	SAMEA3635210	Food/Environment	Food/Environment	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000030	Escherichia coli	PRJEB11514	SAMEA3635256	Sheep	Sheep	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000031	Escherichia coli	PRJNA277984	SAMN04510374	Environment	Bos taurus-raw milk	CFSAN	SRR	●	●	●	●	●	●	
EC_0000032	Escherichia coli	PRJNA277984	SAMN04510400	Environment	soil	CFSAN	SRR	●	●	●	●	●	●	
EC_0000033	Escherichia coli	PRJNA277984	SAMN04510406	Environment	soil	CFSAN	SRR	●	●	●	●	●	●	
EC_0000034	Escherichia coli	PRJNA277984	SAMN04510394	Environment	water	CFSAN	SRR	●	●	●	●	●	●	
EC_0000035	Escherichia coli	PRJNA277984	SAMN04510376	Environment	Bos taurus-raw milk	CFSAN	SRR	●	●	●	●	●	●	
EC_0000036	Escherichia coli	PRJNA299490	SAMN04490312	Clinical	Bovine feces	CFSAN	SRR	●	●	●	●	●	●	
EC_0000037	Escherichia coli	PRJNA299490	SAMN04490312	Clinical	Bovine feces	CFSAN	SRR	●	●	●	●	●	●	
EC_0000038	Escherichia coli	PRJNA230969	SAMN04490950	Environment	Bulk Flour	CFSAN	SRR	●	●	●	●	●	●	
EC_0000039	Escherichia coli	PRJNA218110	SAMN04549544	Clinical	Stool	EDLB-CDC	SRR	●	●	●	●	●	●	
EC_0000040	Escherichia coli	PRJNA268206	SAMN04942588	Environment	Comminuted Beef	USDA-FSIS	SRR	●	●	●	●	●	●	
EC_0000041	Escherichia coli	PRJNA268206	SAMN04621601	Environment	Comminuted Beef	USDA-FSIS	SRR	●	●	●	●	●	●	
EC_0000042	Escherichia coli	PRJEB11514	SAMEA3635202	Food/Environment	Food/Environment	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000043	Escherichia coli	PRJEB11514	SAMEA3635203	Food/Environment	Food/Environment	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000044	Escherichia coli	PRJEB11514	SAMEA3635247	Clinical	Clinical	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000045	Escherichia coli	PRJEB11514	SAMEA3635239	Food/Environment	Food/Environment	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	
EC_0000046	Escherichia coli	PRJEB11514	SAMEA3635316	Food/Environment	Food/Environment	UNIVERSITY OF ABE...	ERR	●	●	●	●	●	●	

Name	Modified date	Level

Name	Modified date

Name	Modified date

Database: WGS_demo_database_for_Escherichia_coli (_DefaultUser_) | Entries: Loaded=60, View=60, Selected=0 | 7 experiments | C:\Users\10026081\Documents\BioNumerics_data\WGS_demo_database_for_Escherichia_coli

Figure 2: The *Escherichia coli* demonstration database: the *Main* window.

3 Importing sequence read sets

You will be importing sequence read set data of *Escherichia coli* strain 397404 from two different sequencing runs performed by Public Health England.

- Short read data from sequencing run *SRR6001344* was generated by Illumina HiSeq 2500 whole genome sequencing.

- Long read data from sequencing run *SRR9987850* was generated by Oxford Nanopore minION whole genome sequencing.

To be able to perform a hybrid de novo assembly the sequencing data from both runs need to be linked to the same entry in the BIONUMERICS database. Therefore, you will first create a new entry in the database.

1. Open the **WGS_demo_database_for_Escherichia_coli** database which has the *WGS tools* plugin installed.

2. Create a new entry in the database by clicking + in the *Database entries* panel of the *Main* window.

3. Type “397404” in the **Database key** text field and click <***OK***>.

A new entry with key **397404** will be created in the database. For this entry you will need to specify the name of the sequencing runs. You will use these later to link the NCBI SRA data to the entry. You will use the existing entry field **Run** for the SRA accession of the short reads sequencing data and create a new entry field **RunLong** for the SRA accession of the long reads sequencing data.

4. Type “SRR6001344” in the entry field **Run** for entry **397404**.

5. Create a new entry field by clicking on + in the *Entry fields* panel.

6. Type “RunLong” in the **Name** text field and click <***OK***>.

7. Type “SRR9987850” in the newly created entry field **RunLong** for entry **397404**.

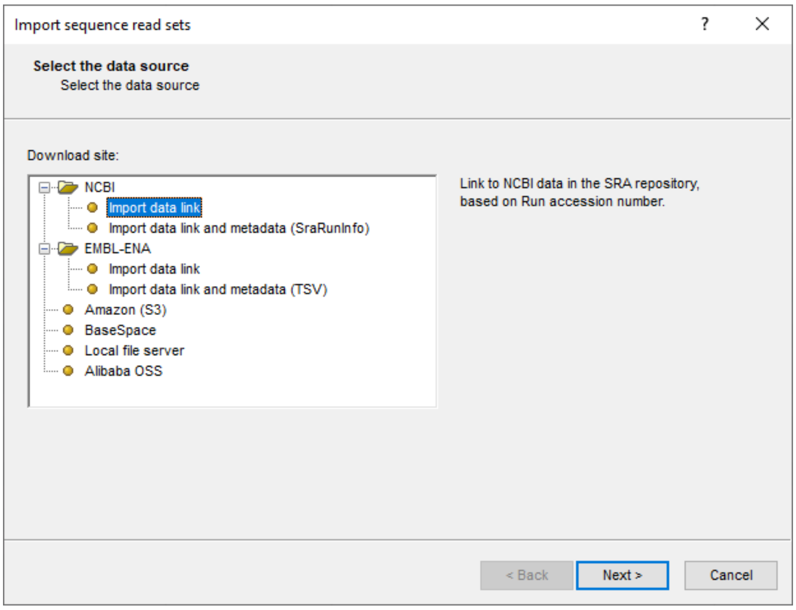
You will now import the sequence read set data and link it to the newly created entry.

8. Make sure entry **397404** is selected by clicking the ballot box in front of the entry’s **Key**.

9. Select ***File*** > ***Import...*** (📂, **Ctrl+I**) to open the *Import data* wizard.

10. Highlight the ***Import sequence read set data as links*** option and press <***Finish***>.

Links to multiple data sources are available, including online and offline data repositories such as: ***NCBI (SRA)***, ***EMBL-EBI (ENA)***, ***Amazon (S3)***, ***BaseSpace***, ***Alibaba OSS*** or ***Local file server*** (see Figure 3). Depending on the choice of import, different parameters may be queried in the next steps.



Link to NCBI data in the SRA repository, based on Run accession number.

Figure 3: Data sources.

In this tutorial, the import of FASTQ files as links from NCBI is covered. For more information about the other options, please consult the *WGS tools plugin* manual.

11. Select ***Import data link*** under **NCBI** as download site and press <***Next***> (see Figure 3).

You will **import the short reads files** first.

12. In the **Pick up accession codes from field** drop-down menu select the entry field **Run** and press <***Fetch***>.

The accession number from the entry field **Run** will now be added to the **Accession code(s)** field.

13. Select <***Next***> to go to the next step.

Now you need to define how the data should be stored in the database. The dialog box allows you to set import rules. For each import source (in this case we have only the accession code), a database destination can be specified.

You will link the accession code to the ***Run*** field.

14. Double-click the first row in the grid, and click the small plus ('+') sign in front of **Entry info field** and select ***Run*** from the list (see Figure 4).

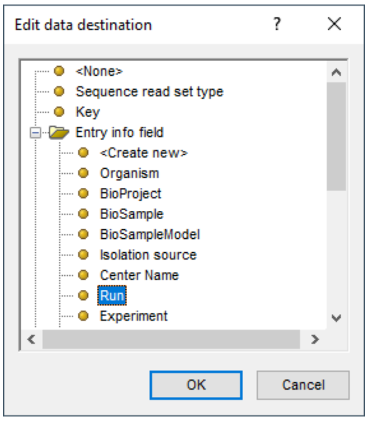


Figure 4: The **import rules** grid after assigning the file name as ***Key***.

15. Press <***OK***>.

The import rules are updated in the grid (see Figure 5).

16. Press <***Preview***> to see what you are about to import.

17. Press the <***Close***> button to close the preview and press <***Next***> to proceed to the *Import links* dialog box.

18. Make sure ***Run*** is checked as link and press <***Finish***>.

The import template needs to be saved to be able to use it again later on.

19. Enter a ***Name*** for the import template (e.g. "short reads") and optionally a ***Description***. Next, press <***OK***>.

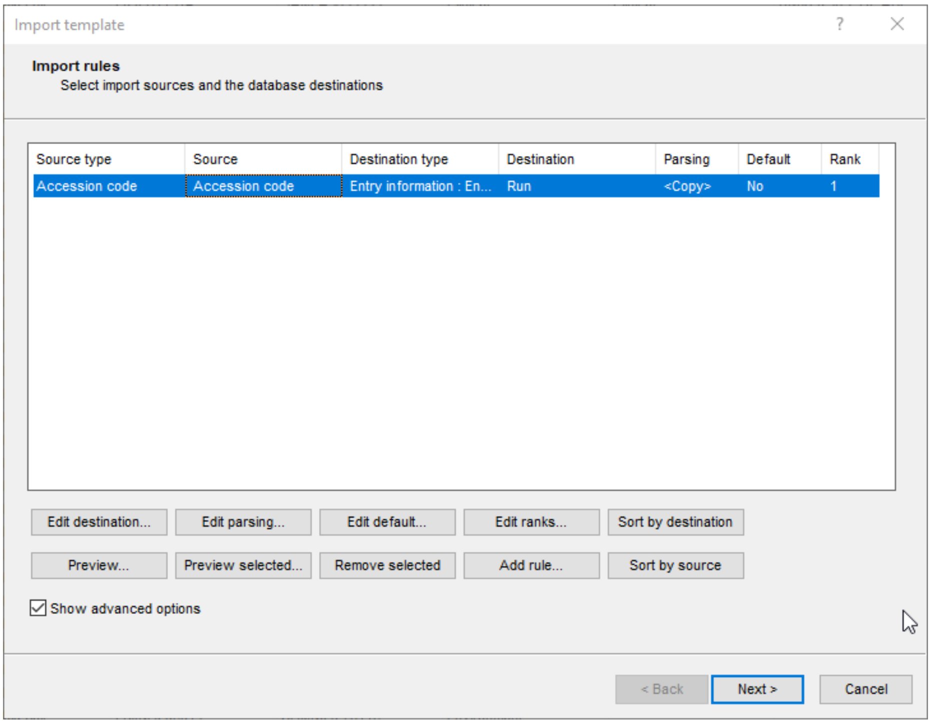
20. In the *Import template* dialog box, highlight the newly created template.

21. Make sure the ***wgs*** experiment is selected and click <***Next***>.

22. Press <***Next***> once more.

In the last step, calculation jobs (e.g. de novo assembly) can be launched on the imported data links (***Open submit jobs dialog after import***). Note that the same dialog can be called from the *Main* window at any time with ***WGS tools*** > ***Submit jobs...*** (▷).

23. Press <***Finish***> to start the import of the data links.



Import rules

Select import sources and the database destinations

Source type	Source	Destination type	Destination	Parsing	Default	Rank
Accession code	Accession code	Entry information : En...	Run	<Copy>	No	1

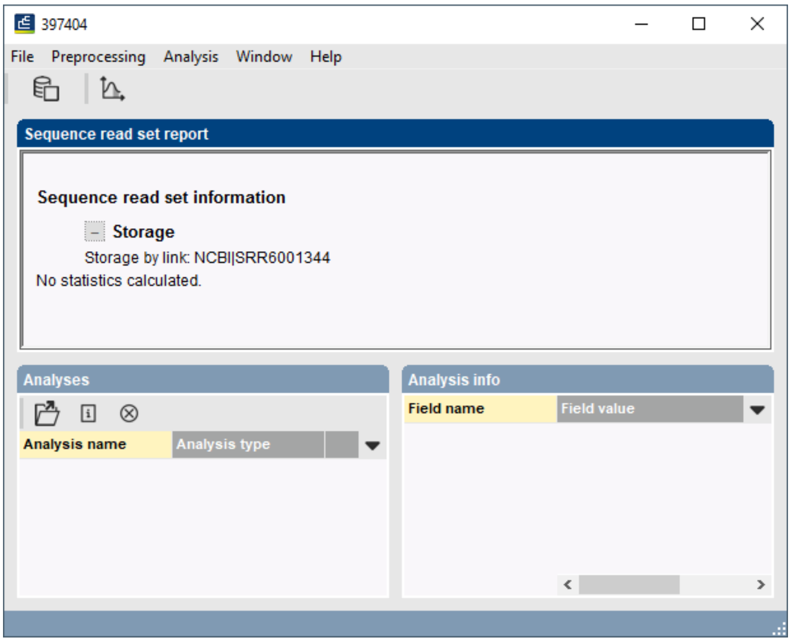
☑ Show advanced options

Figure 5: The **import rules** grid after assigning the file name as ***Key***.

Once the import is completed, the entry **397404** is created/updated and has one green dot next to it in the column of the sequence read set experiment type **wgs.**

24. Click on the green colored dot of the imported entry corresponding to the experiment type **wgs**.

The data links are displayed in the *Sequence read set experiment* window (see Figure 6).



Sequence read set information

Storage

Storage by link: NCBI|SRR6001344

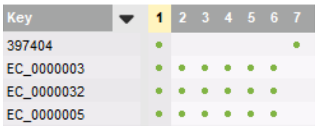
No statistics calculated.

Figure 6: The data links are displayed in the *Sequence read set experiment* window.

25. Close the *Sequence read set experiment* window.

Now, you also need to **import the long reads data** and link it to the same entry using the accession code in the **RunLong** entry field by repeating the previous steps of this part of the tutorial. You should make sure to select the ***wgsLong*** experiment type to import the data to.

After a successful import both the short reads and long reads data should be linked to entry **397404** (see Figure 7).



Key	1	2	3	4	5	6	7
397404	●						●
EC_0000003	●	●	●	●	●	●	
EC_0000032	●	●	●	●	●	●	
EC_0000005	●	●	●	●	●	●	

Figure 7: After a successful import the short reads should be linked to experiment '1' (**wgs**) and the long reads to experiment '7' (**wgsLong**) of entry **397404**.

26. Click on the green colored dot of entry **397404** corresponding to the experiment type **wgsLong** to check whether the data was imported correctly.

27. Close the *Sequence read set experiment* window.

4 Performing a hybrid de novo assembly in the cloud

Launching the hybrid de novo assembly job on the Cloud Calculation Engine is a very easy process.

1. In the *Main* window, select the entries that you want to analyze using the check-boxes next to the entries or with the **Ctrl**- or **Shift**-keys. In this example, make sure entry **397404** is selected.

2. Select ***WGS tools*** > ***Submit jobs...*** (▷) to call the *Submit jobs* dialog box (see Figure 8).

Alternatively check the ***Open submit jobs dialog after import*** option in the *Processing* wizard page during import of the data.

3. To run the de novo assembly job on the Cloud Calculation Engine, check the ***Calculation Engine*** option.

4. If you are only interested in performing a de novo assembly based on the reads obtained after trimming (automated trimming step), check the ***De novo assembly*** option and uncheck all other options.

5. With the ***De novo assembly*** option highlighted, press the <***Settings...***> button.

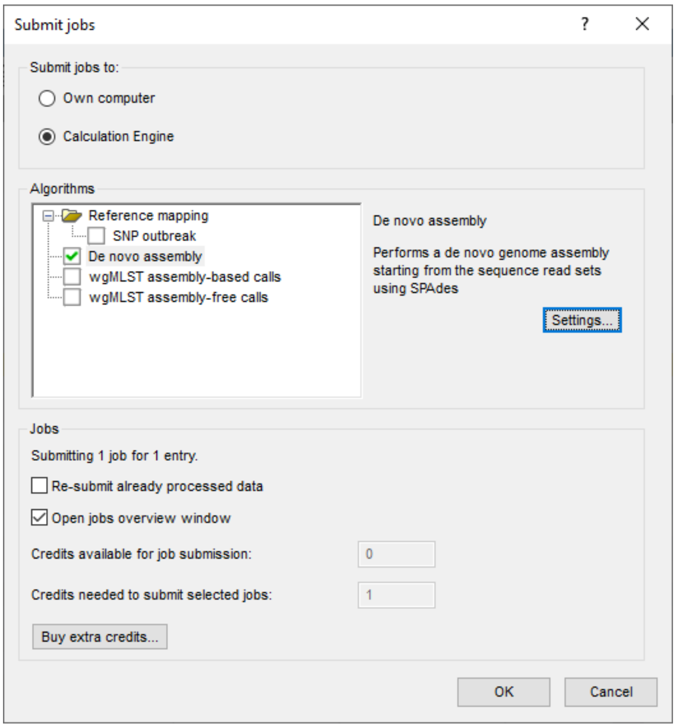
Following de novo assemblers are available on the Cloud Calculation Engine: ***Velvet (Optimizer)***, ***SPAdes*** (default), ***SKESA*** and ***Unicycler*** (see Figure 9). A hybrid de novo assembly can only be performed by ***Unicycler***.

6. Select ***Unicycler*** from the ***Assembler*** drop-down menu (see Figure 10).

7. Check the box before ***Hybrid assembly*** (see Figure 10).

The ***Hybrid assembly*** option will not be available if the entry does not contain any sequence read set data for the **wgs** or **wgsLong** experiment.

If the ***Hybrid assembly*** option is unchecked, the ***Unicycler*** algorithm will create an assembly based on the short reads only.



Submit jobs

Submit jobs to:

○ Own computer

● Calculation Engine

Algorithms

- Reference mapping
  - □ SNP outbreak
- ☑ De novo assembly
- □ wgMLST assembly-based calls
- □ wgMLST assembly-free calls

De novo assembly

Performs a de novo genome assembly starting from the sequence read sets using SPAdes

Settings...

Jobs

Submitting 1 job for 1 entry.

□ Re-submit already processed data

☑ Open jobs overview window

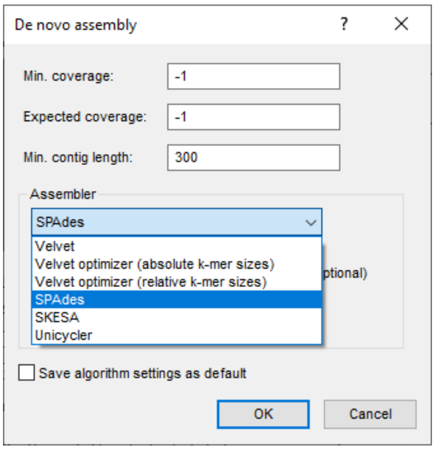
Credits available for job submission: 0

Credits needed to submit selected jobs: 1

Buy extra credits...

OK Cancel

Figure 8: Submit de novo assembly job on the Cloud Calculation Engine.



De novo assembly

Min. coverage: -1

Expected coverage: -1

Min. contig length: 300

Assembler

SPAdes

- Velvet
- Velvet optimizer (absolute k-mer sizes)
- Velvet optimizer (relative k-mer sizes)
- SPAdes
- SKESA
- Unicycler

ptional)

□ Save algorithm settings as default

OK Cancel

Figure 9: Cloud Calculation Engine: available de novo assemblers.

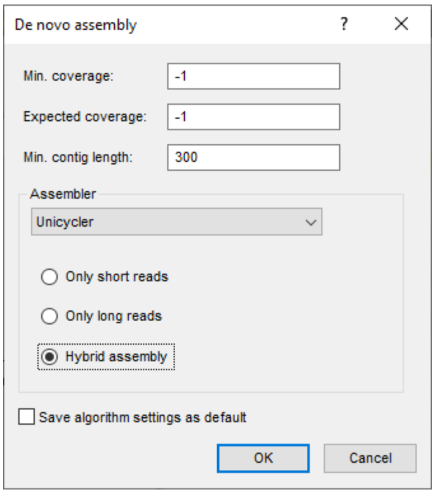
8. Close the *Perform de novo assembly* dialog box.

Jobs that already have been submitted and have been imported successfully, will not be relaunched for analysis, unless the check box in front of ***Re-submit already processed data*** in the ***Jobs*** part is checked.

Credit costs depend on the job that is submitted: 10 credits are counted for 1 hybrid de novo assembly.

9. Press <***OK***> to launch the job on the Cloud Calculation Engine.

When not sufficient credits are available for the submission of the job(s) to the external Cloud Calculation Engine, an error message pops up. Since no credits are assigned to the demo project, this error message will pop up when following this workflow in the demonstration database. Please consult Applied Maths for more information about the purchase of credits.



De novo assembly

Min. coverage: -1

Expected coverage: -1

Min. contig length: 300

Assembler: Unicycler

○ Only short reads

○ Only long reads

● Hybrid assembly

□ Save algorithm settings as default

Figure 10: Settings for the Unicycler assembly algorithm.

When sufficient credits are available for the submission of the job(s) to the external Cloud Calculation Engine, the Cloud Calculation Engine will start the job by retrieving the sequencing read set data directly from the NCBI SRA repository.

10. By default, the *Job overview* window will open after submission of the job(s). The same dialog can be called at any time with ***WGS tools*** > ***Jobs overview...*** (⚙).

The *Entry* key, the *Submitted time*, the job *Status*, a *Description* of the job and its *Progress* and much more can be monitored. In the *Message* field, the run comments are displayed in real time.

On average, the calculation time for a novo assembly on the Cloud Calculation Engine is around **90-120 min**.

11. To refresh the overview, press ***View*** > ***Refresh*** (↻, **F5**).

12. Finished jobs can be imported with a manual action (***Jobs*** > ***Get results*** (⚙)) or through an automatic update: select ***File*** > ***Settings***, check both options and specify an interval (e.g. 10 min).

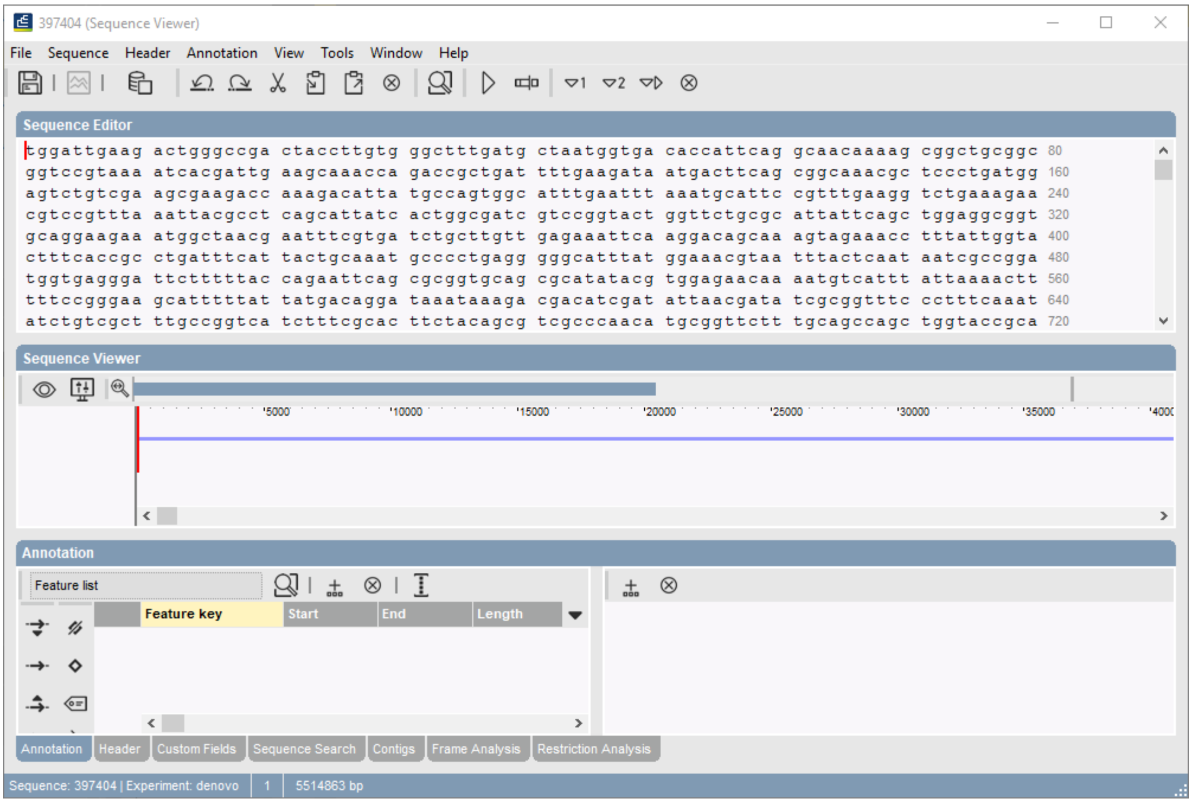
Once the results are imported, the corresponding jobs and their underlying data sets are automatically deleted from the calculation engine and as such, from the *Job overview* window.

The results from the de novo assembly algorithm, i.e. concatenated de novo contig sequences with coverage information are stored in the sequence experiment type **denovo**.

13. Click on the green colored dot for the entry **397404** in the third column in the *Experiment presence* panel. Column 3 corresponds to the third experiment type listed in the *Experiment types* panel, which is **denovo**.

The *Sequence editor* window opens, containing the results from the de novo assembly algorithm, i.e. concatenated de novo contig sequences (see Figure 11).

The de novo assembly can be used for bacterial typing using whole genome Multi Locus Sequence Typing (wgMLST). It can also be used as reference genome in a whole genome Single Nucleotide Polymorphism (wgSNP) analysis, or it can be compared to other whole genome assemblies using minHashing. For all these analyses, please refer to the corresponding tutorials on our website `https://www.bionumerics.com/tutorials`.



```
tggattgaag actgggccga ctaccttgtg ggctttgatg ctaatggtga caccattcag gcaacaaaag cggctgcggc 80
ggtccgtaaa atcacgattg aagcaaacca gaccgctgat tttgaagata atgacttcag cggcaaacgc tccctgatgg 160
agtctgtcga agcgaagacc aaagacatta tgccagtggc atttgaattt aaatgcattc cgtttgaagg tctgaaagaa 240
cgtccgttta aattacgcct cagcattatc actggcgatc gtccggtact ggttctgcgc attattcagc tggaggcggt 320
gcaggaagaa atggctaacg aatttcgtga tctgcttgtt gagaaattca aggacagcaa agtagaaacc tttattggta 400
ctttcaccgc ctgatttcat tactgcaaat gcccctgagg gggcatttat ggaaacgtaa tttactcaat aatcgccgga 480
tggtgaggga ttctttttac cagaattcag cgcggtgcag cgcatatacg tggagaacaa aatgtcattt attaaaactt 560
tttccgggaa gcatttttat tatgacagga taaataaaga cgacatcgat attaacgata tcgcggtttc cctttcaaat 640
atctgtcgct ttgccggtca tctttcgcac ttctacagcg tcgcccaaca tgcggttctt tgcagccagc tggtaccgca 720
```

Sequence: 397404 | Experiment: denovo | 1 | 5514863 bp

Figure 11: The *Sequence editor* window containing the concatenated de novo contig sequences.