

The Bacterial Analysis Pipeline and batch upload

Learning objective:

After this lecture and exercise, you should be able to...

...upload batches of data to the CGE Bacterial Analysis Pipeline, browse the results, and re-run with changed settings for individual services.

A Bacterial Analysis Platform: An Integrated System for Analysing Bacterial Whole Genome Sequencing Data for Clinical Diagnostics and Surveillance

Martin Christen Frølund Thomsen^{1*}, Johanne Ahrenfeldt¹, Jose Luis Bellod Cisneros¹, Vanessa Jurtz¹, Mette Voldby Larsen¹, Henrik Hasman², Frank Møller Aarestrup², Ole Lund¹

¹ Department of Systems Biology, Technical University of Denmark, Kemitorvet Building 208, 2800 Kgs. Lyngby, Denmark, ² National Food Institute, Technical University of Denmark, Søltofts Plads Building 221, 2800 Kgs. Lyngby, Denmark

* mcft@cbs.dtu.dk

Abstract

Recent advances in whole genome sequencing have made the technology available for routine use in microbiological laboratories. However, a major obstacle for using this technology is the availability of simple and automatic bioinformatics tools. Based on previously published and already available web-based tools we developed a single pipeline for batch uploading of whole genome sequencing data from multiple bacterial isolates. The pipeline will automatically identify the bacterial species and, if applicable, assemble the genome, identify the multilocus sequence type, plasmids, virulence genes and antimicrobial resistance genes. A short printable report for each sample will be provided and an Excel spreadsheet containing all the metadata and a summary of the results for all submitted samples.

Bacterial Analysis Pipeline

Service Overview

- * Assembly
- * Species/Lineage prediction
- * Multilocus Sequence Typing
- * Resistance Gene Identification
- * Plasmid Identification
- * Plasmid MLST
- * Virulence Gene Identification

KmerFinder-2.1 - Results

Hit	Score	z-score	Query Coverage [%]	Template Coverage [%]	Depth	Total Query Coverage [%]	Total Template Coverage [%]	Total Depth
Escherichia coli, Escherichia coli O157:H7, Escherichia coli O157:H7 str. Sakai get sequence	11105	543.7	100.00	100.00	1.00	100.00	100.00	1.00

MLST-1.6 Server - Typing Results

Sequence Type: ST-11

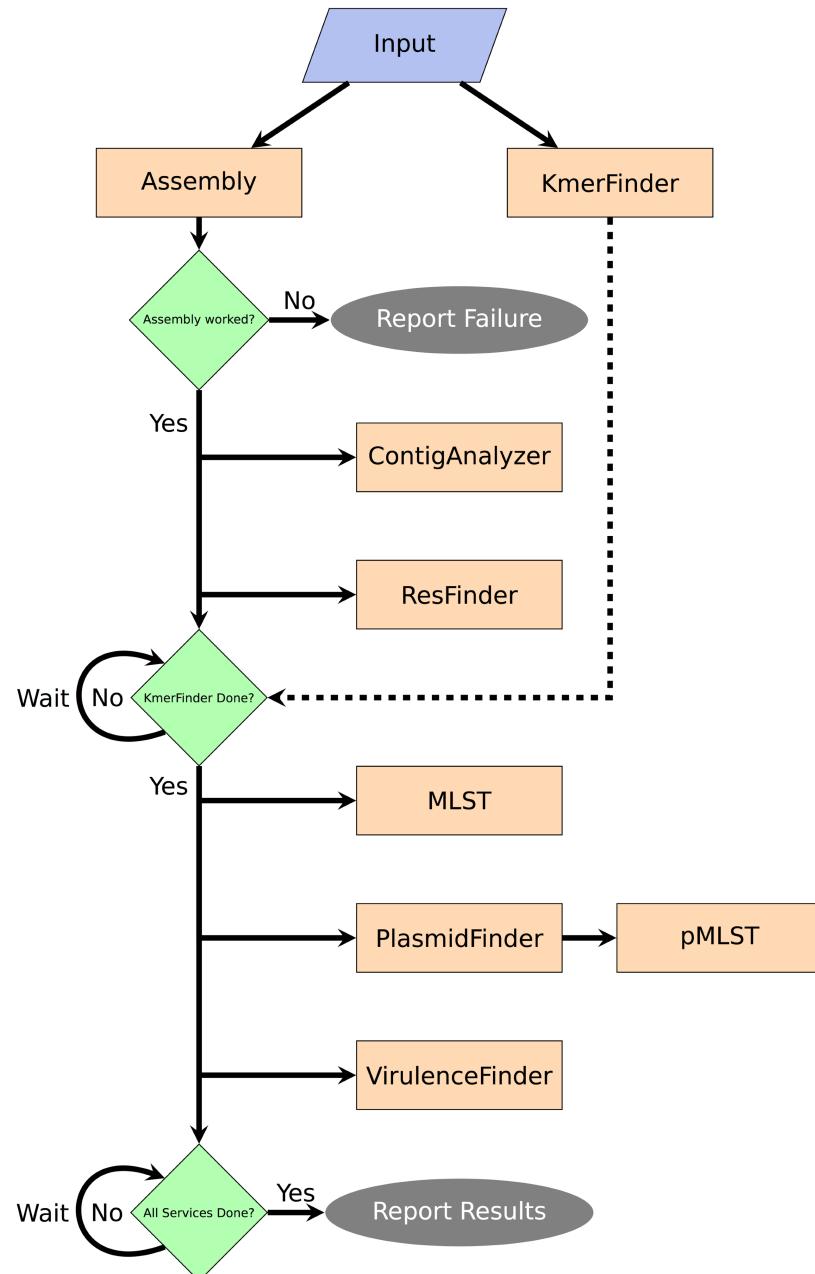
Locus	% Identity	HSP Length	Allele Length	Gaps	Allele
<i>adk</i>	100.00	536	536	0	<i>adk-12</i>
<i>fumC</i>	100.00	469	469	0	<i>fumC-12</i>
<i>gyrb</i>	100.00	460	460	0	<i>gyrb-8</i>
<i>icd</i>	100.00	518	518	0	<i>icd-12</i>
<i>mdh</i>	100.00	452	452	0	<i>mdh-15</i>
<i>pura</i>	100.00	478	478	0	<i>pura-2</i>
<i>reca</i>	100.00	510	510	0	<i>reca-2</i>

Virulence - E. coli

Virulence factor	% identity	Query/HSP length	Contig	Position in contig	Protein function	Accession number
<i>tccP</i>	99.78	894 / 894	gil15829254 reflNC_002695.1I	1209251..1210143	Tir-cytoskeleton coupling protein	AB253557
<i>stx2A</i>	100.00	960 / 960	gil15829254 reflNC_002695.1I	1266965..1267924	Shiga toxin 2, subunit A, variant a	AB048837
<i>stx2B</i>	100.00	270 / 270	gil15829254 reflNC_002695.1I	1267936..1268205	Shiga toxin 2, subunit B, variant a	AE005174
<i>iss</i>	97.96	294 / 294	gil15829254 reflNC_002695.1I	1274890..1275183	Increased serum survival	CP001509

Introducing the pipeline

1. Input sequence file(s) and metadata
2. Create DB entry
3. Run pipeline of services
4. Create analysis report and update DB entry



1. Create a user account and login

<https://cge.cbs.dtu.dk/services/>

The screenshot shows the homepage of the Center for Genomic Epidemiology. At the top right, there is a red login box containing fields for 'Username' and 'Password', and buttons for 'New', 'Reset', and 'Login'. A black circle highlights the 'Login' button. Below the login box, there is a navigation bar with three tabs: 'Home', 'Services', and 'Datasets'. The 'Services' tab is currently active. Under the 'Services' tab, there is a section titled 'Overview of Services' which includes links for 'Workflows' (Bacterial Analysis Pipeline (Batch Upload) (Works)), 'Phenotyping' (ResFinder, PathogenFinder, VirulenceFinder, Restriction-ModificationFinder), and 'Typing' (MLST, pMLST, PlasmidFinder, KmerFinder, SpeciesFinder, Reads2Type, Tapir (OBS! Only works for Firefox), TaxonomyFinder, SerotypeFinder, spaTyper, FimTyper, PAst).

Center for Genomic Epidemiology

Home Services Datasets

Overview of Services

Workflows

[Bacterial Analysis Pipeline \(Batch Upload\) \(Works\)](#)

Phenotyping

[ResFinder](#)
[PathogenFinder](#)
[VirulenceFinder](#)
[Restriction-ModificationFinder](#)

Typing

[MLST](#)
[pMLST](#)
[PlasmidFinder](#)
[KmerFinder](#)
[SpeciesFinder](#)
[Reads2Type](#)
[Tapir \(OBS! Only works for Firefox\)](#)
[TaxonomyFinder](#)
[SerotypeFinder](#)
[spaTyper](#)
[FimTyper](#)
[PAst](#)

User Home

Services

User Overview

Create User Profile

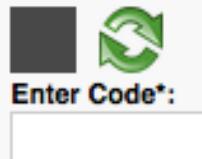
Username*:

Email*:

Password*:

Security: **None**

Confirm Password*:



Support

Technical problems

Copyright DTU 2011 / All rights reserved

Center for Genomic Epidemiology, DTU, Kemitorvet, Building 204, 2800 Kgs. Lyngby, Denmark

Funded by: The Danish Council for Strategic Research

Last modified October 26, 2014 02:42:33 GMT

Center for Genomic Epidemiology

Welcome metteb@cbs.dtu.dk

Map Overview

Batch Uploader

Sample

Manager

Run Analysis

Settings

Log out

Home

Services

Overview of Services

Workflows

[Bacterial Analysis Pipeline \(Batch Upload\)](#) (Works)

Phenotyping

[ResFinder](#)

[PathogenFinder](#)

[VirulenceFinder](#)

[Restriction-ModificationFinder](#)

Typing

[MLST](#)

[pMLST](#)

[PlasmidFinder](#)

[KmerFinder](#)

[SpeciesFinder](#)

[Reads2Type](#)

[Tapir](#) (OBS! Only works for Firefox)

[TaxonomyFinder](#)

[SerotypeFinder](#)

[spaTyper](#)

[FimTyper](#)

[PAst](#)

2. Download and fill out metadata sheet

Center for Genomic Epidemiology

Welcome
metteb@cbs.dtu.dk

Home

Services

Instructions

Example

Article abstract

Bacterial Analysis Pipeline - Batch Upload

The CGE Bacterial Analysis Pipeline executes a workflow of services with predefined parameters and stores the submitted data and result in the database at the user's disposal.

[View the version history](#) of this server.

STEP 1 [Download Metadata Template](#) Template 

STEP 2: Fill out template

STEP 3: [Upload Metadata File](#)

STEP 4: [Select Files](#) 

STEP 5: [Submit](#)

Supported browsers

	 Windows XP	 Windows Vista	 Windows 7	 Windows 8	 OS X Yosemite	 Ubuntu
 ≥ V 8.0						✓
 V 10.0	✓		✓			✓
 edge	✓		✓			✓
 ≥ V 36.0				✓	✓	✓
 ≥ V 41.0	✓	✓	✓	✓	✓	✓

Progress Overview

Name	Size	Progress	Status

The Excel file has two sheets:

1: Empty table into which you should add your metadata

The Excel file has two sheets:

1: Empty table into which you should add your metadata

2: A description of all the fields in the table

Attribute Description

Field Name	Mandatory	Description
sample_name	No	The name of the isolate the user uses to identify the sample
file_names	Yes	Name of all files associated to this sample. Multiple filenames should be seperated by a space
pre_assembled	Yes	Has the uploaded sample data been assembled? yes / no
sequencing_platform	Conditional	Choose between: LS454, Illumina, Ion Torrent or ABI SOLiD. Required if the sample is not pre_assembled.
sequencing_type	Conditional	Choose between: single, paired or mate-paired. Required if the sample is not pre_assembled.
organism	No	Scientific name of the organism. If left blanc default is 'unknown'.
strain	No	<i>Strain type ID</i>
subtype	No	Subspecific genetic linage, i.e MLST, serovar and biotype
country	Yes	
region	No	
city	(No)	
zip_code	(No)	Please provide as much information as possible. Low resolution locations reduce the usability. Our recommendation is to provide either city, zip_code or longitude and latitude coordinates.
longitude	(No)	
latitude		
location_note	No	Additional relevant details about the location
isolation_source	Yes	Choose between: human, water, food, animal, laboratory, other or unknown.
source_note	No	Additional relevant details about the isolation source. i.e. blood, laboratory experiment or urine
pathogenic	No	Is the organism decreed pathogenic? [yes, no, unknown]. If left blanc default is 'unknown'.
pathogenicity_note	No	Additional relevant details about the organism's pathogenicity
collection_date	Yes	The date of the sample collection. Use one of the following format: YYYY-MM-DD, YYYY-MM or YYYY
collected_by	No	Name of the institute or person who took the sample
notes	No	Any additional information can be added

3. Upload filled-out metadata sheet and sequence files

Center for Genomic Epidemiology

Welcome
metteb@cbs.dtu.dk

Home

Services

Instructions

Example

Article abstract

Bacterial Analysis Pipeline - Batch Upload

The CGE Bacterial Analysis Pipeline executes a workflow of services with predefined parameters and stores the submitted data and result in the database at the user's disposal.

[View the version history](#) of this server.

STEP 1: [Download Metadata Template](#) Template

STEP 2: Fill out template

STEP 3: [Upload Metadata File](#)

STEP 4: [Select Files](#)

STEP 5: [Submit](#)

Supported browsers

	Windows XP	Windows Vista	Windows 7	Windows 8	OS X Yosemite	Ubuntu
≥ V 8.0						✓
V 10.0			✓	✓		
edge			✓	✓		
≥ V 36.0				✓	✓	✓
≥ V 41.0	✓	✓	✓	✓	✓	✓

Progress Overview

Name

Size

Progress

Status

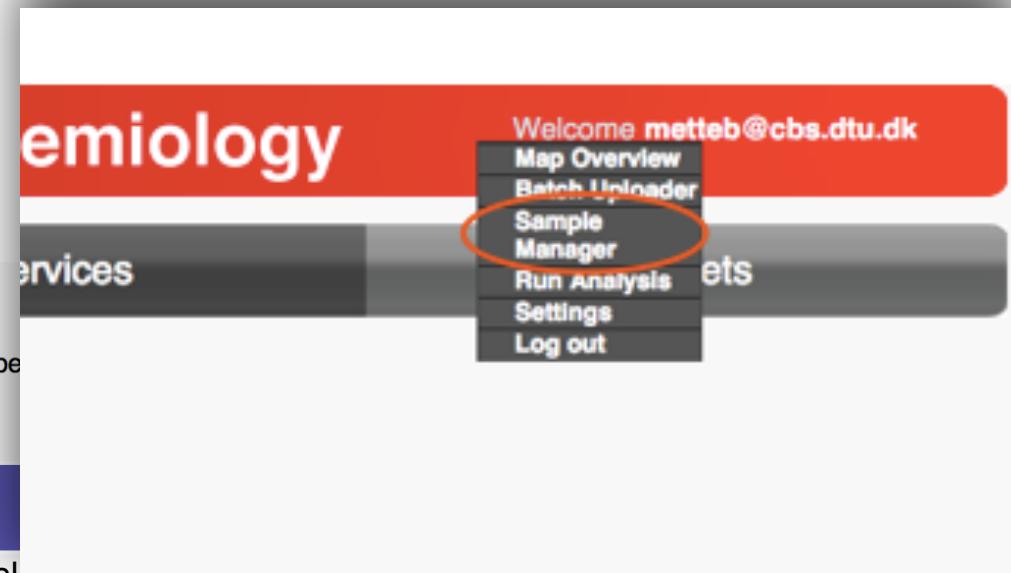
Results Summary

Result Summary:

Following table provides a summary of the results found by the different services in the pipeline. To manage your samples, and to get a full list of results go to [the sample manager](#).

Sample Name	Species	MLST	Plasmids	Antibiotic Resistances	Genes
Test plasmids	Escherichia coli	ST-44	Col(MGD2) Col(MG828) ColRNAI	IncF[F31:A4:B1]	aadA5 strB aac(3)-IIa mph(A) sul1 sul2 dfrA17 tet(B) catB3 blaCTX-M-15 blaOXA-1 aac(6')lb-cr

Your samples were uploaded successfully



Sample manager

1. Single service results
2. Download files
3. Download metadata and result data
4. Rerun of services

Sample Overview

Download all data in an Excel spreadsheet?

Download the resistance data in a more detailed Excel spreadsheet format?

Analyse selected samples Delete selected samples

	<input type="checkbox"/>	Name	Date	Country	City	Origin	Action		
-	<input type="checkbox"/>	Test plasmids	2014-02-28	Denmark	Lyngby	human	Download Analyse Edit Remove		
<hr/>									
	<input type="checkbox"/>	Service	Date	Status	Action				
	KmerFinder-2.1	2016-04-08	Success	Remove					
	ResFinder-2.1	2016-04-08	Success	Remove					
	ContigAnalyzer-1.0	2016-04-08	Success	Remove					
	PlasmidFinder-1.2	2016-04-08	Success	Remove					
	MLST-1.6	2016-04-08	Success	Remove					
	VirulenceFinder-1.2	2016-04-08	Success	Remove					
	pMLST-1.4	2016-04-08	Success	Remove					
<hr/>									
+	<input type="checkbox"/>	unknown_3	2016-01-01	Denmark		unknown	Download Analyse Edit Remove		
+	<input type="checkbox"/>	unknown_2	2016-01-01	Denmark		unknown	Download Analyse Edit Remove		
+	<input type="checkbox"/>	unknown_1	2016-01-01	Denmark		unknown	Download Analyse Edit Remove		

Analyse selected samples

Delete selected samples

Re-running individual services with changed settings

[Home](#)[Services](#)[Instructions](#)[Output](#)[Article abstract](#)

Bacterial Analysis Pipeline - Service Submission

Write the IDs of the samples here (comma separated!):

- Reassemble reads?
- Reanalyze contigs?
- Rerun Species Identification?
- Rerun Plasmid Identification?
- Rerun Multilocus Sequence Typing?
- Rerun Resistance Gene Identification?
- Rerun Virulence Gene Identification?
- Rerun Pathogen Prediction?

[Submit](#)[Clear fields](#)

NEW: Check out our available beta test services [here](#)

Bacterial Analysis Pipeline - Service Submission

OBS! This is our selection of services currently running as beta tests. To access our stable services [click here](#).

Write the IDs of the samples here (comma separated!): 90481

Single Sample Analyses:

Assembly and Input Analysis

- Assemble reads? (Assembler)
- Analyze contigs? (ContigAnalyzer)
- Analyze input data? (InputAnalyzer)
- Map your reads against available databases (MGmapper)

Species Identification and Subtyping

- Identify Species? (KmerFinder)
- Identify Plasmids and pMLST? (PlasmidFinder & pMLST)
- Identify Multilocus Sequence Type? (MLST)
- Identify Serotype (*E. coli*)? (SerotypeFinder)
- Identify Serotype (*Salmonella*)? (SeqSero)
- Identify Serotype (*Salmonella*)? (SalmonellaTypeFinder)
- Identify spa type (*S. aureus*)? (spaTyper)
- Identify Fim type? (FimTyper)

Multi Sample Analyses:

Phylogeny

- Create SNP tree? (NDtree)

Where to get support

Common problem

The file name contains characters other than

a-z

A-Z

0-9

-

-

.

Specifically, spaces are NOT allowed!

cgehelp@cbs.dtu.dk



Support

Scientific problems

Technical problems

What is the price for using the pipeline?

- In the near future, uploaded data will be shared (uploaded to European Nucleotide Archive - ENA) to aid the whole community
- It will be possible to choose a waiting period, for instance 1 year, before data is uploaded to ENA
- Data uploaded to the single-isolate, single-services will not be shared



Illumina BaseSpace

- Cloud-based computing environment for next-generation sequencing (NGS) data management and analysis
- Currently ~70-80 apps
- Mostly by Illumina or BaseSpace, but also from 3rd party developers
- Apps we know: FASTQC, Velvet, SPAdes
- In review: SerotypeFinder (*E. coli* serotyping), Bacterial Analysis Pipeline

BaseSpace Sequence Hub is a cloud-based genomics analysis and storage platform that directly integrates with all Illumina sequencers.

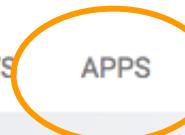
[Learn More](#)

[Sign up](#)

[Log in](#)

- Set up runs for Illumina sequencers and NeoPrep
- Monitor sequencing runs from the web or with the [iOS app](#)
- Stream data to the cloud directly from sequencers
- Analyze genomics data with pre-configured pipelines
- Share data instantaneously with anyone in the world

<https://basespace.illumina.com/home/index>



Dashboard Personal

Your FREE TRIAL has expired



Basic Tier

Usage

iCredits

subscription required

Ready for more? Upgrade to a Professional subscription.

[Learn More](#)

Storage

[Details](#)



8.0 GB

Notifications



Share accepted

Jan 9, 2017

You accepted Ilya Chorny's invitation to the Project MiSeq: Shotgun Met...



Share accepted

Jan 9, 2017

You accepted Lijun Liu's invitation to the Run 2x151PhiX



Applications



16S Metagenomics
Illumina, Inc.



Amplicon DS
Illumina, Inc.



Bacterial Analysis Pipeline
GoSeqIt ApS



BWA Aligner
BaseSpace Labs



BWA Enrichment
Illumina, Inc.



Cancer Variant Caller
Samsung SDS



ChIPSeq
BaseSpace Labs



CrossMap
BaseSpace Labs



Cufflinks Assembly & DE
Illumina, Inc.



DeepChek®-HBV
ABL S.A.



DeepChek®-HCV
ABL S.A.



DeepChek®-HIV
ABL S.A.



DESeq2
BaseSpace Labs



ED Annotator
EONE-DIAGNOMICS Genome Center



eGB
ELASTIC GENOME BROWSER



Enrichment
Illumina, Inc.



Bacterial Analysis Pipeline v1.0.0

GoSeqIt ApS

Analysis Name:

Bacterial Analysis Pipeline 01/13/2017 10:15:00

Input file:

Select File(s):



Save Results To:

Select Project(s):

ResFinder Specific Settings

Minimum %ID:

90



Minimum Length:

0.5



PlasmidFinder Specific Settings

Minimum %ID:

80



VirulenceFinder Specific Settings

Minimum %ID:

90



Disclaimer

- I acknowledge and agree that I have read and understood the End User License Agreement (EULA) related to this app. Specifically, this app is provided AS-IS, without warranty of any kind. The app is for research use only.

Bacterial Analysis Pipeline Version 1.0

pneu2.fa

[Export Summary \(TSV\)](#)

Basic Statistics

Genome Size (bp)	Number of Contigs	N50 ¹
5631011	175	3651106

¹The N50 value is the size of the shortest contig in the set of longest contigs that together cover at least 50% of the total Genome Size

Species

Klebsiella pneumoniae

[Download Detailed Species Results](#)

[Explanations, Detailed Species Results](#)

Multilocus Sequence Type (MLST)

MLST is performed for all species that according to <http://pubmlst.org/data/> have an MLST scheme

Sequence Type: ST38 (MLST Scheme: *Klebsiella pneumoniae*)²

Gene	%Identity	Alignment Length ³	DB Allele Length	Gaps	Best Match
<i>gapa</i>	100	450	450	0	<i>gapa_2</i>
<i>infB</i>	100	318	318	0	<i>infB_1</i>
<i>mdh</i>	100	477	477	0	<i>mdh_2</i>
<i>pgi</i>	100	432	432	0	<i>pgi_1</i>
<i>phoE</i>	100	420	420	0	<i>phoE_2</i>
<i>rpoB</i>	100	501	501	0	<i>rpoB_2</i>
<i>tonB</i>	100	414	414	0	<i>tonB_2</i>

²A "-like" is added to the Sequence Type unless all alleles in the input genome match perfectly to a database allele (%Identity = 100 and Alignment Length = DB Allele Length)

³Length (in bp) of the alignment between the MLST allele in the database (DB) and the corresponding sequence in the input genome

[Results Including Alignments](#)

Resistance Genes

Resistance genes are identified for all species

Resistance gene	DB Allele/Alignment		Contig ID	Position in Contig	Phenotype	Accession no.
	%Identity	Length				
<i>aph(3)-Ic</i>	100	816/816	NODE_315_length_95316_cov_9.030152	13070..13885	Aminoglycoside resistance	X62115
<i>aadB</i>	100	534/534	NODE_315_length_95316_cov_9.030152	23401..23934	Aminoglycoside resistance	JN119852
<i>aadA1</i>	100	789/789	NODE_315_length_95316_cov_9.030152	63994..64782	Aminoglycoside resistance	JQ480156
<i>aac(6')-Ib</i>	100	606/606	NODE_315_length_95316_cov_9.030152	64852..65457	Aminoglycoside resistance	M21682
<i>fosA</i>	97.38	420/420	NODE_259_length_1578568_cov_16.819324	409319..409738	Fosfomycin resistance	NZ_AFB001000747
<i>sul1</i>	100	927/927	NODE_315_length_95316_cov_9.030152	19809..20735	Sulphonamide resistance	CP002151
<i>sul2</i>	100	816/816	NODE_315_length_95316_cov_9.030152	4666..5481	Sulphonamide resistance	GQ421466
<i>oqxA</i>	99.4	1176/1176	NODE_61_length_3651056_cov_15.427067	2166270..2167445	Quinolone resistance	EU370913
<i>oqxB</i>	98.86	2450/2450	NODE_61_length_3651056_cov_15.427067	2167469..2169918	Quinolone resistance	EU370913
<i>aac(6')Ib-cr</i>	99.42	519/519	NODE_315_length_95316_cov_9.030152	64852..65370	Fluoroquinolone and aminoglycoside resistance	EF636461
<i>tet(D)</i>	100	1185/1185	NODE_315_length_95316_cov_9.030152	91379..92563	Tetracycline resistance	AF467077
<i>cmlA1</i>	99.13	1260/1260	NODE_315_length_95316_cov_9.030152	21181..22440	Phenicol resistance	AB212941
<i>blaSHV-11</i>	100	861/861	NODE_123_length_1427_cov_22.562719	594..1454	Beta-lactam resistance	EF035558
<i>blaTEM-1A</i>	99.88	861/861	NODE_315_length_95316_cov_9.030152	61896..62756	Beta-lactam resistance	HM749966
<i>blaOXA-9</i>	99.88	840/840	NODE_315_length_95316_cov_9.030152	63110..63949	Beta-lactam resistance	JF703130
<i>blaSHV-11</i>	100	861/861	NODE_61_length_3651056_cov_15.427067	693558..694418	Beta-lactam resistance	EF035558

[Results Including Alignments](#)

[Databases Allele Sequence\(s\)](#)

[Input Genome Allele Sequence\(s\)](#)

Plasmid Replicons

Plasmid replicons are identified for Enterobacteriaceae and gram positive bacteria

Plasmid replicon	%Identity	DB Allele/Alignment Length	Contig ID	Position in Contig	Note	Accession no.
<i>IncFII(pRSB107)</i>	97.7	261/261	Supercontig_1.8	38667..38927	pRSB107	AI851089
<i>IncFIB(AP001918)</i>	96.92	682/682	Supercontig_1.8	43597..44278		AP001918
<i>IncI1</i>	97.89	142/142	Supercontig_1.7	59412..59553	Alpha	AP005147
<i>IncQ1</i>	100	450/450	Supercontig_1.2	834376..834825		HE654726
<i>IncP</i>	99.44	534/535	Supercontig_1.2	848856..849390	alpha	L27758
<i>Col(MG828)</i>	91.98	262/262	Supercontig_1.9	937..1198		NC_008486

[Results Including Alignments](#)[Databases Allele Sequence\(s\)](#)[Input Genome Allele Sequence\(s\)](#)

Plasmid Multilocus Sequence Type (pMLST)

pMLST is performed for plasmids of the type *incF*, *IncH1*, *IncH2*, *IncI1*, or *IncN*

Sequence Type: [F48:A-B25] (pMLST Profile: IncF)

Gene	%Identity	Alignment Length ⁴	DB Allele Length	Gaps	Best match
<i>fia</i>	-	-	-	-	-
<i>fli</i>	100	157	157	0	fli_48
<i>fib</i>	100	373	373	0	fib_25

⁴Length (in bp) of the alignment between the pMLST allele in the database (D8) and the corresponding sequence in the input genome

[Results Including Alignments](#)

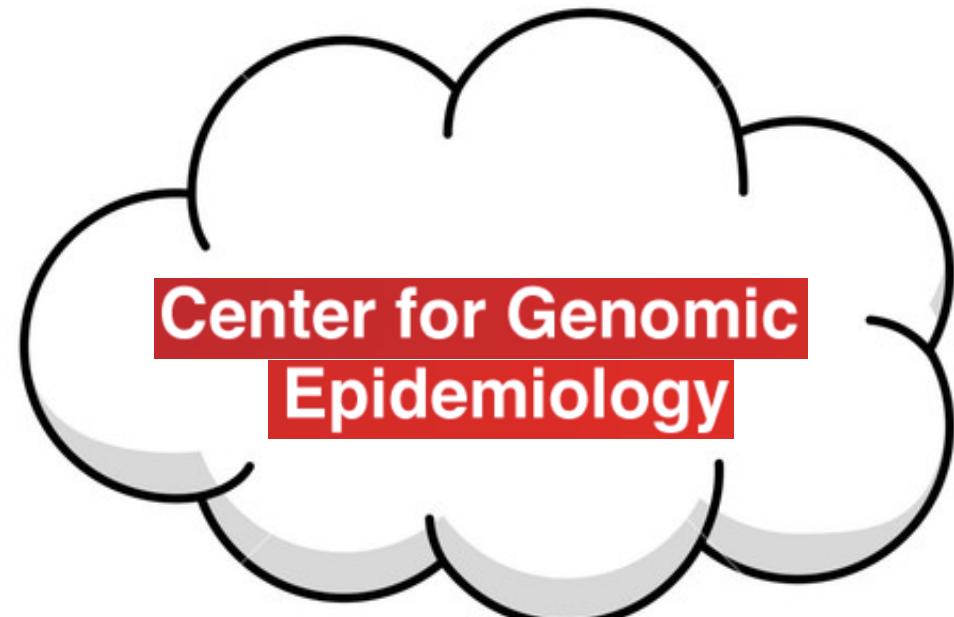
Virulence Genes

Virulence genes are identified for *Escherichia coli*, *Staphylococcus aureus*, and *Enterococcus*

Virulence factor	%Identity	DB Allele/Alignment Length	Contig ID	Position in Contig	Protein Function	Accession no.
<i>stx2A</i>	100	960/960	Supercontig_1.6	1455019..1455978		AY143336
<i>stx2B</i>	100	270/270	Supercontig_1.6	1455990..1456259		AE005174
<i>iha</i>	100	2091/2091	Supercontig_1.6	1540113..1542203		CP003289
<i>mchB</i>	96.39	294/305	Supercontig_1.6	1553612..1553916		AE014075
<i>mchC</i>	100	1551/1551	Supercontig_1.6	1554188..1555738		CP003289
<i>mchF</i>	100	2103/2103	Supercontig_1.6	1557636..1559738		CP003301
<i>sepA</i>	99.88	4095/4096	Supercontig_1.8	17173..21268		CP003299
<i>pic</i>	100	4119/4119	Supercontig_1.5	183..4301		CP003289
<i>capU</i>	100	1089/1089	Supercontig_1.6	188824..189912		CU928145
<i>gad</i>	99.93	1401/1401	Supercontig_1.6	2088066..2089466		CP003301
<i>sigA</i>	100	3858/3858	Supercontig_1.6	2540..6397		AE005674
<i>aap</i>	100	351/351	Supercontig_1.8	255..605		Z32523
<i>aaiC</i>	100	507/507	Supercontig_1.2	26777..27283		cp003301
<i>aar</i>	100	201/201	Supercontig_1.8	2864..3064		SSI_AA784
<i>aggD</i>	99.21	759/759	Supercontig_1.8	54326..55084		U12894
<i>aggC</i>	99.64	2529/2529	Supercontig_1.8	55098..57626		U12894
<i>aggB</i>	100	438/438	Supercontig_1.8	57640..58077		AFRH01000026
<i>aggA</i>	100	504/504	Supercontig_1.8	58179..58682		AFRH01000026
<i>ORF3</i>	100	1029/1029	Supercontig_1.8	63740..64768		HE610901
<i>ORF4</i>	100	540/540	Supercontig_1.8	64772..65311		AFRH01000026
<i>aggR</i>	100	798/798	Supercontig_1.8	69040..69837		afrh01000026
<i>aatA</i>	100	1239/1239	Supercontig_1.8	8486..9724		AFRH01000026
<i>lpfA</i>	100	573/573	Supercontig_1.2	951003..951575		CU928145

[Results Including Alignments](#)[Databases Allele Sequence\(s\)](#)[Input Genome Allele Sequence\(s\)](#)

GoSeqIt Tools



- The CGE methods in the cloud
- Currently only the Bacterial Analysis Pipeline is available (KmerFinder, ResFinder, MLST, PlasmidFinder, pMLST, VirulenceFinder)
- Currently only pre-assembled draft genomes in FASTA format are accepted
- In the coming months all CGE methods will be made available via GoSeqIt Tools and it will be able to accept raw reads
- Price: App. EURO 1 per method per isolate (currently we are in beta and it is free)

<http://tools.goseqit.com/#/login>



Email address

We'll never share your email with anyone else.

Password

Log in

Registration

[Forgot your password?](#)



Reports

Filters: All ▾

Report name	Last update	Project	File	Quality (N50)	
ERR070040-P12 09-03-2017 08:16:52 UTC	9/3/17 9:16:52	Final test ...	ERR070040...	156394	Reanalyse file Show report
Llactis6005contigs 07-03-2017 11:34:22 UTC	7/3/17 12:34:22	Final test ...	Llactis6005c...	45735	Reanalyse file Show report
Clostridium_beijerinckii_NCIMB_8052_uid12637 07-03-2017 11:34:20 UTC	7/3/17 12:34:20	Final test ...	Clostridium_...	6000632	Reanalyse file Show report

Run analysis

Drop files
(.fa .fna .fsa .fasta)Drop files for analysis or [browse your computer](#)

Activity

In progress

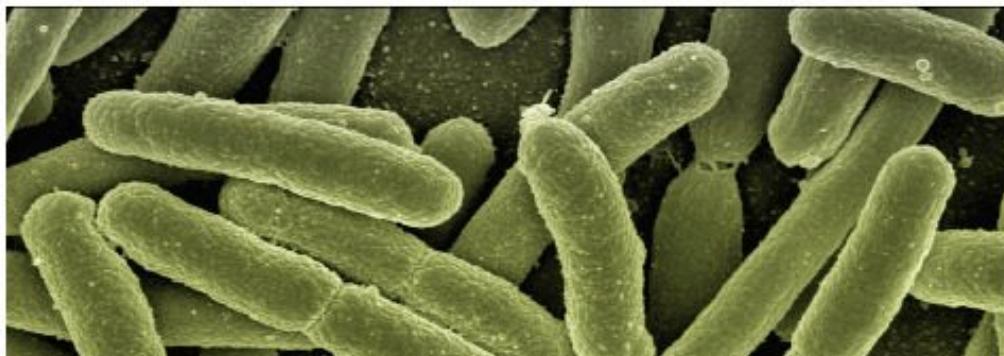
 ERR070040-P12 1...
Final test before release t...

Analysis

Results

- When the analysis is complete, you will receive a report in pdf format via mail
- You can also browse the results via GoSeqIt Tools

Detailed Results				MLST	ResFinder	VirulenceFinder	PlasmidFinder	pMLST
Identification of Acquired Antimicrobial Resistance Genes (ResFinder)								
Resistance gene	% Identity	Query/HSP length	Contig	Position in contig	Phenotype	Accession no.		
aac(6')-aph(2")	100.00	1440/1440	NODE_68_length_2008_cov_16.922810	490..1929	Aminoglycoside resistance	M13771		
aadD	99.87	771/771	NODE_27_length_4644_cov_22.306202	3743..4513	Aminoglycoside resistance	AF181950		
blaZ	95.86	846/846	NODE_77_length_17228_cov_18.322962	16288..17133	Beta-lactam resistance	AP003139		
erm(C)	100.00	735/735	NODE_2_length_2393_cov_381.271210	397..1131	Macrolide resistance	M13761		
mecA	100.00	2010/2010	NODE_37_length_4979_cov_22.601526	1453..3462	Beta-lactam resistance	AB505628		
norA	92.37	1167/1167	NODE_74_length_406518_cov_18.662951	218491..219657	Fluoroquinolone resistance	M97169		
Results including alignments (txt)			Results only (tab)	Input genome gene sequences (fsa)			Database gene sequences (fsa)	



Exercise5

The Bacterial Analysis Pipeline and batch upload

The purpose of this exercise is to use batch upload and the Bacterial Analysis Pipeline (BAP) to investigate the content of three samples of unknown origin.

Data

Choose *one* of the following three files to work with:

Day_3/Pipeline/unknown_1.fasta