BIONUMERICS Tutorial:

Annotating sequences by Prokka

1 Introduction

This tutorial demonstrates the annotation of genomes by Prokka [1] in BIONUMERICS. The Prokka algorithm is only available on the Calculation Engine.

To avoid issues with gaps in reference mapped (i.e. aligned) sequences, the Prokka algorithm is limited to sequence types that are *not* reference mapped.

2 Preparing the database

2.1 Introduction

The **Annotation by Prokka** pipeline can only be performed in BIONUMERICS after installation of the *WGS tools plugin* in the BIONUMERICS database (***File*** > ***Install / remove plugins...*** (🔌)).

As the Prokka job is only available on the Cloud Calculation engine make sure to select the options ***Use default Cloud Calculation Engine*** and ***Enable running jobs on Cloud Calculation Engine*** during installation of the *WGS tools plugin*. The Calculation engine option requires credits for running jobs on the Applied Maths cloud calculation engine. Credits are linked to credentials that you need to enter when installing the *WGS tools plugin*.

In this tutorial the **WGS demo database for *Salmonella*** will be used in which the *WGS tools plugin* is already installed. No credits are assigned to the demo project so no Prokka jobs can be launched on the external calculation engine. Please contact Applied Maths to obtain more information.

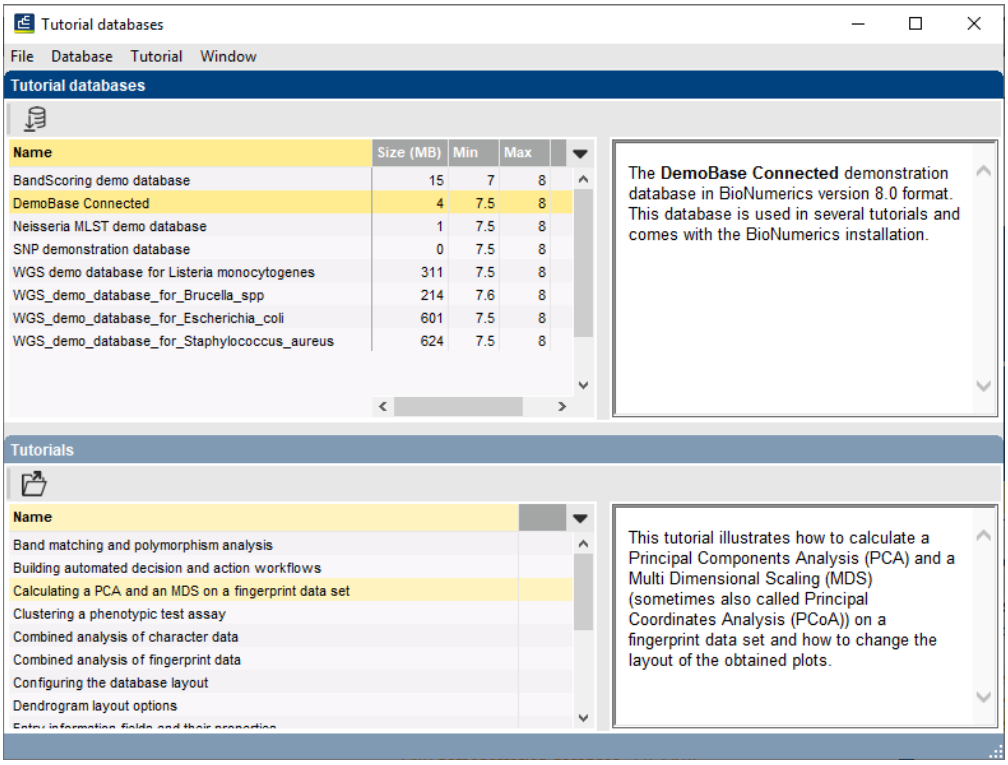
The **WGS demo database for *Salmonella*** can be downloaded directly from the *BIONUMERICS Startup* window (see 2.2), or restored from the back-up file available on our website (see 2.3).

2.2 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).

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



Name	Size (MB)	Min	Max
BandScoring demo database	15	7	8
DemoBase Connected	4	7.5	8
Neisseria MLST demo database	1	7.5	8
SNP demonstration database	0	7.5	8
WGS demo database for Listeria monocytogenes	311	7.5	8
WGS_demo_database_for_Brucella_spp	214	7.6	8
WGS_demo_database_for_Escherichia_coli	601	7.5	8
WGS_demo_database_for_Staphylococcus_aureus	624	7.5	8

The **DemoBase Connected** demonstration database in BioNumerics version 8.0 format. This database is used in several tutorials and comes with the BioNumerics installation.

Name
Band matching and polymorphism analysis
Building automated decision and action workflows
Calculating a PCA and an MDS on a fingerprint data set
Clustering a phenotypic test assay
Combined analysis of character data
Combined analysis of fingerprint data
Configuring the database layout
Dendrogram layout options
Entry information fields and their properties

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.

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_Salmonella_enterica** appears in the *BIONUMERICS Startup* window.

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

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

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

6. Download the file `WGS_Salm.bnbk` file from https://www.applied-maths.com/download/sample-data, under 'WGS_demo_database_for_Salmonella_enterica'.

In contrast to other browsers, some versions of Internet Explorer rename the `WGS_Salm.bnbk` database backup file into `WGS_Salm.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...***.

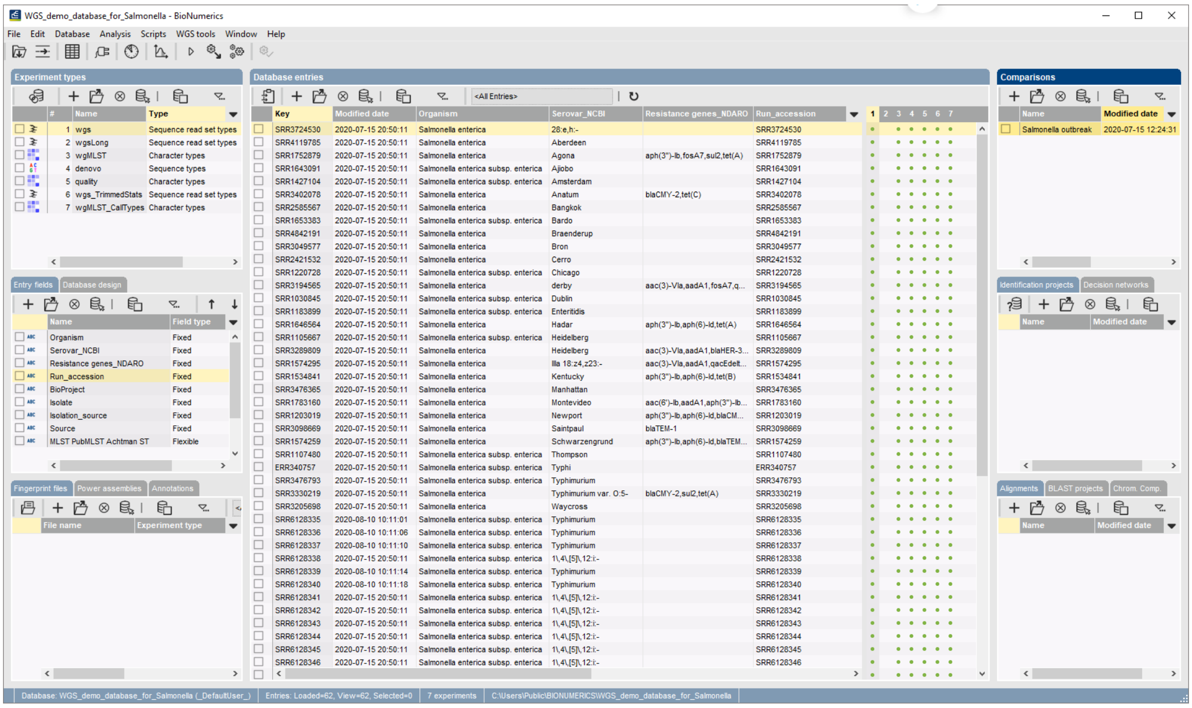
8. Browse for the downloaded file and select ***Create copy***. Note that, if ***Overwrite*** is selected, an existing database will be overwritten.

9. Specify a new name for this demonstration database, e.g. “WGS_Salmonella_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	wgsLong	Sequence read set types
3	wgMLST	Character types
4	denovo	Sequence types
5	quality	Character types
6	wgs_TrimmedStats	Sequence read set types
7	wgMLST_CallTypes	Character types

Name	Field type
Organism	Fixed
Serovar_NCBI	Fixed
Resistance genes_NDARO	Fixed
Run_accession	Fixed
BioProject	Fixed
Isolate	Fixed
Isolation_source	Fixed
Source	Fixed
MLST PubMLST Achtman ST	Flexible

File name	Experiment type

Key	Modified date	Organism	Serovar_NCBI	Resistance genes_NDARO	Run_accession
SRR3724530	2020-07-15 20:50:11	Salmonella enterica	28:e,h:-		SRR3724530
SRR4119785	2020-07-15 20:50:11	Salmonella enterica	Aberdeen		SRR4119785
SRR1752879	2020-07-15 20:50:11	Salmonella enterica	Agona	aph(3")-Ib,fosA7,sul2,tet(A)	SRR1752879
SRR1643091	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Ajiobo		SRR1643091
SRR1427104	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Amsterdam		SRR1427104
SRR3402078	2020-07-15 20:50:11	Salmonella enterica	Anatum	blaCMY-2,tet(C)	SRR3402078
SRR2585567	2020-07-15 20:50:11	Salmonella enterica	Bangkok		SRR2585567
SRR1653383	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Bardo		SRR1653383
SRR4842191	2020-07-15 20:50:11	Salmonella enterica	Braenderup		SRR4842191
SRR3049577	2020-07-15 20:50:11	Salmonella enterica	Bron		SRR3049577
SRR2421532	2020-07-15 20:50:11	Salmonella enterica	Cerro		SRR2421532
SRR1220728	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Chicago		SRR1220728
SRR3194565	2020-07-15 20:50:11	Salmonella enterica	derby	aac(3)-VIa,aadA1,fosA7,q...	SRR3194565
SRR1030845	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Dublin		SRR1030845
SRR1183899	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Enteritidis		SRR1183899
SRR1646564	2020-07-15 20:50:11	Salmonella enterica	Hadar	aph(3")-Ib,aph(6)-Id,tet(A)	SRR1646564
SRR1105667	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Heidelberg		SRR1105667
SRR3289809	2020-07-15 20:50:11	Salmonella enterica	Heidelberg	aac(3)-VIa,aadA1,blaHER-3...	SRR3289809
SRR1574295	2020-07-15 20:50:11	Salmonella enterica	IIIa 18:z4,z23:-	aac(3)-VIa,aadA1,qacEdelt...	SRR1574295
SRR1534841	2020-07-15 20:50:11	Salmonella enterica	Kentucky	aph(3")-Ib,aph(6)-Id,tet(B)	SRR1534841
SRR3476365	2020-07-15 20:50:11	Salmonella enterica	Manhattan		SRR3476365
SRR1783160	2020-07-15 20:50:11	Salmonella enterica	Montevideo	aac(6')-Ib,aadA1,aph(3")-Ib...	SRR1783160
SRR1203019	2020-07-15 20:50:11	Salmonella enterica	Newport	aph(3")-Ib,aph(6)-Id,blaCM...	SRR1203019
SRR3098669	2020-07-15 20:50:11	Salmonella enterica	Saintpaul	blaTEM-1	SRR3098669
SRR1574259	2020-07-15 20:50:11	Salmonella enterica	Schwarzengrund	aph(3")-Ib,aph(6)-Id,blaTEM...	SRR1574259
SRR1107480	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Thompson		SRR1107480
ERR340757	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Typhi		ERR340757
SRR3476793	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	Typhimurium		SRR3476793
SRR3330219	2020-07-15 20:50:11	Salmonella enterica	Typhimurium var. O:5-	blaCMY-2,sul2,tet(A)	SRR3330219
SRR3205698	2020-07-15 20:50:11	Salmonella enterica	Waycross		SRR3205698
SRR6128335	2020-08-10 10:11:01	Salmonella enterica subsp. enterica	Typhimurium		SRR6128335
SRR6128336	2020-08-10 10:11:06	Salmonella enterica subsp. enterica	Typhimurium		SRR6128336
SRR6128337	2020-08-10 10:11:10	Salmonella enterica subsp. enterica	Typhimurium		SRR6128337
SRR6128338	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128338
SRR6128339	2020-08-10 10:11:14	Salmonella enterica subsp. enterica	Typhimurium		SRR6128339
SRR6128340	2020-08-10 10:11:18	Salmonella enterica subsp. enterica	Typhimurium		SRR6128340
SRR6128341	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128341
SRR6128342	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128342
SRR6128343	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128343
SRR6128344	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128344
SRR6128345	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128345
SRR6128346	2020-07-15 20:50:11	Salmonella enterica subsp. enterica	1\,4\,[5]\,12:i:-		SRR6128346

Name	Modified date
Salmonella outbreak	2020-07-15 12:24:31

Database: WGS_demo_database_for_Salmonella (_DefaultUser_) | Entries: Loaded=62, View=62, Selected=0 | 7 experiments | C:\Users\Public\BIONUMERICS\WGS_demo_database_for_Salmonella

Figure 2: The *Salmonella* demonstration database: the *Main* window.

3 Launch a Prokka job on the Calculation Engine

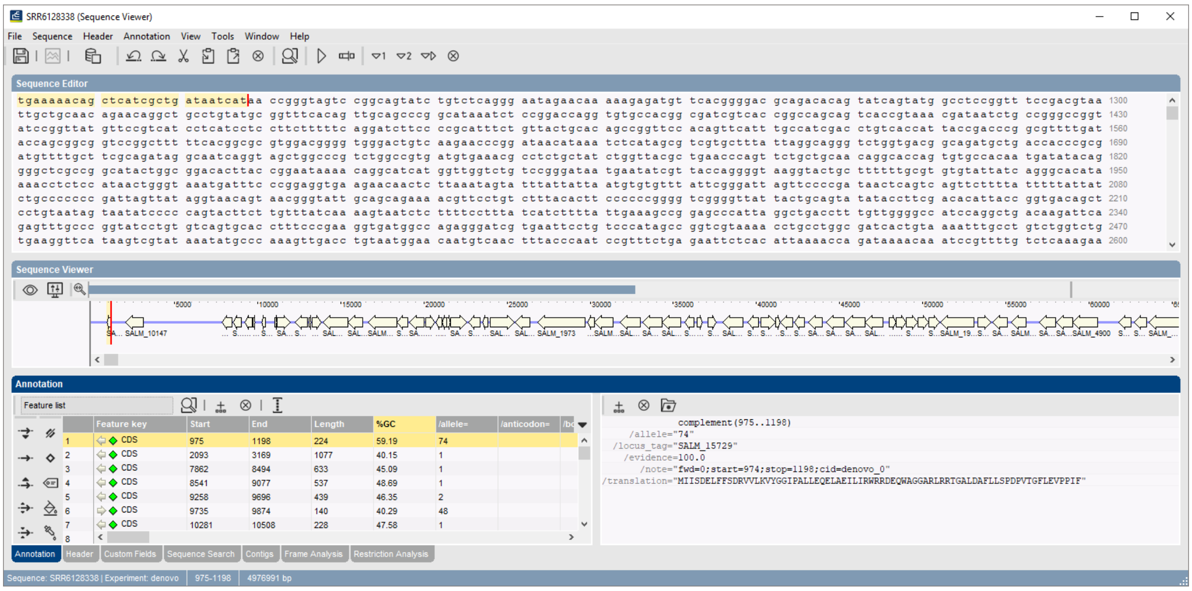
The **Annotation by Prokka** pipeline can be launched from the *Main* window on one or multiple selected entries.

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, select the entry with key ”SRR6128338”.

2. In the *Experiment presence* panel click on the green dot corresponding to the denovo experiment for the selected entry to open the *Sequence editor* window (see Figure 3).

The genome sequence of the entry with key ”SRR6128338” is shown in the *Sequence Editor* panel. The annotation obtained from the wgMLST analysis is presented graphically in the *Sequence Viewer* panel and listed in the *Annotation* panel.

3. Highlight a feature in the feature list of the *Annotation* panel to make the annotation information available in the panel on the right (see Figure 3).


SRR6128338 (Sequence Viewer)
File Sequence Header Annotation View Tools Window Help
Sequence Editor
Sequence Viewer
'5000 '10000 '15000 '20000 '25000 '30000 '35000 '40000 '45000 '50000 '55000 '60000
SALM_10147
SALM_1973
SALM_19...
SALM_4900
Annotation
Feature list
Annotation Header Custom Fields Sequence Search Contigs Frame Analysis Restriction Analysis

```
tgaaaaacag ctcatcgctg ataatcataa ccgggtagtc cggcagtatc tgtctcaggg aatagaacaa aaagagatgt tcacggggac gcagacacag tatcagtatg gcctccggtt tccgacgtaa 1300
ttgctgcaac agaacaggct gcctgtatgc ggtttcacag ttgcagcccg gcataaatct ccggaccagg tgtgccacgg cgatcgtcac cggccagcag tcaccgtaaa cgataatctg ccgggccggt 1430
atccggttat gttccgtcat cctcatcctc cttctttttc aggatcttcc ccgcatttct gttactgcac agccggttcc acagttcatt tgccatcgac ctgtcaccat taccgacccg gcgttttgat 1560
accagcggcg gtccggcttt ttcacggcgc gtggacgggg tgggactgtc aagaacccgg ataacataaa tctcatagcg tcgtgcttta ttaggcaggg tctggtgacg gcagatgctg accacccgcg 1690
atgttttgct tcgcagatag gcaatcaggt agctggcccg tctggccgtg atgtgaaacg cctctgctat ctggttacgc tgaacccagt tctgctgcaa caggcaccag tgtgccacaa tgatatacag 1820
gggctcgccg gcatactggc ggacacttac cggaataaaa caggcatcat ggttggtctg tccgggataa tgaatatcgt taccaggggt aaggtactgc ttttttgcgt gtgtattatc agggcacata 1950
aaacctctcc ataactgggt aaatgatttc ccggaggtga agaacaactc ttaaatagta tttattatta atgtgtgttt attcgggatt agttccccga taactcagtc agttctttta tttttattat 2080
ctgccccccc gattagttat aggtaacagt aacgggtatt gcagcagaaa acgttcctgt ctttacactt cccccggggg tcggggttat tactgcagta tataccttcg acacattacc ggtgacagct 2210
cctgtaatag taatatcccc cagtacttct tgtttatcaa aagtaatctc ttttccttta tcatctttta ttgaaagccg gagcccatta ggctgacctt tgttggggcc atccaggctg acaagattca 2340
gagtttgccc ggtatcctgt gtcagtgcac ctttcccgaa ggtgatggcc agagggatcg tgaattcctg tcccatagcc ggtcgtaaaa cctgcctggc gatcactgta aaatttgcct gtctggtctg 2470
tgaaggttca taagtcgtat aaatatgccc aaagttgacc tgtaatggaa caatgtcaac tttacccaat ccgtttctga gaattctcac attaaaacca gataaaacaa atccgttttg tctcaaagaa 2600
```

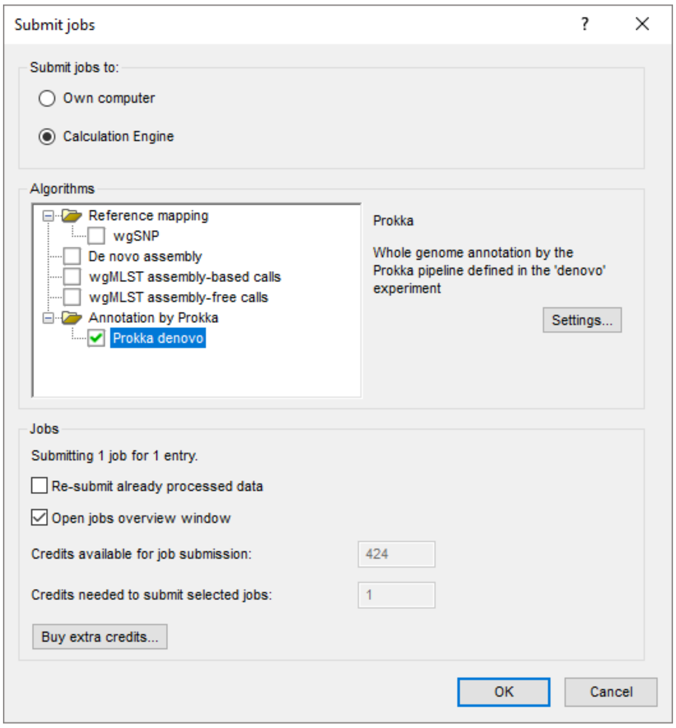
	Feature key	Start	End	Length	%GC	/allele=	/anticodon=	/bc
1	CDS	975	1198	224	59.19	74		
2	CDS	2093	3169	1077	40.15	1		
3	CDS	7862	8494	633	45.09	1		
4	CDS	8541	9077	537	48.69	1		
5	CDS	9258	9696	439	46.35	2		
6	CDS	9735	9874	140	40.29	48		
7	CDS	10281	10508	228	47.58	1		
8								

```
              complement(975..1198)
      /allele="74"
  /locus_tag="SALM_15729"
    /evidence=100.0
       /note="fwd=0;start=974;stop=1198;cid=denovo_0"
/translation="MIISDELFFSDRVVLKVYGGIPALLEQELAEILIRWRRDEQWAGGARLRRTGALDAFLLSPDPVTGFLEVPPIF"
```

Sequence: SRR6128338 | Experiment: denovo | 975-1198 | 4976991 bp

Figure 3: The *Sequence editor* window for the selected entry before annotation by Prokka.

4. Close the *Sequence editor* window and select ***WGS tools*** > ***Submit jobs...*** (▷) to call the *Submit jobs* dialog box (see Figure 4).



Submit jobs to:

○ Own computer

● Calculation Engine

Algorithms

- Reference mapping
  - □ wgSNP
- □ De novo assembly
- □ wgMLST assembly-based calls
- □ wgMLST assembly-free calls
- Annotation by Prokka
  - ☑ Prokka denovo

Prokka

Whole genome annotation by the Prokka pipeline defined in the 'denovo' experiment

Jobs

Submitting 1 job for 1 entry.

□ Re-submit already processed data

☑ Open jobs overview window

Credits available for job submission: 424

Credits needed to submit selected jobs: 1

Figure 4: The *Submit jobs* dialog box.

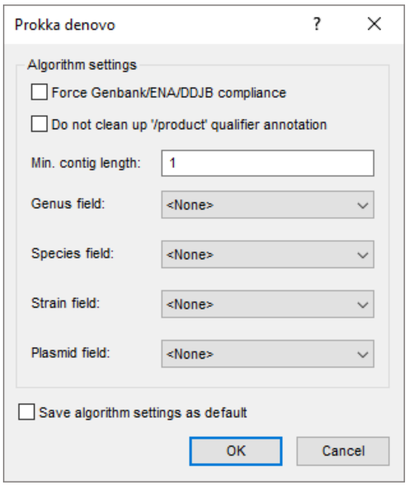
5. To run an **Annotation by Prokka** pipeline job on the cloud calculation engine, check the ***Calculation Engine*** option in the ***Submit jobs to*** panel and the **Prokka denovo** option in the ***Algorithms*** panel.

Note that if the selected entry contains sequence data in additional non-reference mapped sequence experiment types in the database these experiment types would also be available under the **Annotation by Prokka** algorithm.

The settings for the Prokka job can be defined by highlighting the job type and pressing <***Settings...***>.

6. Highlight the **Prokka denovo** job type and press <***Settings...***>.

This action displays the *Prokka settings* dialog box (see Figure 5).



Prokka denovo

Algorithm settings

□ Force Genbank/ENA/DDJB compliance

□ Do not clean up '/product' qualifier annotation

Min. contig length: 1

Genus field: <None>

Species field: <None>

Strain field: <None>

Plasmid field: <None>

□ Save algorithm settings as default

Figure 5: The *Prokka settings* dialog box.

Checking ***Force GenBank/ENA/DDJB compliance*** will make the annotations compliant with submission criteria from the GenBank, ENA and DDJB online repositories: add 'gene' features for each 'CDS' feature and a minimum contig length of 200 bp.

By default, Prokka tries to clean up the '/product' names to ensure they are compliant with GenBank/ENA conventions. Checking ***Do not clean up '/product' qualifier annotation*** will prevent this behavior.

The minimum size of a contig to be considered for annotation (***Min. contig length***) can be entered in bp.

In case the corresponding information is already present in the BIONUMERICS database, optionally a ***Genus field***, ***Species field***, ***Strain field*** and/or ***Plasmid field*** can be specified. Entry information contained in these fields will then be included in the annotation, which facilitates later submission to online repositories.

When altering these settings, one can save the updated settings as defaults to the database with ***Save algorithm settings as default***.

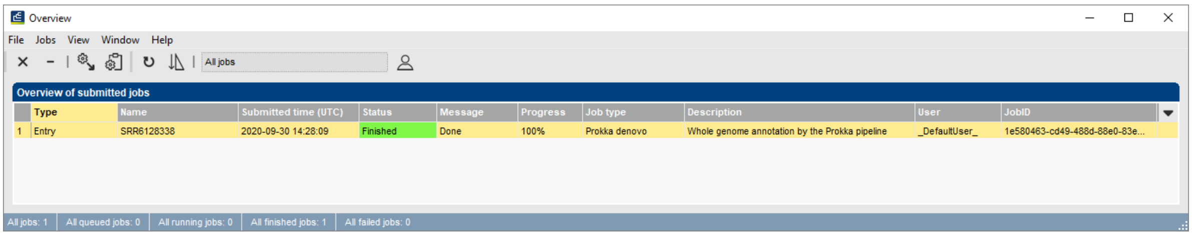
7. Check the ***Force GenBank/ENA/DDJB compliance*** option and leave the other settings as default.

8. Press the <***OK***> twice to launch the **Annotation by Prokka** pipeline job on the cloud calculation engine.

The job is submitted to the Calculation Engine and the *Job overview* window opens. In the *Job overview* window, the job type, job name, time of submission, job status, a description of the job, its progress and much more can be monitored.

4 Import and analyze Prokka job results

Once the job has been finished (see Figure 6), the results can be imported in the database by selecting ***Jobs*** > ***Get results*** (⚙) from the *Job overview* window.



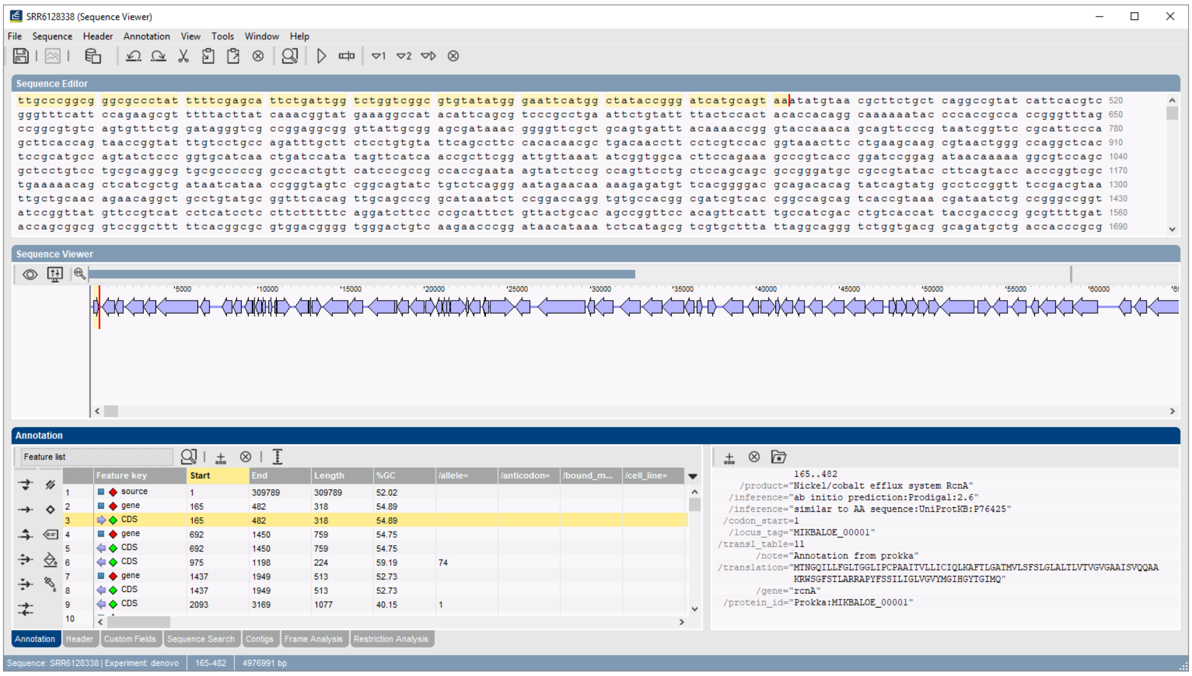
	Type	Name	Submitted time (UTC)	Status	Message	Progress	Job type	Description	User	JobID
1	Entry	SRR6128338	2020-09-30 14:28:09	Finished	Done	100%	Prokka denovo	Whole genome annotation by the Prokka pipeline	_DefaultUser_	1e580463-cd49-488d-88e0-83e...

Figure 6: The *Job overview* window.

1. When the job is finished, highlight the job and select ***Jobs*** > ***Get results*** (⚙) to import the results in the database.

The annotation will be imported in the BIONUMERICS database and saved with the corresponding sequence experiment. A newer Prokka annotation will replace any earlier Prokka annotation on the same sequence, but manually created features or annotation features from other tools will not be overwritten: Prokka features will be added, even if they are defined on exactly the same positions.

2. In the *Experiment presence* panel click on the green dot corresponding to the denovo experiment for the selected entry (i.e. entry with key "SRR6128338") to open the *Sequence editor* window (see Figure 7).



SRR6128338 (Sequence Viewer)
File Sequence Header Annotation View Tools Window Help
Sequence Editor

```
ttgcccggcg ggcgccctat ttttcgagca ttctgattgg tctggtcggc gtgtatatgg gaattcatgg ctataccggg atcatgcagt aaatatgtaa cgcttctgct caggccgtat cattcacgtc 520
gggtttcatt ccagaagcgt ttttacttat caaacggtat gaaaggccat acattcagcg tcccgcctga attctgtatt ttactccact acaccacagg caaaaaatac cccaccgcca ccgggtttag 650
ccggcgtgtc agtgtttctg gatagggtcg ccggaggcgg gttattgcgg agcgataaac ggggttcgct gcagtgattt acaaaaccgg gtaccaaaca gcagttcccg taatcggttc cgcattccca 780
gcttcaccag taaccggtat ttgtcctgcc agatttgctt ctcctgtgta ttcagccttc cacacaacgc tgacaacctt cctcgtccac ggtaaacttc ctgaagcaag cgtaactggg ccaggctcac 910
tccgcatgcc agtatctccc ggtgcatcaa ctgatccata tagttcatca accgcttcgg attgttaaat atcggtggca cttccagaaa gcccgtcacc ggatccggag ataacaaaaa ggcgtccagc 1040
gctcctgtcc tgcgcaggcg tgcgcccccg gcccactgtt catcccgccg ccaccgaata agtatctccg ccagttcctg ctccagcagc gccgggatgc cgccgtatac cttcagtacc acccggtcgc 1170
tgaaaaacag ctcatcgctg ataatcataa ccgggtagtc cggcagtatc tgtctcaggg aatagaacaa aaagagatgt tcacggggac gcagacacag tatcagtatg gcctccggtt tccgacgtaa 1300
ttgctgcaac agaacaggct gcctgtatgc ggtttcacag ttgcagcccg gcataaatct ccggaccagg tgtgccacgg cgatcgtcac cggccagcag tcaccgtaaa cgataatctg ccgggccggt 1430
atccggttat gttccgtcat cctcatcctc cttctttttc aggatcttcc ccgcatttct gttactgcac agccggttcc acagttcatt tgccatcgac ctgtcaccat taccgacccg gcgttttgat 1560
accagcggcg gtccggcttt ttcacggcgc gtggacgggg tgggactgtc aagaacccgg ataacataaa tctcatagcg tcgtgcttta ttaggcaggg tctggtgacg gcagatgctg accacccgcg 1690
```

	Feature key	Start	End	Length	%GC	/allele=	/anticodon=	/bound_m...	/cell_line=
1	source	1	309789	309789	52.02				
2	gene	165	482	318	54.89				
3	CDS	165	482	318	54.89				
4	gene	692	1450	759	54.75				
5	CDS	692	1450	759	54.75				
6	CDS	975	1198	224	59.19	74			
7	gene	1437	1949	513	52.73				
8	CDS	1437	1949	513	52.73				
9	CDS	2093	3169	1077	40.15	1			
10									

```
                 165..482
     /product="Nickel/cobalt efflux system RcnA"
   /inference="ab initio prediction:Prodigal:2.6"
   /inference="similar to AA sequence:UniProtKB:P76425"
  /codon_start=1
    /locus_tag="MIKBALOE_00001"
 /transl_table=11
        /note="Annotation from prokka"
 /translation="MTNGQILLFGLTGGLIPCPAAITVLLICIQLKAFTLGATMVLSFSLGLALTLVTVGVGAAISVQQAA
              KRWSGFSTLARRAPYFSSILIGLVGVYMGIHGYTGIMQ"
        /gene="rcnA"
  /protein_id="Prokka:MIKBALOE_00001"
```

Figure 7: The *Sequence editor* window after annotation by Prokka.

The genome sequence of the entry with key "SRR6128338" is shown in the *Sequence Editor* panel. The features detected by Prokka are added to the feature list and can be recognized by the value ”Annotation by Prokka” in the information field /**note=**.

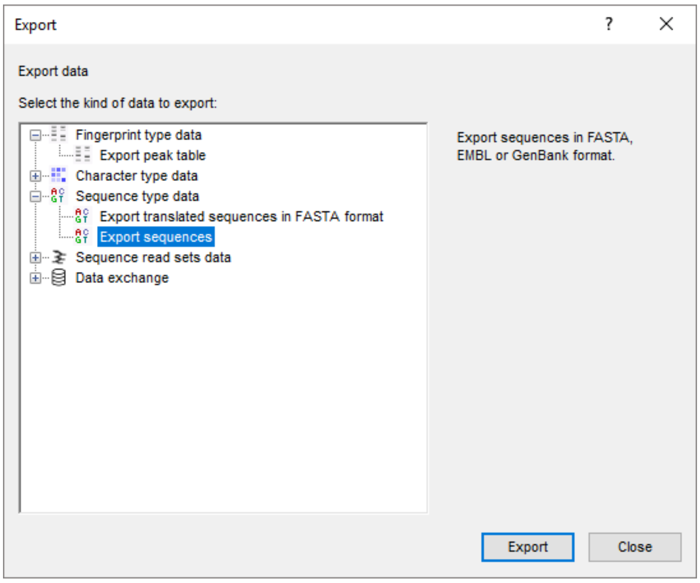
3. Right-click on the **Start** information field and select **Arrange by field** to sort the features based on the start position on the genome. Select a feature in the feature list to make the annotation information available in the panel on the right (see Figure 7).

4. Close the *Sequence editor* window.

The genome sequences can be exported in EMBL or GenBank format.

5. In the *Main* window select ***File*** > ***Export...***.

The *Export* dialog box opens (see Figure 8).



Export sequences in FASTA, EMBL or GenBank format.

Figure 8: The *Export* dialog box.

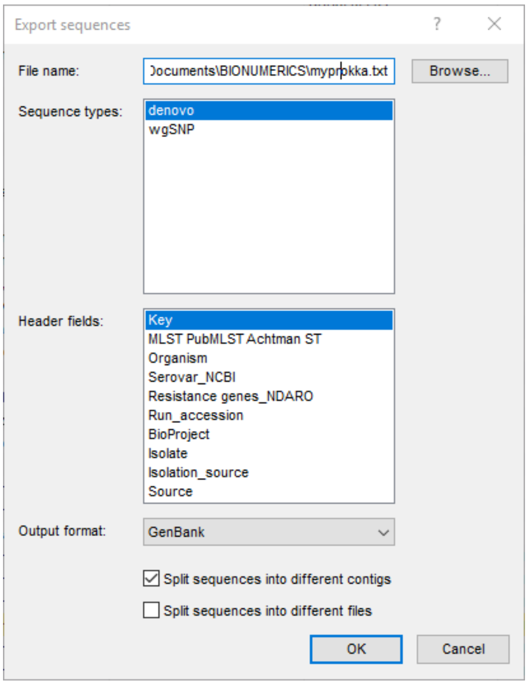
6. In the *Export* dialog box select ***Export sequences*** under ***Sequence type data*** and <***Export***>.

The *Export sequences* dialog box open (see Figure 9).

7. Browse for the preferred file location, enter a name and press <***Open***>.

8. Select **denovo** as sequence type, **Key** as header field and **GenBank** as output format. Leave the other settings as default and press <***OK***>.

The genome sequence with the Prokka annotation will be exported in GenBank format to the provided file location.



Export sequences

File name: Documents\BIONUMERICS\myprokka.txt

Sequence types: denovo, wgSNP

Header fields: Key, MLST PubMLST Achtman ST, Organism, Serovar_NCBI, Resistance genes_NDARO, Run_accession, BioProject, Isolate, Isolation_source, Source

Output format: GenBank

☑ Split sequences into different contigs

☐ Split sequences into different files

Figure 9: The *Export sequences* dialog box.

Bibliography

[1] Torsten Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069, 2014.