

BIONUMERICS Tutorial:

E. coli functional genotyping: predicting phenotypic traits from whole genome sequences

1 Aim

In this tutorial we will screen whole genome sequences of *Escherichia coli* samples for phenotypic traits using the *E. coli functional genotyping plugin*. This plugin contains knowledge bases for serotype, virulence and antibiotic resistance prediction, as well as plasmid and phage detection. An *in silico* PCR tool is also implemented, making it possible to detect Shiga toxin gene subtypes and virulence genes, mimicking the wet lab PCR.

The different steps are illustrated using the whole genome demonstration database of *Escherichia coli*. This database is available for download on our website (see 2) and contains 60 publicly available sequence read sets of *Escherichia coli* with already calculated de novo assemblies.

2 Preparing the database

2.1 Introduction to the demonstration database

We provide a **WGS demo database** for *Escherichia coli* containing sequence read set data links for 60 samples, calculated de novo assemblies and wgMLST results (allele calls and quality information).



The wgMLST workflow and results will not be discussed in this tutorial.

The **WGS_demo_database_for_Escherichia_coli** can be downloaded directly from the *BIONU-MERICS Startup* window (see 2.2), or restored from the back-up file available on our website (see 2.3).

Installation of the *E. coli functional genotyping plugin* is only possible when no spaces are present in the BIONUMERICS home directory and in the name of the database. Before downloading or restoring the **WGS demo database** for *Escherichia coli*, please check if your BIONUMERICS home directory does not contain any spaces:

- 1. Click on the Obstration button, located in the toolbar in the *BIONUMERICS Startup* window and select **Change home directory...** to call the *Home directory* dialog box.
- 2. In case the currently specified home directory contains spaces, update the path to a path containing no spaces and close the *Home directory* dialog box.

2.2 Option 1: Download demo database from the Startup Screen

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

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

🖆 Tutorial databases					>	×
File Database Tutorial Window						
, Tutorial databases						
B						
Name	Size (MB)	Min	Max	•		
BandScoring demo database	15	7	8.1	^	The DemoBase Connected demonstration	\sim
DemoBase Connected	4	7.5	8.1		database. This database is used in several	
Neisseria MLST demo database	1	7.5	8.1		tutonais.	
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			
						\sim
				÷	P	
Tutorials						
<i>₽</i>						
Name				•		_
Band matching and polymorphism analysis				^	This tutorial illustrates how to calculate a	
Entry information fields and their properties					Principal Components Analysis (PCA) and a	
Configuring the database layout					(comotimos also called Principal	
Selecting entries in a database					Coordinates Analysis (PCoA)) on a	
Combined analysis of fingerprint data					fingerprint data set and how to change the	
Calculating a PCA and an MDS on a fingerprint data set					layout of the obtained plots.	\checkmark
Clustering a phenotypic test assay				~		

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

- 4. Select WGS_demo_database_for_Escherichia_coli from the list and select *Database* > *Download* ().
- 5. Confirm the installation of the database and press < OK > after successful installation of the database.
- 6. Close the *Tutorial databases* window with *File* > *Exit*.

The WGS_demo_database_for_Escherichia_coli appears in the BIONUMERICS Startup window.

7. Double-click the **WGS_demo_database_for_Escherichia_coli** 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 *Escherichia coli* is also available on our website. This backup can be restored to a functional database in BIONUMERICS.

8. 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.

- 9. In the *BIONUMERICS Startup* window, press the button. From the menu that appears, select **Restore database...**.
- 10. Browse for the downloaded file and select *Create copy*. Note that, if *Overwrite* is selected, an existing database will be overwritten.
- 11. Specify a new name for this demonstration database, e.g. "WGS_Ecoli_demobase".
- 12. Click < OK > to start restoring the database from the backup file.
- 13. Once the process is complete, click < Yes> to open the database.

The *Main* window is displayed.

3 About the demonstration database

The **WGS_demo_database_for_Escherichia_coli** contains data for a set of 60 samples. The sample information, stored in entry info fields (Isolation source, Center Name, etc.) was collected from the publications. Seven experiments are present in the demo database and are listed in the *Experiment types* panel (see Figure 2).

Experim	ent ty	pes						
କ୍ର	·	+ 🗗 ⊗		Ê	∽	î		
		Name				-		
≣ ≣	1	wgs		Sequence	e read set t	ypes		
	2	wgMLST		Character	types			
🗆 🛟	3	denovo		Sequence	types			
	4	quality		Character types				
≣ ≩	5	wgs_TrimmedSta	ats	Sequence	e read set t	ypes		
	6	wgMLST_CallTyp	pes	Character	types			
≣ ≨	7	wgsLong		Sequence	e read set t	ypes		

Figure 2: The Experiment types panel in the Main window.

1. Click on the green colored dot for one of the entries in the first column in the *Experiment presence* panel. Column 1 corresponds to the first experiment type listed in the *Experiment types* panel, which is **wgs**.

In the *Sequence read set experiment* window, the link to the sequence read set data on NCBI (SRA) with a summary of the characteristics of the sequence read set is displayed: *Read set size*, *Sequence length statistics*, *Quality statistics*, *Base statistics* (see Figure 3).

2. Close the Sequence read set experiment window.

G EC_0000001 -		×
File Preprocessing Analysis Window Help		
Sequence read set report		
		~
Sequence read set information		
- Storage		
Storage by link: NCBIJSRR3242188		
 Read set size Number of sequences: 469412 		
Number of paired-end sequences: 469412		
Number of bases: 223023791, 1st end 111487506, 2nd end 111536283		
- Sequence length statistics		
Standard deviation of the sequence length: 38.61, 1st end 237.50, 2nd end 28.56		
Minimum sequence length: 35, 1st end 35, 2nd end 35		
Maximum sequence length: 251, 1st end 251, 2nd end 251		
Quality statistics Average base guality: 35.28, 1st end 36.52, 2nd end 34.05		
Standard deviation of the base quality: 6.44, 1st end 4.77, 2nd end 7.56		
Minimum base quality: 2, 1st end 2, 2nd end 2 Maximum base quality: 40, 1st end 40, 2nd end 40		
Q20: 207726245 (93.14%), 1st end 107815218 (96.71%), 2nd end 99911027 (89.56	3%)	
Q25: 205115896 (91.97%), 1st end 107133421 (96.09%), 2nd end 97982475 (87.85 Q30: 200557075 (89.93%), 1st end 105857364 (94.95%), 2nd end 94699711 (84.90	5%) 1%)	
	, ,0)	•
Analyses Analysis info		
☐ i ⊗ Field name Field value		-
Analysis name Analysis type 🗨		
<		>
		.:

Figure 3: The sequence read set experiment card for an entry.

3. Click on the green colored dot for one of the entries 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 4).

4. Close the Sequence editor window.

The sequence read set experiment type **wgs_TrimmedStats** contains some data statistics about the reads retained after trimming, used for the de novo assembly.

The sequence read set experiment type **wgsLong** contains the links to long read sequence read data (typically PacBio or MinION datasets). In this demo database, no links are defined for this experiment.

The other three experiments contain data related to the wgMLST analysis performed on the samples:

- Character experiment type wgMLST contains the allele calls for detected loci in each sample, where the consensus from assembly-based and assembly-free calling resulted in a single allele ID.
- Character experiment type **quality** contains quality statistics for the raw data, the de novo assembly and the different allele identification algorithms.

≝ EC_	0000001	(Sequence	e Viewer))																					-		>
ile Se	equence	Header	Annot	ation	View 1	Fools	Winde	ow He	lp																		
		ß	0	Q L	۲. ۲	ß	\otimes	Q		⊟		▽1	√2	∇b	\otimes												
							Ŭ					-	_		0												
Seque	ence Edi	tor																									
CAG	GGCCA	GG TG	TGCA	ICGC	CGGA	ACG	CGT	TTGT	TACT	GG	AA	GAGA	GCA	AT 1	IGCC	GAT	GAA	TTCI	TAGO	ссс	ΤG	TTA	AAAC	A 31	440		
GCA	GGCGC	AA AA	CTGG	CAGC	CGGG	TCA	TCC	ACTT	GATC	:cc	GCI	AACO	ACC	A :	IGGG	CAC	СТТ	AATC	GACI	GC	GC	CCAC	GCC	G 31	1520		
ACT	CGGTG	CA TA	GCTT	TATT	CGTG	AAG	GCG	AAAG	CAAA	GG	GCI	AACI	GCI	G	CTGG	ATGO	GCC	GTAA	CGCC	GA	GC	TGG	TGC	C 31	1600		
GCC.	ATCGG	CC CG	ACCA	ICTT	TGTG	GAT	GTA	GACC	CGAA	TG	CGI	rcci	TAA	A G	CCGC	GAA	GAG	ATTI	TCGO	GCC	CG	GTGI	TGG	T 31	680		
GGT	CACGO	GT TT	CACA	ICAG	AAGA	TCA	GGC	GCTT	CAGC	TT	GCI	TAAC	GAC	A (GCCA	GTA	CGG	CCTC	GGC	GCG	GC	GGTI	ATGG	A 31	760		
CGC	GCGAC	CT CT	CCCG	CGCG	CACC	GCA	TGA	GCCG	TCGC	CT	GAI	AAGO	CGG	T :	ICCG	TCT	TCG	TCAA	TAA	CTA	CA	ACGI	ACGG	C 31	840		
GAT.	ATGAC	CG TG	CCGT	IIGG	CGGC	TAT	AAG	CAGA	GCGG	CA	ACO	GGTC	GCG	GA (CAAA	TCCO	CTG	CATO	ccci	ΓTG	AA.	AAA	TCA	C 31	920		
rga.	ACTGA	AA AC	CATC	IGGA	TAAG	CCT	GGA	GGCC	TGA	AT	GAO	CCGA	ACA	AT 2	ACCA	GCA	GTT	ACTA	CGCC	CGC	CA	GTGO	GAA	T 32	2000		
AAA	TATGO	AC CA	TTCG	ACAC	ACTG	AAT	GAG	TCAA	TCAC	CT	GCO	GACO	TTT	G (CGTG	GTT	GGC	GGCG	GCTI	ATA	СС	GGGG	CTCT	C 32	2080		
CTC	CGCGC	TG CA	TCTG	GCGG	AAGC	GGG	CTT	TGAC	GTAG	TG	GTI	CTC	GAA	AG (CCTC	ACGO	CAT	CGGC	TTTC	GGC	GC.	AAGO	GGGG	C 32	2160		
GCA	ATGGC	GG GC	AGCT	TGTG	AACT	сст	ACA	GCCG	CGAC	ТА	CG7	ACGT	GAT	C (GAAA	AAA	GCT	ACGO	CATO	GA	C A	CCG	icce.	т 32	240		
Seque	ence Vie	wer																									
		<															,										•
Annot	tation																										
Feat	ture list					±	\otimes	Т							t	. 0		7									
		Feat	ure kev		Start	000	End		Len	ath		%GC		-	00	0 -		3047		57							
₹	1/1 20		CDS		20470		210	57	149		_	56.06	_			/a1	lele=	"26"									
	. 31		CDS		31050		3333	20	170	1		59.09			/1	locus	_tag=	"EC_6	805"								
-	32		CDS		33277		3454	42	126			57.71				/ev10	aence note=	=100.0) I•etar	t=26	234.	eton	-27722	·cid	denc	wo 1"	
•	GE 33		CDS		35806		3643	≠∠ 7.4	660			40.25			/tra	anslat	tion=	"MNFHI	ILAYWO	DKAL	SLA1	ENRL	TINGEY	ТААА	AENETE	TETVDP	v
 -	34		CDS		36528		3674	52	225			43.23						TQAP	AKIAR	GKSV	DIDF	AVSA	ARGVFE	RGDW	SLSSE	AKRKA	v
÷	À 35	- i i	CDS		36752		371	11	360			50.14						LNKL	ADLMEA	HAEE	LALI	ETLD	IGKPIR	HSLF	RDDIPG	GAARAI	R
	. 36		CDS		37120		3734	41	222			53.85						WYAE	AIDKVY	GEVA	TTSS	HELAN	4IVREP	VGVI	IAAIVE	WNFPL	L
÷	37	-	CDS		37416		377	30	315			50.96						VI.NV	TGFGH	HAGN EAGO	ALSE	NUDTI	DATAFT	GSTR	RTGKOT	TKDPC	D
	38				51 110		0.11		0.0			50.00	,					SNMK	RVWLEA	GGKS	ANIV	FADCI	DLQKA	ASAT	TAAGIE	YNQGQ	v
Z							0																				-
Annota	ation He	ader Cus	stom Field	Is Seq	uence S	earch	Contig	IS Fran	ne Analy	/SIS	Rest		Analys	sis j													
					_		_						_														

Figure 4: The Sequence editor window.

• Character experiment type wgMLST_CallTypes: contains details on the call types.

4 Installing the *E. coli* functional genotyping plugin

- 1. Call the *Plugins and Scripts* dialog box from the *Main* window by selecting *File* > *Install / remove plugins...* (,).
- 2. Select the *E. coli functional genotyping plugin* and press the *<Install*> button (see Figure 5).
- 3. Confirm the installation of the plugin.

During installation, the plugin downloads online knowledge bases from https://www.bionumerics.com, which requires a connection to the internet.

4. Click on < *Yes*> to review the settings.

The *E. coli genotyping settings* dialog box pops up, consisting of 8 tabs (see Figure 6).

In the General tab the following general settings need to be specified:

- **Included info fields**: In this list the entry information fields that will be displayed in the genotyping report can be specified.
- *Exports directory*: With < *Browse...*> you can specify an export directory to store all exports from the genotyping reports.



Figure 5: Install the plugin.

- *Input Sequence experiment*: From the drop-down list you can specify the sequence experiment that holds the (whole) genome sequences that will be screened.
- **Enabled features**: This list contains all offered features of the genotyping plugin. Features which are not required can be disabled in this list to save on processing time and omit the corresponding sections from the report. By default, all features are enabled.
 - In our demonstration database, the assembled sequences are stored in the *denovo* sequence experiment. Make sure this experiment is selected from the drop-down list and check the *Run* number to include in the report (see Figure 6).

The other tabs group the settings for each possible search: Resistance (Acquired/mutational resistance and resistance typing), Virulence (Acquired virulence and Virulence islands), PCR extraction, Serotype, Plasmid, Phage and Species confirmation.

All feature tabs contain a *Knowledgebase* and *Results* panel:

- Knowledgebase: in this panel the Version and Name of the knowledge base that is being used for this feature is shown. A different knowledge base version can be selected by pressing the < Change...> button. With Check for updates on startup checked, BIONUMERICS will check if a newer knowledge base version is available online for this feature each time the database is opened.
- Results: in this panel the output database information fields and experiments to which the screening results will be written can be dictated. Use the drop-down list to choose an existing experiment type or field, or the <*Create*> option to create new experiments and fields. A default name for the experiment or information field is suggested, but you can adjust this if you want to. Check *Annotate sequence experiment* to annotate the input sequence with the detected genotyping features.

In the *Resistance* tab there is an additional panel (*Resistance typing*) where you can specify the information fields to which the ESBL and CPE typing information should be stored.

coll genotyping se	ettings					ſ	
Plasmid			Phage	S	Species Cont	irmation	
General	Resistanc	e	Virulence	PCR ext	traction	Serotype	
he Escherichia coli Reporting	genotyping) model. nter Nam	e		•		
	☐ Ru ☐ Exp ☐ Loa ☐ Rel ☐ SR	n periment adDate leaseDat A Sample	e e		v		
Exports directory	C:\Use	rs\10029	961\Documents	BIONUMERICS	58.1\W(Browse	
Processing							
Input sequence exp	periment	denovo				\sim	
Enabled features		 ✓ Res ✓ Vin ✓ PCI ✓ PCI ✓ Ser ✓ Plas ✓ Pha 	sistance ulence R extraction otype smid uge			^	
					ОК	Cano	el

Figure 6: The Settings dialog: General tab.

In the *Virulence* tab there is an additional panel (*Virulence islands*) where you can specify the minimum percentage of virulence island loci that needs to be detected (*Minimum loci (%)*) before the presence of the virulence island is shown in the results.

In the *Results* panel of the *PCR extraction* tab you can click on *<Change...>* next to *Sequence extraction* to open the *Change sequence experiment* dialog box. For each PCR target in the list a new or existing sequence experiment type can be selected from the drop-down list.

Sourmash is used to screen the genome for the presence of reference plasmids. In the *Plasmids* panel of the *Plasmid* tab, two settings for the sourmash algorithm can be specified:

- *Min plasmid containment (%)* is the minimum containment score (expressed as a percentage) of a plasmid sequence in the target assembly sequence. If the minimum plasmid containment score is set to e.g. 95% and less than 95% of the plasmid sequence is contained in the query sequence, the plasmid will not be reported.
- *Min contig containment (%)* is the minimum containment score (expressed as a percentage) of a contig sequence in the detected plasmid sequence. If the minimum contig containment score is set to e.g. 95% and less than 95% of the contig sequence is contained in the detected plasmid sequence, the contig will not be reported.

For the BLAST-based searches (i.e. acquired resistance, resistance typing, acquired virulence, virulence islands, serotype determination, plasmid ori detection and phage detection) an additional

BLAST panel is available. In this panel two settings for the BLAST algorithm can be specified; the *Minimum identity (%)* and the *Minimum coverage (%)* of your query sequence against the knowledge base's reference sequences. If the option *Combine fragments* is checked, genes that occur fragmented in the genome (i.e. split over two contigs) can still be detected.

Please note that this panel is called Ori instead of BLAST in the Plasmid tab.

- In this tutorial, specify the experiment types and information fields in all tabs by selecting the <*Create*> option in the drop-down lists and accepting the default names. Leave the other settings unaltered.
- 7. In the *Results* panel of the *PCR extraction* tab click on <*Change...*> next to *Sequence extraction* to open the *Change sequence experiment* dialog box. In the drop-down list of the *e_coli-det* PCR target select the <*Create*> option and accept the default name. Click on <*OK*> to close the *Change sequence experiment* dialog box (see Figure 7).

_					
1	Change PCR extraction Seque	nce experiment		?	×
	Name	Sequence experiment			^
	stx2d_F1_R1	<none></none>		\sim	
	stx2d_F1_055-R	<none></none>		\sim	
	stx2d_F1_R2	<none></none>		\sim	
	stx2e	<none></none>		\sim	
	stx2f	<none></none>		\sim	
	stx2g	<none></none>		\sim	
	eaeA	<none></none>		\sim	
	ehxA	<none></none>		\sim	
	rpoB-AEM	<none></none>		\sim	
	rpoB-CDC	<none></none>		\sim	
	іраН	<none></none>		\sim	
	e_coli-det	e_coli-det		\sim	
	e_albertii-det	<none></none>		\sim	
	e_fergusonii-det	<none></none>		\sim	
					¥
			ОК		Cancel

Figure 7: The Change sequence experiment dialog box.

- 8. Click on *<OK>* in the *E. coli genotyping settings* dialog box.
- 9. When the *E. coli functional genotyping plugin* is successfully installed, a confirmation message pops up. Press <*OK*>.
- 10. Press < *Close* > to close the *Plugins and Scripts* dialog box.
- 11. Close and reopen the database to activate the features of the *E. coli functional genotyping plugin*.

The *E. coli functional genotyping plugin* installs menu items in the main menu of the software under *E. coli* (see Figure 8).



The settings specified during installation of the plugin can be called again at any time with *E. coli* > *Settings...*.

5 Screening of entries

The screening can be done on any selection of entries in the database.

1. Select a single entry in the *Database entries* panel by holding the **Ctrl**-key and left-clicking on the entry. Alternatively, use the **space bar** to select a highlighted entry or click the ballot box next to the entry.



Figure 8: New menu-items after installation of the *E. coli functional genotyping plugin.*

Selected entries are marked by a checked ballot box () and can be unselected in the same way.

2. In order to select a group of entries, hold the Shift-key and click on another entry.

A group of entries can be unselected the same way.

3. Make sure a few entries are selected in the *Database entries* panel of the demonstration database.

Screening for the phenotypic traits can be done for all tools checked in the *E. coli genotyping* settings dialog box (*E. coli* > *Analysis* > *All Enabled*) or for each tool separately (*E. coli* > *Analysis* > *...*).

4. Select *E. coli* > *Analysis* > *All Enabled* to screen the selected entries for all enabled traits.

A progress bar appears. The analysis time depends on the number of selected entries. When the analysis is finished, the progress bar disappears. The detected traits for the screened entries are stored in the database.

The predicted **CPE** and **ESBL** resistance types, **Total islands**, **Pathotype**, **H-antigens**, **O-antigens** and **Species confirmation** are written to the information fields in the *Database entries* panel (see Figure 9). Please note that the shown names of the information fields are those created per default, but can be different in your case depending on whether you have chosen an alternative name during installation.



Figure 9: Example output of the CPE, ESBL, Total islands, Pathotype, Hantigens, O-antigens and Species confirmation information fields.

The character experiment types for **Resistance**, **Virulence**, **PCR extraction**, **Plasmid** and **Phage** detection are created and updated with the predicted traits. Please note that the shown names of the experiment types are those created per default, but can be different in your case depending on whether you have chosen an alternative name during installation.

5. Open a character card for one of the analyzed entries by clicking on the corresponding green colored dot in the *Experiment presence* panel.



The characters in the characters experiments are displayed in the same order they are listed in their knowledge base. However, it might be more convenient for interpretation to have them displayed alphabetically. This can be done in the *Character type* window with the option *Characters* > *Arrange characters by field...* (\downarrow).

Below, the interpretation of the results gathered in the character experiment types is given.

Acquired resistance (see Figure 10):

- **Resistance_traits**: contains the results for each antibiotic: 0 = not detected (sensitive), 1 = detected (resistant).
- **Resistance_loci**: contains the results for each resistance gene: 0 = not detected (sensitive), when detected (resistant) the % identity of the best hit is shown.

EC_0000001				EC_0000001			
Character	Value	Mapping	-	Character	Value	Mapping	-
Virginiamycin S	0	<->	^	dfrA29	0	<->	^
Virginiamycin M	0	<->		dfrA18	0	<>	
Unknown Tetracycline	1	<+>		dfrA9	0	<->	
Unknown Rifamycin	1	<+>		dfrA13	0	<>	
Unknown Phenicol	1	<+>		dfrA26	0	<>	
Unknown Macrolide	1	<+>		dfrA25	0	<>	
Unknown Fluoroquin	1	<+>		dfrA23	0	<>	
Unknown Aminoglyc	1	<+>		dfrA30	0	<>	
unknown	0	<->		erm(44)v	0	<>	
Trimethoprim	0	<->		msr(E)	0	<>	
Tobramycin	0	<->		blaIMI-5	0	<>	
Tobramycin	0	<->	~	blaVIM-38	0	<->	~
Press Insert to add chara	acter			Press Insert to add	I character		

Figure 10: Example output of the *Resistance_traits* and the *Resistance_loci* experiment types for sample EC_0000001.

Mutational resistance (see Figure 11):

• **Resistance_mutations**: contains the results for each resistance mutation: -2 = partially indecisive, -1 = fully indecisive, 0 = not detected (sensitive), 1 = detected (resistant).

Character	Value	Mapping	•
gyrA_pA51V	0	<->	^
gyrA_pA67S	0	<->	
gyrA_pG81C	0	<->	
gyrA_pG81D	0	<->	
gyrA_pD82G	0	<->	
gyrA_pS83L	0	<->	
gyrA_pS83W	0	<->	
gyrA_pS83A	0	<->	
gyrA_pS83V	0	<->	
gyrA_pA84P	0	<->	
gyrA_pA84V	0	<->	
gyrA_pD87N	0	<->	~

Figure 11: Example output of the *Resistance_mutations* experiment type for sample EC_0000001.

Acquired virulence (see Figure 12):

Virulence_loci: contains the results for each virulence gene: 0 = not detected, when detected the % identity of the best hit is shown.

EC_000001				EC_0000001			
Character	Value	Mapping	-	Character	Value	Mapping	-
adherence	1	<+>	^	yfcV	0	<->	^
Type VI translocated	0	⇔		virF	0	<->	
iron uptake	1	<+>		vat	0	<->	
[not specified]	0	<->		usp	0	<->	
survival	1	<+>		tsh	0	<->	
toxin	1	<+>		traT	100	<+>	
antiphagocytosis	0	<->		toxB	0	<->	
protease	0	<->		tir	0	<->	
complement protease	1	<+>		terC	100	<+>	
type III translocated	0	<->		tcpC	0	<->	
type II translocated p	0	<->		tccP	0	<->	
invasion	1	<+>	~	subA	0	<->	~
Press Insert to add chara	acter			Press Insert to add	character		

Figure 12: Example output of the *Virulence_loci* and the *Virulence_traits* experiment types for sample EC_0000001.

• *Virulence_traits*: contains the results for each virulence type: 0 = not detected, 1 = detected.

Virulence islands (see Figure 13):

- *island_counts*: contains the number of detected loci associated to a pathogenicity island.
- island_percentages: contains the percentage of detected loci associated to a pathogenicity island.

EC_0000001				EC_0000001			
Character	Value	Mapping	-	Character	Value	Mapping	-
PALIII	0	<->	^	PALIII	0	<->	^
SE-PAI	0	<->		SE-PAI	0	<->	
PALV	0	<->		PALV	0	<->	
LIM	0	<->		LIM	0	<->	
OI-122	0	<->		OI-122	0	<->	
PALIV	0	<->		PAIN	0	<->	
HPI	0	<->		HPI	0	<->	
ETT2	25	<+>		ETT2	68	<+>	
espC PAI	0	<->		espC PAI	0	<->	
LEE	0	<->		LEE	0	<->	
PALL	0	<->		PALL	0	<->	
LEE II	0	<->	~	LEE II	0	<->	~
Press Insert to add c	haracter			Press Insert to add	l character		

Figure 13: Example output of the *island_counts* and the *island_percentages* experiment types for sample EC_0000001.

PCR extraction (see Figure 14):

 PCR extraction_amplicons: contains the results for each in silico PCR: 0 = no amplicon, 1 = amplicon generated.

Plasmid detection (see Figure 15):

- **Plasmid**: contains the results of the full plasmids detection: 0 = not detected, when detected the % containment of the detected plasmid is shown.
- **Ori**: contains the results of the plasmid ori detection: 0 = not detected, when detected the % BLAST identity with the ori reference sequence is shown.

Phage detection (see Figure 16):

EC_000001			
Character	Value	Mapping	-
e_fergusonii-det	0	<->	^
e_albertii-det	0	<->	
e_coli-det	1	<+>	
іраН	0	<->	
rpoB-CDC	0	<->	
rpoB-AEM	1	<+>	
ehxA	0	<->	
eaeA	0	<->	
stx2g	0	\Leftrightarrow	
stx2f	0	<->	
stx2e	0	\Leftrightarrow	
stx2d_F1_R2	0	\Leftrightarrow	~
Press Insert to add chara	icter		

Figure 14: Example output of the *PCR extraction_amplicons* experiment type for sample EC_0000001.

EC_000001				EC_000001			
Character	Value	Mapping	-	Character	Value	Mapping	-
NZ_WYDM02000007	0	<->	^	repB_KLEB_VIR	0	<->	^
NZ_WYDM02000006	0	<->		repA_3	0	<->	
NZ_WYDM02000005	0	<->		repA_2	0	<->	
NZ_WXZA01000057	0	<->		repA_1	0	<->	
NZ_WXZA01000048	0	<->		IncFIB(H89-PhagePl	0	<->	
NZ_WXYZ01000005	0	<->		IncFIB(pN55391)	0	<->	
NZ_WXYY01000004	0	<->		Col(pHAD28)	0	<->	
NZ_WXYX01000005	0	<->		Incl	0	<->	
NZ_WXYX01000004	0	<->		Col440II	0	<->	
NZ_WXYW01000006	0	<->		Col440I	0	<->	
NZ_WXYW01000004	0	<->		FIA(pBK30683)	0	<->	
NZ_WXYV01000007	0	<->	~	FII(pBK30683)	0	<->	~
Press Insert to add char	acter			Press Insert to add char	acter		

Figure 15: Example output of the *Plasmid* and the *Ori* experiment types for sample EC_0000001.

- **Phage_seq_ids**: contains the results of the phages detection by sequence IDs: 0 = not detected, when detected the % of the detected full phage is shown.
- **Phage_categories**: contains the results of the phages detection by phage categories: 0 = not detected, when detected the % of the detected full phage is shown.

EC_000001					EC_000001			
Character	Value	Mapping		•	Character	Value	Mapping	-
Escherichia virus phiX174	0	<->		^	Sinsheimervirus	0	<->	^
Escherichia virus phiEco32	0	<->			unclassified Tevenvi	0	<->	
Escherichia virus TLS	0	<->			unclassified Tevenvi	0	<->	
Escherichia virus T5	0	<->			Ravinvirus	0	<->	
Escherichia virus T4	0	<->			Inovirus	0	<->	
Escherichia virus SSL2009a	0	<->			Lineavirus	0	<->	
Escherichia virus Rtp	0	<->			Allolevivirus	0	<->	
Escherichia virus RB43	0	<->			Levivirus	0	<->	
Escherichia virus RB16	0	<->			Peduovirus	0	<->	
Escherichia virus P2	0	<->			Gaprivervirus	0	<->	
Escherichia virus P1	0	<->		~	Hendrixvirus	0	<->	
Escherichia virus N4	<		>		Vequintavirus	0	<->	~
Press Insert to add character					Press Insert to add chara	acter		

Figure 16: Example output of the *Phage_seq_ids* and *Phage_categories* experiment types for sample EC_0000001.

- 6. Close the character card(s) by clicking in the top left corner of the card.
- 7. Open the **PCR extraction_amplicons** character card for one of the analyzed entries by clicking on the corresponding green colored dot in the *Experiment presence* panel.

The **PCR extraction_amplicons** character card (see Figure 17) lists all *in silico* PCR sequences that passed the search criteria.

8. Close the character card by clicking in the top left corner of the card.

Character	Value	Mapping	•
e_fergusonii-det	0	<->	^
e_albertii-det	0	<->	
e_coli-det	1	<+>	
ipaH	0	<->	
rpoB-CDC	0	<->	
rpoB-AEM	1	<+>	
ehxA	0	<->	
eaeA	0	<->	
stx2g	0	<->	
stx2f	0	<->	
stx2e	0	<->	
stx2d_F1_R2	0	<->	~

Figure 17: Example output of the PCR extraction_amplicons experiment type for sample EC_0000001.

The predicted *In silico* PCR sequences are stored in the corresponding sequence type experiments if these experiments have been created in the *E. coli genotyping settings* dialog box.

Click on the green colored dot of the e_coli-det sequence experiment for the entry with Key EC_0000001. The Sequence editor window opens and displays the extracted sequence (see Figure 18).

EC_0000001 (Sequence Viewer)	-	×
File Sequence Header Annotation View Tools Window Help		
🗒 🖄 ♎ ♫ ♫ ♫ ㉓ ㉓ ⊗ Q] ▷ ▽1 ▽2 ▽▷ ⊗		
Sequence Editor		
ccaggcaaag agtttatgtt gaataaaaaa atgttggcat gattttagta attccaactg ctctgaaaat aattgccaat gttgctcctc ggtaagttgc gctataaccc gactcgtcgg gatacgcacc gtgccatctt cactggagaa atgggtaatt aactcaacgc cgacaaggac ttgctggttg tctcttatcg gcaggaaata gc	60 120 180	
Sequence Viewer		
<		>
Annotation		
Feature list \bigcirc \pm \otimes \pm \otimes		
→ W Feature key Start End →		
Annotation Header Custom Fields Sequence Search Contigs Frame Analysis Restriction Analysis		
Sequence: EC_0000001 Experiment: e_coli-det 1 212 bp		.::

Figure 18: Example output of the **e**_**coli-det** experiment type for sample EC_0000001.

10. Close the Sequence editor window.

1. Open the genotype report for the selected entries with *E. coli* > *Reports...*.

The *Report* window contains a genotype report for each of the selected entries (see Figure 19).

2. Select another entry in the *Entries* panel to update the results in the *Genotype report* panel.

The creation date of the report (*Date*), the Key (*Name*), and information fields checked in the *E. coli genotyping settings* dialog box are displayed in the *Genotype report* panel.

🖆 Genotyping E. coli reports		- 0	×
File Entries Report Window Help			
Entries	E. coli - functional genotyping report		
Кеу 🗸	1. Summary		
C_0000001	·		
C_0000002	EC 000001		
C_0000003	EC_0000001		
	E. coli functional genotyping repo	ort	
	The contents of this report are fo See below for full disclaimer.	r research purposes only and not intended for clinical decision making.	
	Date:	2021-09-27	
	Name:		
	Run:	SRR3242188	
	Acquired resistance :	Unknown Aminoalvcoside	
	,	Unknown Fluoroquinolone	
		Unknown Macrolide	
		Unknown Phenicol	
		Unknown Rifamycin	
	Mutational resistance :	Unknown letracycline	
	FSBI ·	False	
	CPE :	False	
	Acquired virulence :	adherence	
		complement protease	
		invasion	
		iron uptake	
		regulation	
		toxin	
	Virulence islands :	ETT2	
	Pathotypes :	STEC	
	PCR extraction :	e_coli-det	
		rpoB-AEM	
		stx2-det_F4_R1	
	Serotune :	SIX28_F2_R3 088:H19	
	Ori :	IncEIB(AP001918)	
		IncFII	
		Incl2(Delta)	
	plasmids :	NZ_CP010124	- 11
		NZ_CP014271	
		NZ_CP030301 NZ_CP043415	
		NZ_RRSZ01000123	
	Phage :	Lambdavirus	
	0	Uetakevirus	
			-

Figure 19: Functional genotyping report.

3. Select *Report styles* in the *Report* window and make sure the option *Summary* is selected.

A summary of the results of all analyzed traits is displayed in the *Report* window.

4. Select *Report* > *Report styles* in the *Report* window and select the option *Complete* (see Figure 20).

In the *Complete* view, the summarized results as well as all available details are shown. All hits that passed the settings for *Acquired Resistance*, *Mutational Resistance*, *ESBL*, *CPE*, *Acquired Virulence*, *Virulence Islands*, *Pathotypes*, *PCR extraction*, *Serotype*, *Ori*, *Plasmids*, *Phage* and *Species confirmation* screening are listed and described.

E. coli - functional genotyping report		
1. Si	ummary	
~	1. Summary	nce :
	2. Default	
	3. Complete	

Figure 20: Report styles in the *Report window*.

- 5. Click on a hyperlink of one of the enabled features to display the detailed results in the *Genotype report* panel (see Figure 21).
- 6. Select *File* > *Exit* to close the *Report* window.

For more detailed information on the genotyping analyses and interpretation of the reported results, please check the genotyping plugin manual.



Figure 21: Report details.