



Computational Practical 7 & 8

Online tools

Agilakumari Pragasam
OUCRU, Ha Noi, Viet Nam



Database on antimicrobial resistance mechanisms

- AMR finder plus
- Antimicrobial Drug Databases (AMDD)
- ARDB (no longer maintained)
- ARGminer
- BacMet
- Beta-Lactamases Databases (BLAD)
- CBMAR
- Enterobase
- The Comprehensive Antibiotic Resistance Database
- FARME
- INTEGRALL
- LcED
- MEGARes
- MULBII-TB-DB
- Mustard Database
- MvirDB
- PathoPhenoDB
- PATRIC dataase
- RAC: Repository of Antibiotic resistance Cassettes
- ResFinder
- TBDReaMDB
- U-CARE
- VFDB

The 'Big Three'

- ✓ NCBI's Pathogen Detection Reference Gene catalog
- ✓ **Comprehensive Antibiotic Resistance Database**
- ✓ **Resfinder**



RGI Resistance Gene Identifier

RGI can be used to predict resistomes from protein or nucleotide data based on homology and SNP models. Analyses can be performed via this web portal (20 Mb limit), via the command line, or via use of a [Galaxy wrapper](#). The command line version can be obtained from the [Download section of the CARD website](#). You can additionally install RGI from Conda or run RGI from Docker.

This web portal supports analysis of genomes, genome assemblies, metagenomic contigs, or proteomes. The command line tool additionally supports analysis of metagenomic reads and k-mer prediction of pathogen-of-origin for AMR genes.

Web portal - RGI 6.0.3, CARD 3.3.0: Open Reading Frame (ORF) prediction using [Prodigal](#), homolog detection using [DIAMOND](#), and Strict significance based on CARD curated bitscore cut-offs. Options included for percent identity filtering, low quality/coverage assemblies, merged metagenomic reads, small plasmids or assembly contigs (<20,000 bp).

Online RGI results cached for 7 days. As the CARD curation evolves, the results of the RGI evolve. RGI targets, reference sequences, and significance cut-offs are under constant curation. Full documentation for the RGI can be found at [GitHub](#).

Curation paradigm

- ✓ Peer reviewed publication
- ✓ Clear experimental evidence of elevated MIC (some exceptions)
- ✓ Public sequence record in GenBank

Curation – 1-3 months

Upload genome

Use RGI:

Enter a GenBank accession(s):

Enter accessions separated by commas

Nucleotide sequences will undergo ORF calling to generate predicted protein sequences. Examples: JN420336.1, AY123251.1, HQ451074.1, AL123456

Upload FASTA sequence file(s):

Choose Files A1-11_S11_L001.fasta

Upload a **plain text file** containing DNA or protein sequence(s) in FASTA format (20 Mb limit). The file can contain more than one FASTA formatted sequence, such as assembly contigs or multiple proteins. Each file will be treated as a single sample.

CARD:Live

☐ Consent, allowing CARD to collect pathogen, AMR gene list, and geotemporal data for these isolates.

If consenting, please select geographical source of isolates

The CARD:Live project collects pathogen identification, MLST, AMR gene list, date, and geographical region for genome sequences submitted to RGI online, providing a dynamic view of the antibiotic resistant isolates being analyzed around the world. **No sequence information is collected.** Full information at [CARD:Live](#).

Submit

Select Data Type:

- ☒ DNA sequence
☐ Protein sequence

Select Criteria:

- ☒ Perfect and Strict hits only
☐ Perfect, Strict and Loose hits

Nudge $\geq 95\%$ identity Loose hits to Strict:

- ☒ Exclude nudge
☐ Include nudge

Sequence Quality:

- ☒ High quality/coverage¹
☐ Low quality/coverage²

¹ Complete genomes, plasmids, or high quality assemblies (includes contigs > 20,000 bp). Excludes prediction of partial genes.

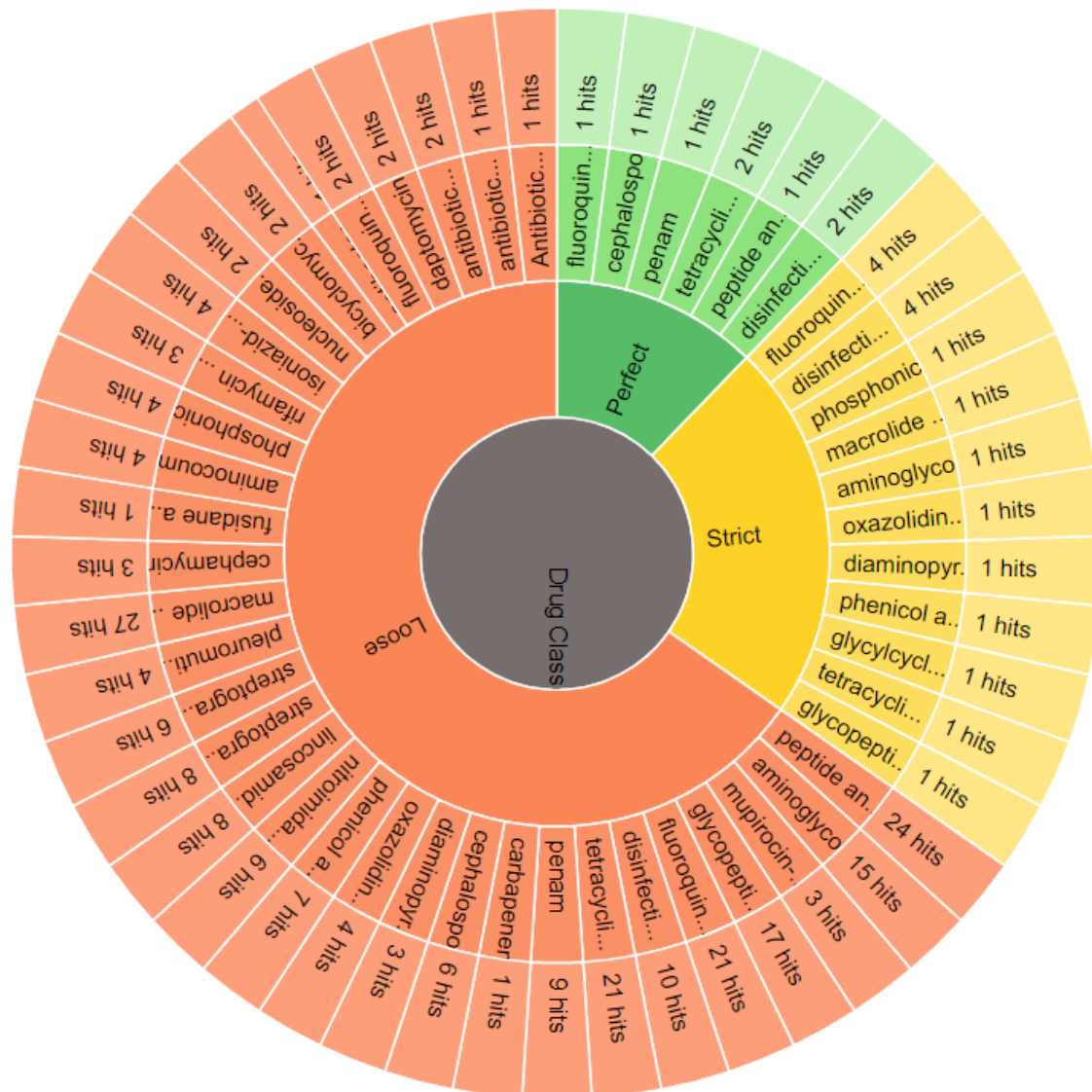
² Low quality/coverage assemblies, metagenomic merged reads, small plasmids or assembly contigs (<20,000 bp). Includes prediction of partial genes.

Upload external RGI json results and visualize:

Upload a **JSON file** containing RGI results generated using the command-line version. File size limited to 20 Mb. Note that only **Loose** hits of **e-10** or better can be visualized.

 Upload JSON

CARD results: AMR – Gene family



- Perfect – Perfectly matches with references sequences

“Known Known”



- Strict – Similarity within the references sequences, functional variant

“Known Unknown”



- Loose – Similarity outside of model; needs validation

“Unknown Unknown”

Exercise

Klebsiella pneumoniae

- A6-6_S6_L001.fasta
- A6-3_S3_L001.fasta


Staphylococcus aureus

- A1-11_S11_L001.fasta
- A1-14_S14_L001.fasta



Resfinder

- ✓ ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance in total or partial DNA sequence of bacteria.

 **Center for Genomic Epidemiology**

Services

Contact

ResFinder

Version

4.6.0 ▼

ResFinder identifies acquired genes and/or finds chromosomal mutations mediating antimicrobial resistance in total or partial DNA sequence of bacteria.

ResFinder software: (2024-03-22)
ResFinder database: (2024-03-22)
PointFinder database: (2024-03-08)
DisinFinder database: (2023-05-31)

<http://genepi.food.dtu.dk/resfinder>

WCS ACORN - Integrated AMR: From Genomic
Analysis to Clinical Application



Upload genome

Other
Campylobacter jejuni*
Campylobacter coli*
Campylobacter spp.*
Enterococcus faecalis*
Enterococcus faecium*
Escherichia coli*
Helicobacter pylori*
Klebsiella*
Mycobacterium tuberculosis*
Neisseria gonorrhoeae*
Plasmodium falciparum*
Salmonella spp.*
Staphylococcus aureus*

Chromosomal point mutations:

Threshold for %ID

90% ▼

Minimum length

60% ▼

- ☐ Show unknown mutations
- ☐ Ignore premature stop codons:
- ☐ Ignore frameshift indels:

Acquired antimicrobial resistance genes:

Threshold for %ID

90% ▼

Minimum length

60% ▼

Species and input data type:

Select species

Other ▼

Select input type

FASTA (Assembled Genome/Contigs) ▼

FASTA (Assembled Genome/Contigs)

FASTQ (Non-nanopore Reads)

FASTQ (Nanopore Reads)

Disinfectant:

- ☐ Run disinfectant

Threshold for %ID

90% ▼

Minimum length

60% ▼

Upload and submit job:

Email (Get email, when finished - Optional):

Enter your email address...

Files (The sum of uploaded file sizes cannot exceed 1 gb):

Choose File No file chosen

Choose File No file chosen

Exercise

Analyze 2 organisms each from K. pneumoniae and S. aureus to the Resfinder

Klebsiella pneumoniae

- A6-6_S6_L001.fasta
- A6-3_S3_L001.fasta

Staphylococcus aureus

- A1-11_S11_L001.fasta
- A1-14_S14_L001.fasta

Use default parameters for all

✓ A1-14_S14_L001.fasta – Choose Unknown mutations



A Global Platform for Genomic Surveillance

- ✓ MLST prediction is available for over 100 species using schemes from PubMLST, Pasteur, and Enterobase
- ✓ cgMLST calling and clustering is available for the following schemes:

[Acinetobacter baumannii](#)

[Candida auris](#)

[Klebsiella pneumoniae](#)

[Listeria](#)

[Mycobacterium canettii](#)

[Neisseria meningitidis](#)

[Staphylococcus aureus](#)

[Campylobacter coli](#)

[Enterococcus faecium](#)

[Klebsiella quasipneumoniae](#)

[Mycobacterium africanum](#)

[Mycobacterium tuberculosis](#)

[Salmonella enterica](#)

[Vibrio cholerae](#)

[Campylobacter jejuni](#)

[Escherichia](#)

[Klebsiella variicola](#)

[Mycobacterium bovis](#)

[Neisseria gonorrhoeae](#)

[Shigella](#)

Further Analyses

Organism	AMR Prediction	Core SNP Trees	Other
<i>Campylobacter</i>	✓		
<i>Candida auris</i>	✓	✓	
<i>Klebsiella</i>	✓		Kleborate
<i>Klebsiella pneumoniae</i>	✓	✓	Kleborate LIN codes
<i>Neisseria gonorrhoeae</i>	✓	✓	NG-MAST NG-STAR
<i>Renibacterium salmoninarum</i>		✓	
<i>Salmonella</i>			SISTR
<i>Salmonella Typhi</i>	✓	✓	Genotypi
SARS-CoV-2		✓	Pangolin
<i>Staphylococcus aureus</i>	✓	✓	
<i>Streptococcus equi</i>		✓	
<i>Streptococcus pneumoniae</i>	✓		PopPUNK 2 SeroBA PBP Typing
<i>Vibrio cholerae</i>	✓	✓	Vista Genotypes PopPUNK 2
Zika virus		✓	

Upload genome

NEW UPLOAD

PREVIOUS UPLOADS

What would you like to upload?



Single Genome FASTAs

One or more FASTA files, one genome per FASTA file.
(e.g. bacterial genomes)

Upload FASTA(s)



Multi-genome FASTAs

Multiple genomes per file, one genome per record.
(e.g. viral genomes)

Upload FASTA(s)



FASTQ

One or more pairs of read files in FASTQ format.

Upload FASTQ(s).

Exercise

Klebsiella pneumoniae

- A6-6_S6_L001.fasta
- A6-3_S3_L001.fasta

Staphylococcus aureus

- A1-11_S11_L001.fasta
- A1-14_S14_L001.fasta



A Global Platform for Genomic Surveillance



Proksee - Genome Analysis

- ✓ Genome assembly, annotation and visualization, featuring interactive circular and linear genome maps

Input :

Assemblies (fasta)

Genbank


EMBL

Assembly for Reads

<https://www.ncbi.nlm.nih.gov/nuccore/57158257,57158258>



Start a new Project

Proksee [New Project](#) [My Projects](#) [About](#) [Help](#)  [Tutorial](#)

0%

Welcome to Proksee

Proksee is an expert system for genome assembly, annotation and visualization. To begin using Proksee, provide a complete genome sequence, sequencing reads or a CGView/Proksee map JSON file.

Create a New Map

Genome

Reads Beta

JSON

Proksee Status

C Running...

You will be redirected to your results after the map is generated. You can also bookmark this page and check back later.

Run an Example

NC_000907

Haemophilus influenzae Rd KW20
1,830,138 bp

NC_000913

Escherichia coli str. K-12 substr. MG1655
4,641,652 bp

CP002121

Escherichia coli B171 plasmid pB171_90
90,229 bp

CP002122

Escherichia coli DA2

Project Page

Proksee [New Project](#) [My Projects](#) [About](#) [Help](#)

[Tutorial](#)

3%

 [Saved](#)[About](#) [Map](#)

ID: d36b67a8-d391-4bcb-9c4a-4e786c19d133

1.00 x

1 bp

Mark

WCS AGORN - Integrated AMR: From Genomic Analysis to Clinical Application

GCF_019335785.1.fa

[Display](#) [Regions](#) [Download](#) [Tools](#) [Jobs](#)

Genome Annotation5

ORFsAdd

ProkkaStart

CARD Resistance Gene IdentifierStart

pLannotateStart

CRISPR/Cas FinderStart

User Data2

PlotsAdd

FeaturesAdd

Sequence Comparison2

FastANI NewStart

BLASTStart

Mobile Genetic Elements3

Alien Hunter NewStart

mobileOG-db NewStart

VirSorter NewStart

Sequence Composition2

Job Tab

Proksee

New ProjectMy ProjectsAboutHelp

Tutorial3%

GCF_019335785.1.fa

Saved

AboutMapProkka X

Job: Prokka

running

Name

Prokka

Size

0 B

ID

79705

Created

Today, 12:45:13

Run Time

NA

Feature Source ID

prokka-1.*

Start Time

Today, 12:45:13

End Time

-

Inputs

genus: Other, contigs: input/contigs, jobName: Prokka, kingdom: Bacteria, genetic_code: Bacterial and Plant Plastid

Log

[19:45:13] Date: 2022-11-29

[19:45:13] Project: d36b67a8-d391-4bcb-9c4a-4e786c19d133

[19:45:13] Tool: Prokka [1.1.0]

[19:45:13] Software: Prokka 1.14.6

[19:45:13] Working directory created

[19:45:13] Writing CGView data to files:

[19:45:13] - contigs: PROJECT_DIR/prokka/79705/input/contigs

[19:45:14] Data Synced to worker: 'ps-worker-2'

[19:45:14] Starting Job - Prokka...

[19:45:14] Inputs: {"genus"=>"other", "contigs"=>"input/contigs", "jobName"=>"Prokka", "kingdom"=>"Bacteria", "genetic_code"=>"Bacterial and Plant Plastid"}

[19:45:14] Command: cd PROJECT_DIR/prokka/79705 || exit 1

[19:45:14] prokka prokka --working-dir PROJECT_DIR/prokka/79705

DisplayRegionsDownloadToolsJobs1

Search...

Genome Annotation5

ORFsAdd

ProkkaStart

CARD Resistance Gene IdentifierStart

pLannotateStart

CRISPR/Cas FinderStart

User Data2

PlotsAdd

FeaturesAdd

Sequence Comparison2

FastANI>NewStart

BLASTStart

Mobile Genetic Elements3

Alien Hunter>NewStart

mobileOG-db>NewStart

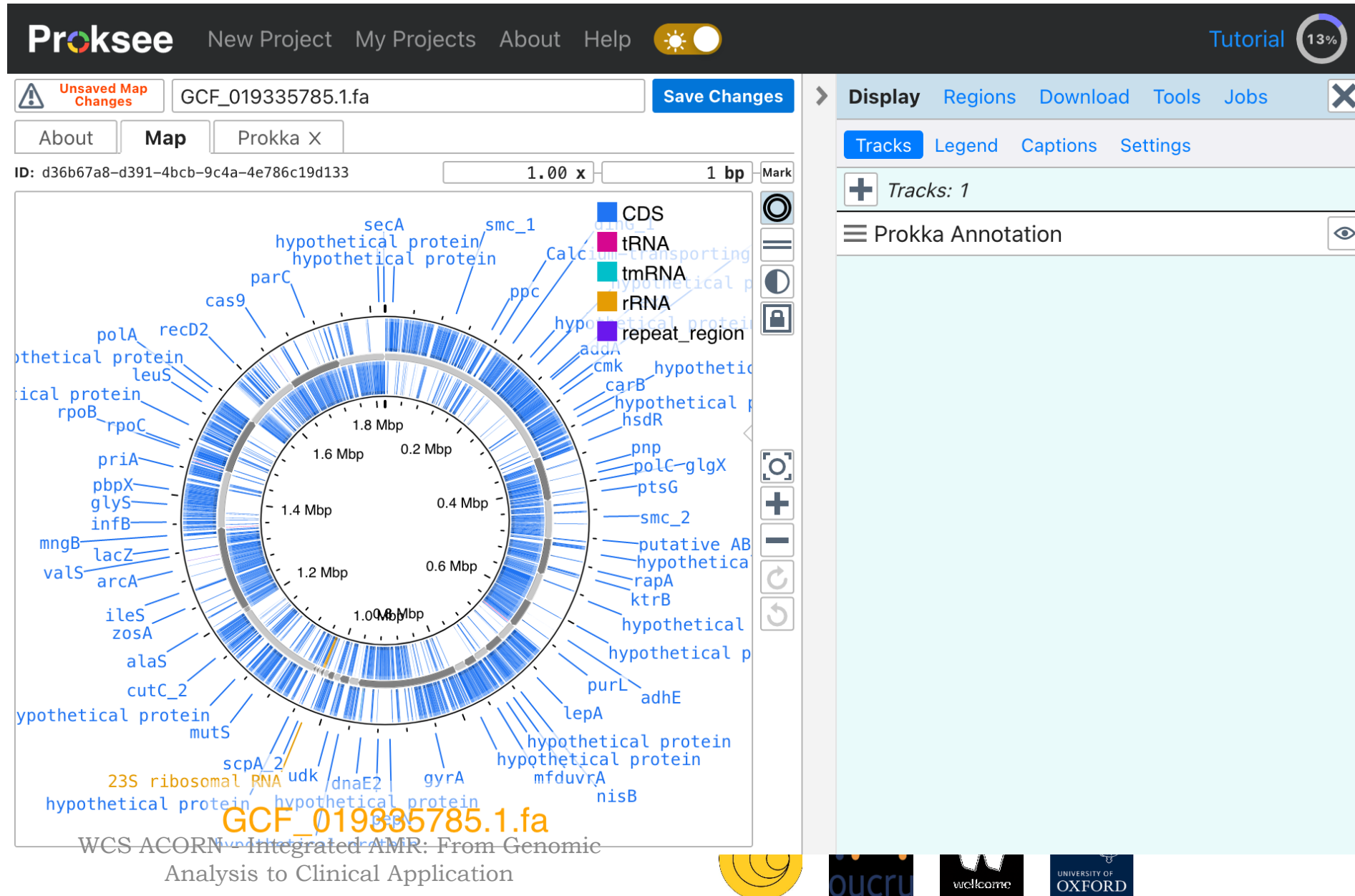
VirSorter>NewStart

Sequence Composition2



WGS ACORN - Integrated AMP-Res. Genomics

Analysis to Clinical Application

Job Report



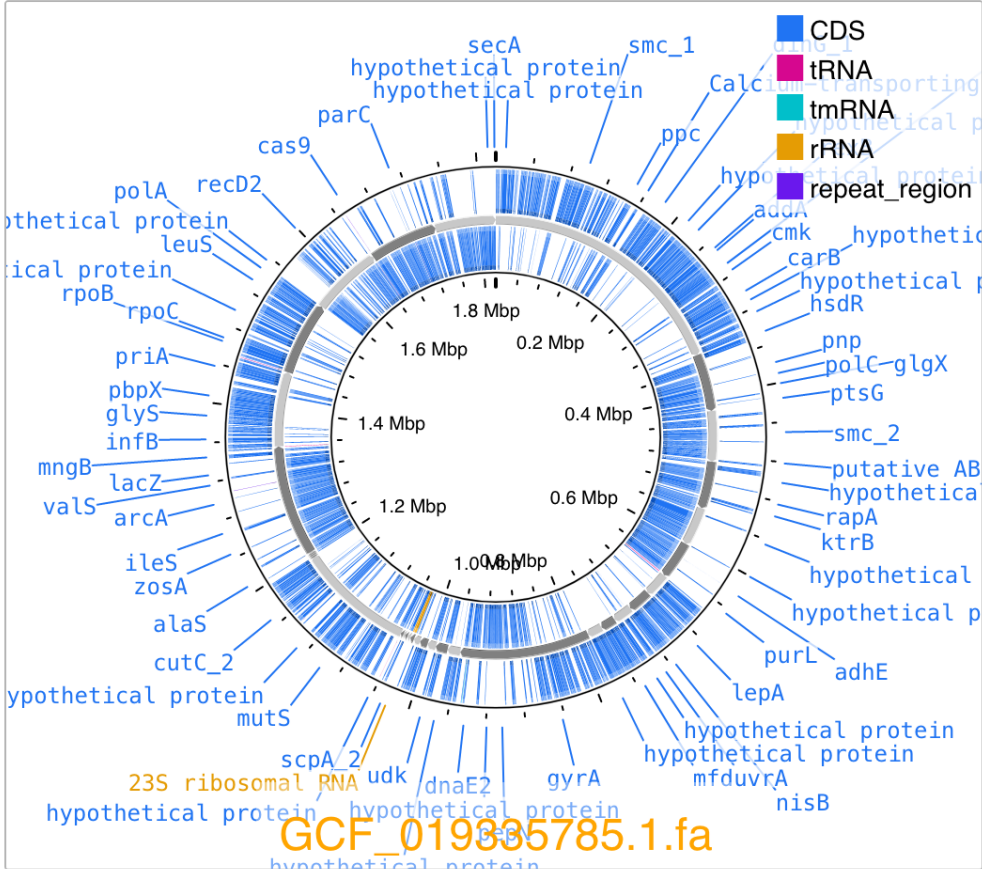
Downloading Results

Proksee New Project My Projects About Help  Tutorial 

⚠ Unsaved Map Changes GCF_019335785.1.fa Save Changes

About Map Prokka X

ID: d36b67a8-d391-4bcb-9c4a-4e786c19d133 1.00 x 1 bp Mark



Download the map as a png image. The maximum size for the width/height is 8000px.

	Download Size	Current Size
Width:	521 px	521 px
Height:	463 px	463 px
Percent:	100 %	

Download

Image - SVG **New**

JSON (Reloadable Map Data)

Download the CGView JSON data. This file can be loaded later to view and edit the map.

Download

Sequences

Exercise

Go to the NCBI database and visualize the Gene bank **accession ID: AP006726.1**

- Download the fasta and genbank file
- Annotate the plasmid and run the tools below:
 - CARD
 - Mobile-OG db
 - GC content
 - GC skew

<https://www.ncbi.nlm.nih.gov/nucore/57158257,57158258>

Phenotyping

[ResFinder](#)

Identification of acquired antibiotic resistance genes.

[ResFinderFG](#)

Identification of functional metagenomic antibiotic resistance determinants.

[LRE-finder](#)

Identification of genes and mutations leading to linezolid resistance.

[KmerResistance](#)

Identification of acquired antibiotic resistance genes using Kmers.

[PathogenFinder](#)

Prediction of a bacteria's pathogenicity towards human hosts.

[VirulenceFinder](#)

Identification of acquired virulence genes.

[Restriction-ModificationFinder](#)

Determination of Restriction-Modification sites (based on [REBASE](#).)

[SPIFinder](#)

SPIFinder identifies Salmonella Pathogenicity Islands.

[ToxFinder](#)

ToxFinder identifies genes involved in mycotoxin synthesis.

Phylogeny

[MINTyper](#)

Identification of SNPs with automatic filtering, masking and site validation together with inferred phylogeny based on both long and short sequencing data.

[CSIPhylogeny](#)

CSI Phylogeny calls SNPs, filters the SNPs, does site validation and infers a phylogeny based on the concatenated alignment of the high quality* SNPs.

[NDtree](#)

NDtree constructs phylogenetic trees from Single-End or Pair-End FASTQ files.

[Evergreen](#)

Evergreen generates a forest of constantly updated phylogenetic trees with publicly available whole-genome sequencing data from foodborne, bacterial isolates that were deposited in the short sequencing read archives (NCBI SRA/ENA).

[TreeViewer](#)

Phylogeny Tree Viewer.

Plasmid finder

Center for Genomic Epidemiology

[Home](#)[Services](#)[Publications](#)[Contact](#)

PlasmidFinder 2.1

[Service](#)[Instructions](#)[Output](#)[Article abstract](#)[Citations](#)

Software version: 2.0.1 ([2020-07-01](#))

Database version: ([2023-01-18](#))

[Test sequence](#)

The database is curated by:

Henrik Hasman and Alessandra Carattoli

([click to contact](#))

Select database

Gram Positive

Enterobacteriales



Exercise

Klebsiella pneumoniae

- A6-6_S6_L001.fasta

Staphylococcus aureus

- A1-14_S14_L001.fasta

Don't know much about your bacterial species ??

Don't worry, Let DIVE in
together

BacDive

Total strains: **97,334**
Type strains: **20,060**

<https://bacdive.dsmz.de/>



GLOBAL
CORE
BIODATA
RESOURCE



BacDive

Explore **Bacterial Diversity**



BacDive is the worldwide largest database for standardized bacterial information. Its mission is to mobilize and integrate research data on strain level from diverse sources and make it freely accessible. [Read more.](#)



BacDive

The information you
need to become a
microbiologist !!

Include predictions ☐ [Copy query link](#)

Query

Species name (e.g. *E. coli*) contains

AND

Antibiotic sensitivity

+ Add AND

+ Add OR

Submit

Klebsiella pneumoniae

designation_header: 240

strain number: DSM 2026, ATCC 15380, NCIB 418, NCTC 418, CIP 103733, NCDO 711, NCTC 240

synonym

Hyalococcus pneumoniae

Antibiotic sensitivity

0

▼	Name and taxonomic classification
▼	Morphology
▼	Culture and growth conditions
▼	Physiology and metabolism
▼	Isolation, sampling and environmental information
▼	Safety information
▼	Sequence information
▼	Genome-based predictions

You can look for your reference genome here

You need more ???

Let's see the virulence factors

<http://www.mgc.ac.cn/VFs/main.htm>



Let's investigate more about Kleb

Comparative pathogenomics of *Klebsiella* (Total 12 genomes available [Customize comparison](#))

Virulence factors	Related genes	<i>K. oxytoca</i> E718	<i>K. oxytoca</i> KCTC 1686	<i>K. pneumoniae</i> 342	<i>K. pneumoniae</i> CG43	<i>K. pneumoniae</i> JM45	<i>K. pneumoniae</i> KCTC 2242		<i>K. pneumoniae</i> subsp. <i>pneumoniae</i> 1084	<i>K. pneumoniae</i> subsp. <i>pneumoniae</i> HS11286	<i>K. pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578
		chromosome NC_018106	chromosome NC_016612	chromosome NC_011283	pLVPK NC_005249	chromosome NC_022082	chromosome NC_017540	pKCTC2242 NC_017541	chromosome NC_018522	chromosome NC_016845	chromosome NC_009648

WCS ACORN - Integrated AMR: From Genomic
Analysis to Clinical Application

