RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

anses

EURL CPS
European Union Reference Laboratory for
Coagulase Positive Staphylococci
http://eurl-staphylococci.anses.fr

BLAST-based tools for the characterization of assembled genomes

EURL CPS
Marina Cavaiuolo

Joint Training Course of the inter EURLs Working Group on NGS

Characterisation of bacterial genomes

Search for genetic features
e.g. Antimicrobial resistance genes
Virulence gene

Typing
e.g. MLST
Serotyping, virulotyping

Principle of the analyses:

Alignment of sequences:
e.g. whole genomes, genes

Mapping of NGS reads

specific algorithms →

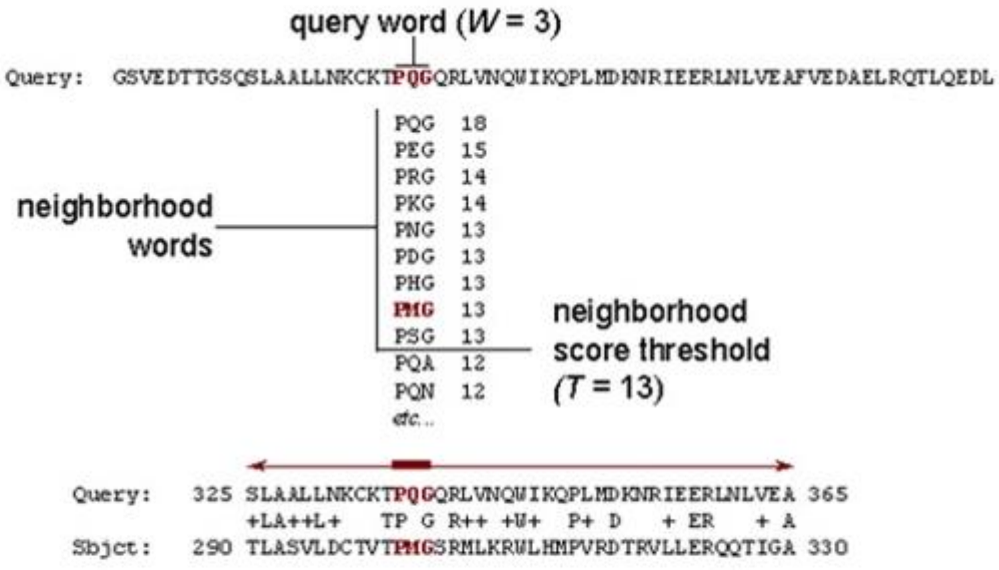
database of reference gene sequences

BLAST algorithm

Basic Local Alignment Search Tool (BLAST)

BLAST is a sequence comparison algorithm to find homologous genes and proteins

BLAST takes a **query** (single or multiple sequences) in fasta format and **compares** it to a **subject** (single or multiple sequences) by constructing local alignments



```
Query:  GSVEDTTGSQSLAALLNKCKTPQGQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDL
```

```
Query:   325 SLAALLNKCKTPQGQRLVNQWIKQPLMDKNRIEERLNLVEA 365
             +LA++L+   TP G R++ +W+  P+ D   + ER   + A
Sbjct:   290 TLASVLDCTVTPMGSRMLKRWLHMPVRDTRVLLERQQTIGA 330
```

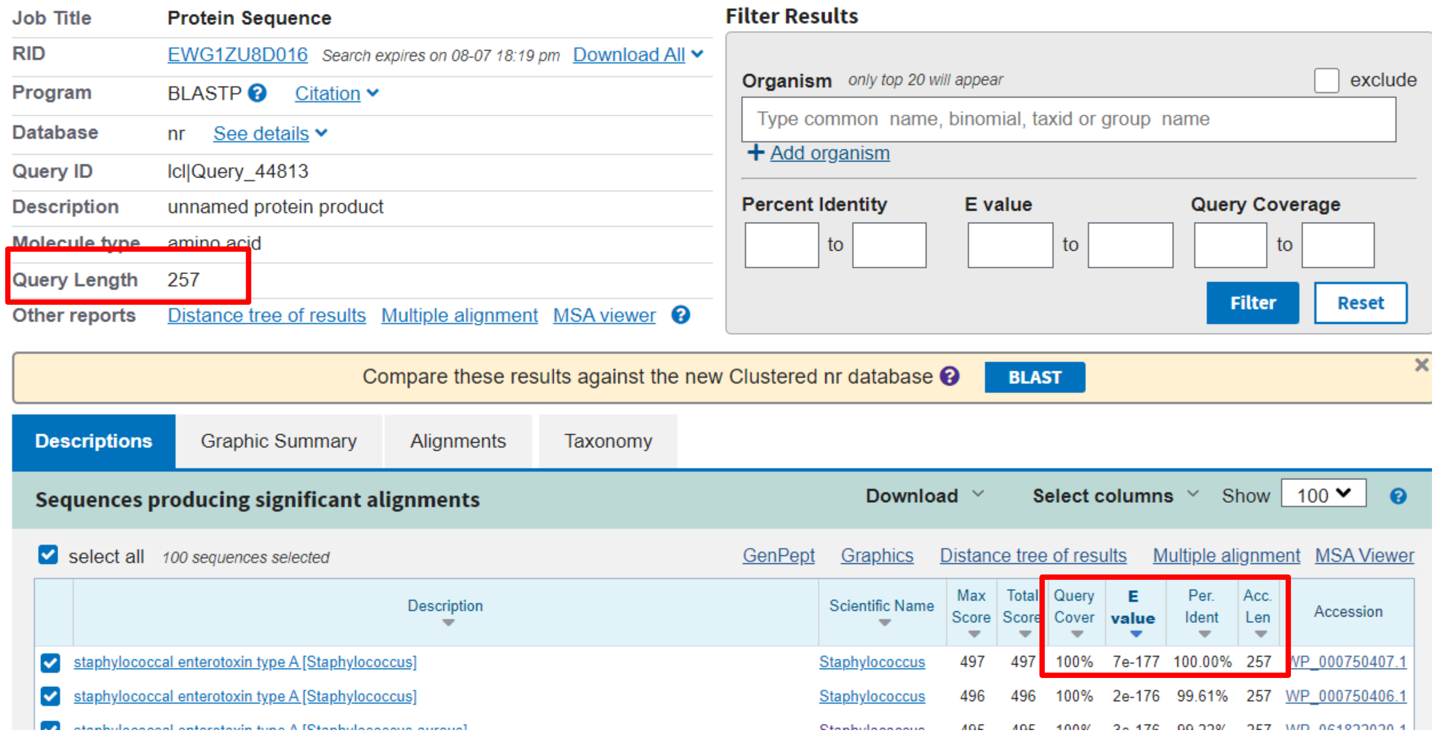
BLAST programs

Program	Database (Subject)	Query
BLASTN	Nucleotide	Nucleotide
BLASTP	Protein	Protein
BLASTX	Protein	Nt. → Protein
TBLASTN	Nt. → Protein	Protein
TBLASTX	Nt. → Protein	Nt. → Protein

How to interpret the results of a BLAST search?

Blast output: Alignment. The result of matching up nucleotide/amino acids of two or more biological sequences to achieve maximal levels of identity and conservation (for proteins), for the purpose of assessing the degree of similarity and the possibility of homology.

For NCBI's web-page, the default format for output is HTML



Job Title	Protein Sequence
RID	EWG1ZU8D016 *Search expires on 08-07 18:19 pm* Download All
Program	BLASTP Citation
Database	nr See details
Query ID	lcl|Query_44813
Description	unnamed protein product
Molecule type	amino acid
Query Length	257
Other reports	Distance tree of results Multiple alignment MSA viewer

Filter Results

Organism *only top 20 will appear* □ exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity [ ] to [ ] E value [ ] to [ ] Query Coverage [ ] to [ ]

Filter Reset

Compare these results against the new Clustered nr database BLAST

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

☑ select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	staphylococcal enterotoxin type A [Staphylococcus]	Staphylococcus	497	497	100%	7e-177	100.00%	257	WP_000750407.1
☑	staphylococcal enterotoxin type A [Staphylococcus]	Staphylococcus	496	496	100%	2e-176	99.61%	257	WP_000750406.1
☑	staphylococcal enterotoxin type A [Staphylococcus aureus]	Staphylococcus	495	495	100%	3e-176	99.22%	257	WP_061822020.1

1) How good is the match?

e-value: the number of expected hits of similar quality (score) that could be found just by chance

E = 10^{-4} is considered the cutoff point

E = 0 means that the sequences are identical

2) How long is the alignement ?

Coverage: the % of the query length that aligns with the subject.

3) How similar are the aligned segments?

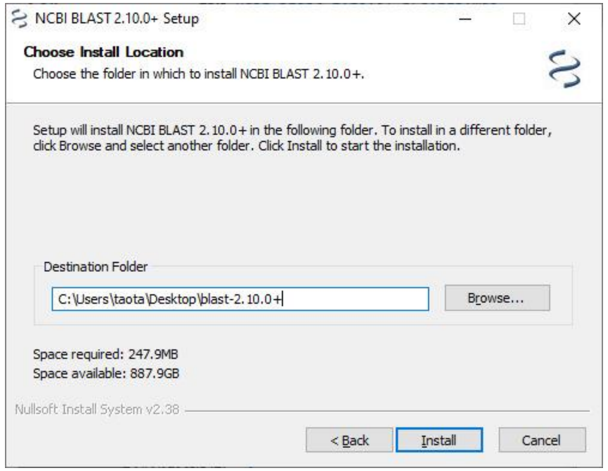
Identity: The extent to which two sequences have the same residues at the same positions in an alignment.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

anses

How to use BLAST?

1) BLAST+ standalone suite



Choose Install Location

Choose the folder in which to install NCBI BLAST 2.10.0+.

Setup will install NCBI BLAST 2.10.0+ in the following folder. To install in a different folder, click Browse and select another folder. Click Install to start the installation.

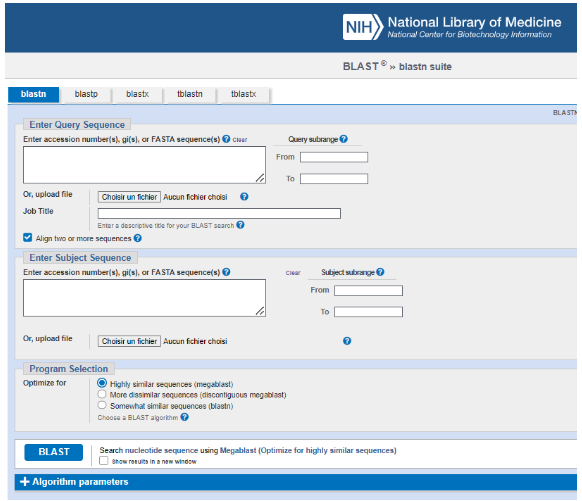
Destination Folder: C:\Users\taota\Desktop\blast-2.10.0+

Space required: 247.9MB
Space available: 887.9GB

Nullsoft Install System v2.38

https://www.ncbi.nlm.nih.gov/books/NBK52637/

2) BLAST online



Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Query subrange: From / To

Or, upload file: Aucun fichier choisi

Job Title: Enter a descriptive title for your BLAST search

☑ Align two or more sequences

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Subject subrange: From / To

Or, upload file: Aucun fichier choisi

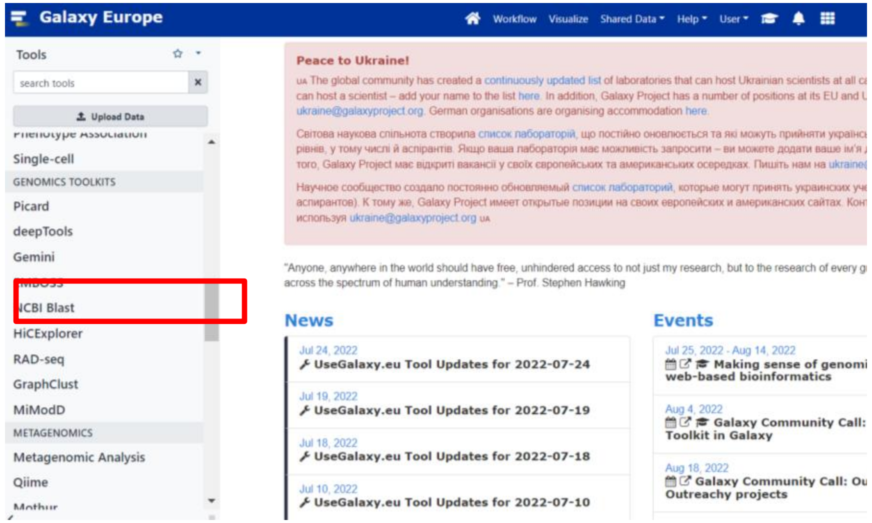
Program Selection

Optimize for
● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)
Choose a BLAST algorithm

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)
□ Show results in a new window

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

3) NCBI BLAST on Galaxy



Peace to Ukraine!

UA The global community has created a continuously updated list of laboratories that can host Ukrainian scientists at all ca can host a scientist – add your name to the list here. In addition, Galaxy Project has a number of positions at its EU and U ukraine@galaxyproject.org. German organisations are organising accommodation here.

Світова наукова спільнота створила список лабораторій, що постійно оновлюється та які можуть прийняти українсь рівнів, у тому числі й аспірантів. Якщо ваша лабораторія має можливість запросити – ви можете додати ваше ім'я д того, Galaxy Project має відкриті вакансії у своїх європейських та американських осередках. Пишіть нам на ukraine@

Научное сообщество создало постоянно обновляемый список лабораторий, которые могут принять украинских уче аспирантов). К тому же, Galaxy Project имеет открытые позиции на своих европейских и американских сайтах. Конт используя ukraine@galaxyproject.org UA

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every g across the spectrum of human understanding." – Prof. Stephen Hawking

News

Jul 24, 2022
UseGalaxy.eu Tool Updates for 2022-07-24

Jul 19, 2022
UseGalaxy.eu Tool Updates for 2022-07-19

Jul 18, 2022
UseGalaxy.eu Tool Updates for 2022-07-18

Jul 10, 2022
UseGalaxy.eu Tool Updates for 2022-07-10

Events

Jul 25, 2022 - Aug 14, 2022
Making sense of genomi web-based bioinformatics

Aug 4, 2022
Galaxy Community Call: Toolkit in Galaxy

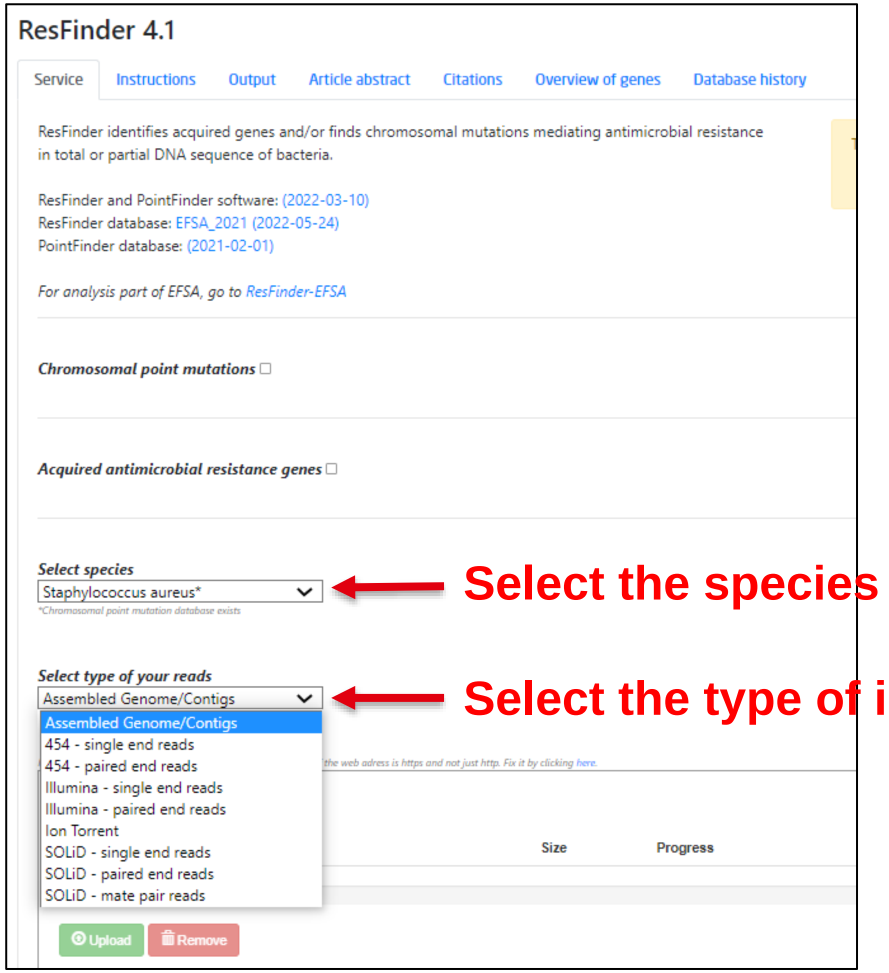
Aug 18, 2022
Galaxy Community Call: Ou Outreachy projects

4) Many web-tools and software implement BLAST as search engine:

1. Upload of the assembled genomes
2. Blast analysis against a database

Search for antimicrobial resistance genes using BLAST web-tools

Center for Genomic Epidemiology - CGE



ResFinder 4.1

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

ResFinder and PointFinder software: (2022-03-10)
ResFinder database: EFSA_2021 (2022-05-24)
PointFinder database: (2021-02-01)

For analysis part of EFSA, go to ResFinder-EFSA

Chromosomal point mutations □

Acquired antimicrobial resistance genes □

Select species
Staphylococcus aureus*
**Chromosomal point mutation database exists*

Select type of your reads
Assembled Genome/Contigs

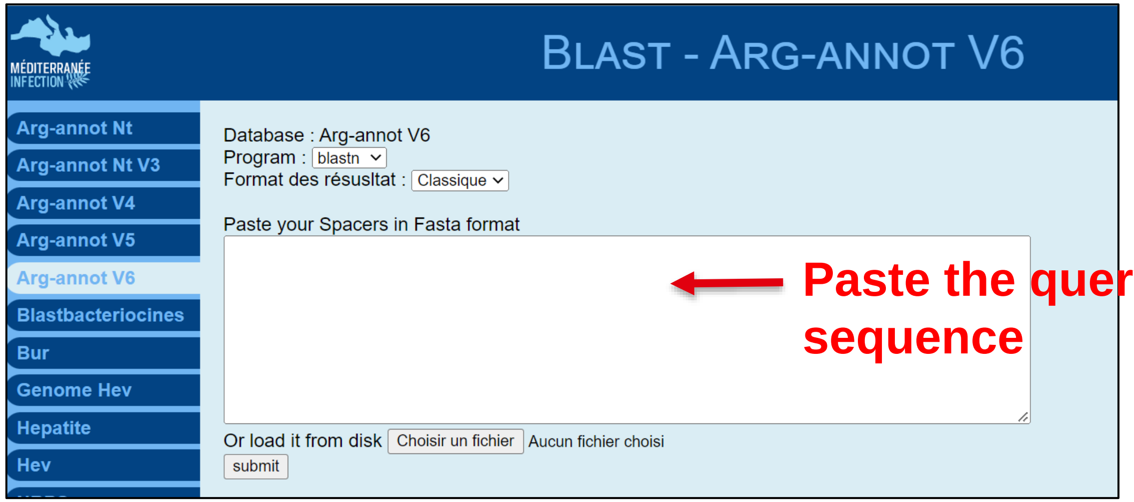
the web adress is https and not just http. Fix it by clicking here.

Select the species

Select the type of input

https://cge.food.dtu.dk/services/ResFinder/

Institut Hospitalier Universitaire Méditerranée Infection
IHU – Méditerranée Infection



BLAST - ARG-ANNOT V6

Database : Arg-annot V6
Program : blastn
Format des résusltat : Classique

Paste your Spacers in Fasta format

Or load it from disk Choisir un fichier Aucun fichier choisi

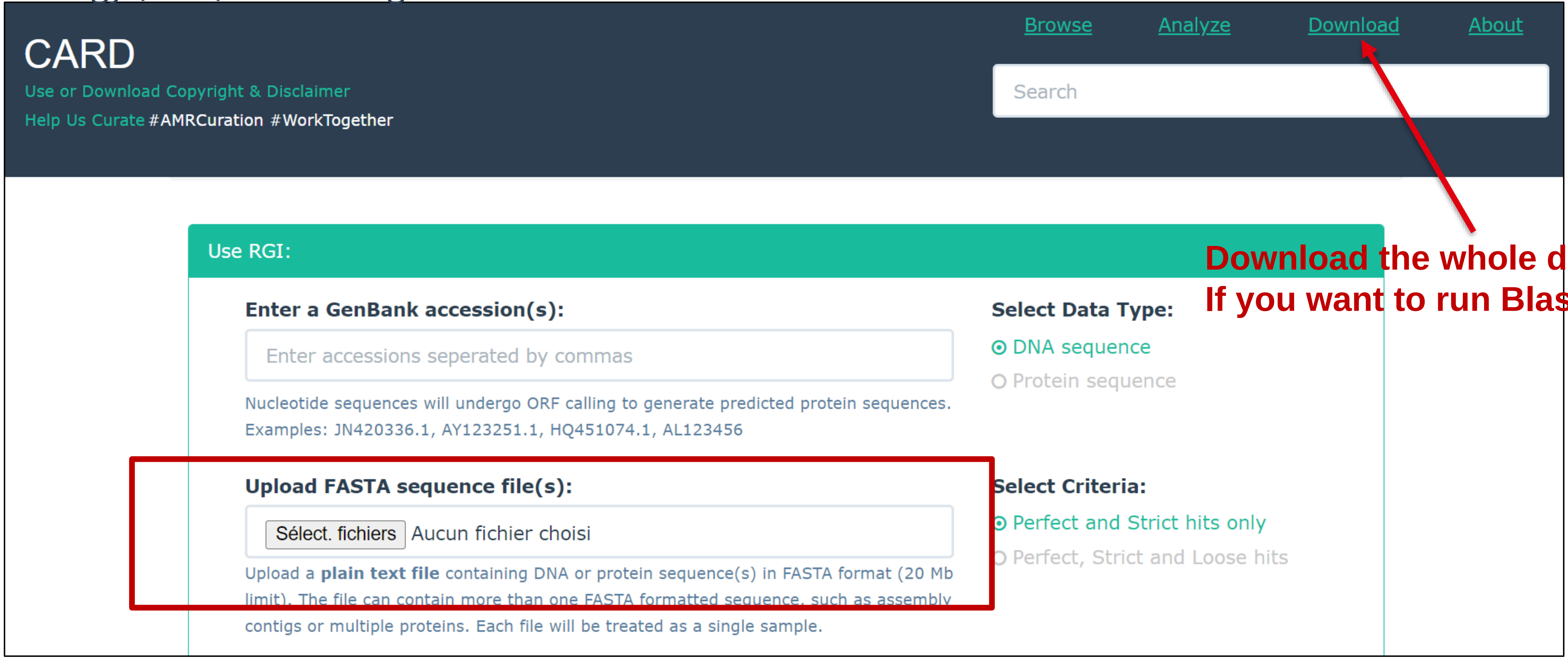
Paste the query sequence

https://ifr48.timone.univ-mrs.fr/blast/arg-annot_v6.html

No registration needed

Search for antimicrobial resistance genes using BLAST web-tools

The Comprehensive Antibiotic Resistance Database: a rigorously curated collection of characterized, peer-reviewed resistance determinants and associated antibiotics, organized by the Antibiotic Resistance Ontology (ARO) and AMR gene detection models.



CARD

Use or Download Copyright & Disclaimer

Help Us Curate #AMRCuration #WorkTogether

Use RGI:

Enter a GenBank accession(s):

Enter accessions seperated 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):

Sélect. fichiers Aucun fichier choisi

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.

Select Data Type:

● DNA sequence

○ Protein sequence

Select Criteria:

● Perfect and Strict hits only

○ Perfect, Strict and Loose hits

Download the whole database
If you want to run Blast locally

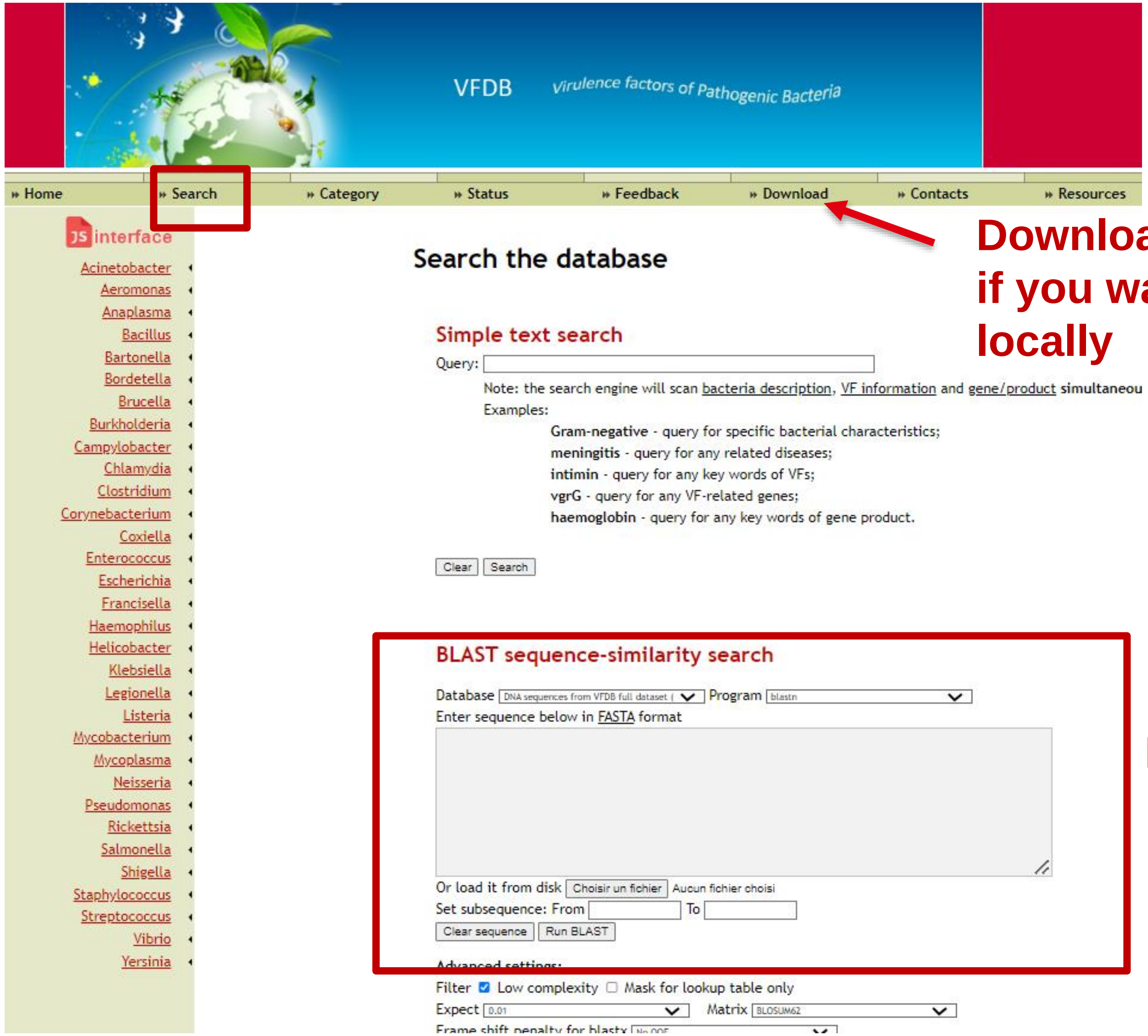
https://card.mcmaster.ca/

Search for virulence factors using BLAST web-tools

The virulence factor database (VFDB) is an integrated and comprehensive online resource for curating information about virulence factors of bacterial pathogen

http://www.mgc.ac.cn/cgi-bin/VFs/genus.cgi?Genus=Staphylococcus

No registration needed



Search the database

Simple text search

Query:

Note: the search engine will scan bacteria description, VF information and gene/product simultaneou

Examples:

Gram-negative - query for specific bacterial characteristics;
meningitis - query for any related diseases;
intimin - query for any key words of VFs;
vgrG - query for any VF-related genes;
haemoglobin - query for any key words of gene product.

BLAST sequence-similarity search

Database DNA sequences from VFDB full dataset (Program blastn

Enter sequence below in FASTA format

Or load it from disk Aucun fichier choisi

Set subsequence: From To

Advanced settings:

Filter ☑ Low complexity ☐ Mask for lookup table only

Expect 0.01 Matrix BLOSUM62

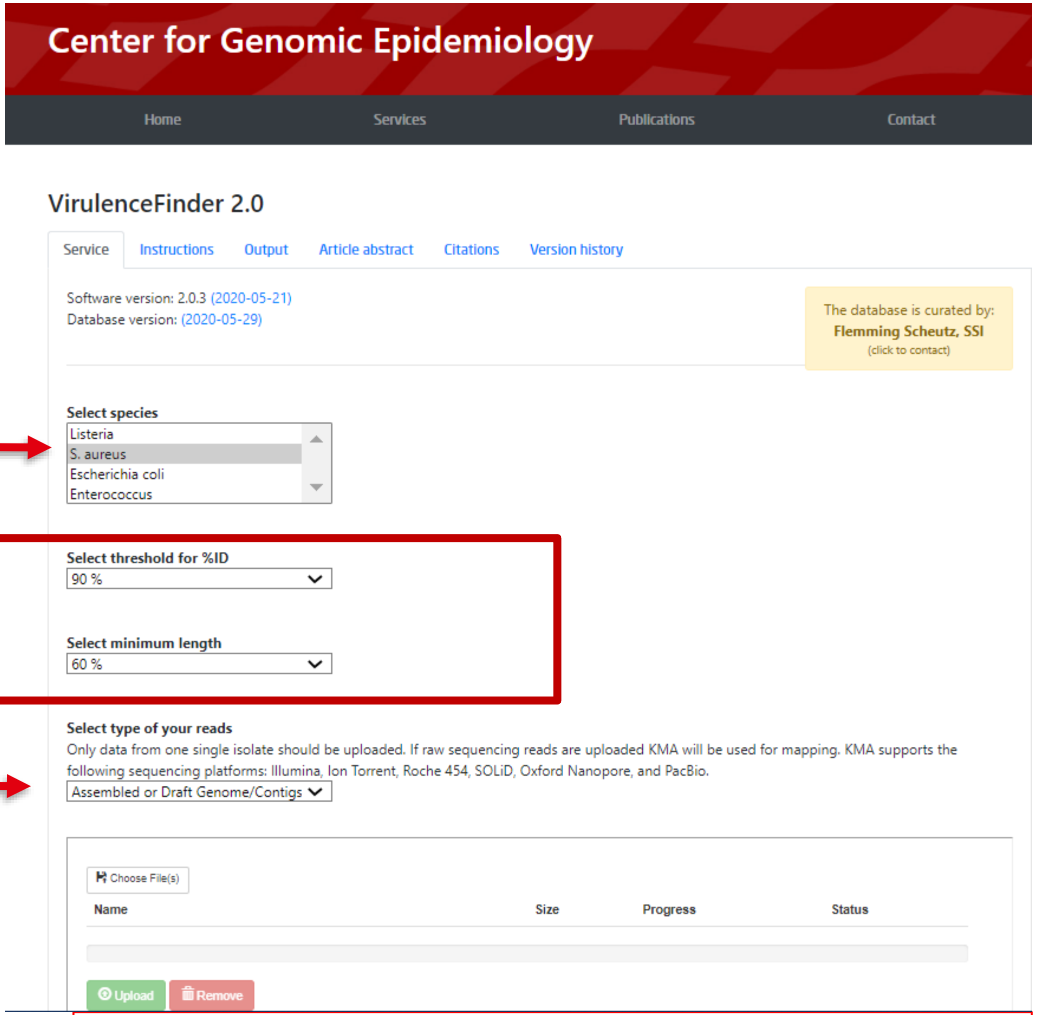
Frame shift penalty for blastx No OOF

Download the whole database if you want to run the blast locally

Paste query sequence

Search for virulence factors using BLAST web-tools

CGE tools



Center for Genomic Epidemiology

VirulenceFinder 2.0

Software version: 2.0.3 (2020-05-21)
Database version: (2020-05-29)

The database is curated by:
Flemming Scheutz, SSI
(click to contact)

Select species
Listeria
S. aureus
Escherichia coli
Enterococcus

Select threshold for %ID
90 %

Select minimum length
60 %

Select type of your reads
Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.
Assembled or Draft Genome/Contigs

Select species

Define coverage and % identity

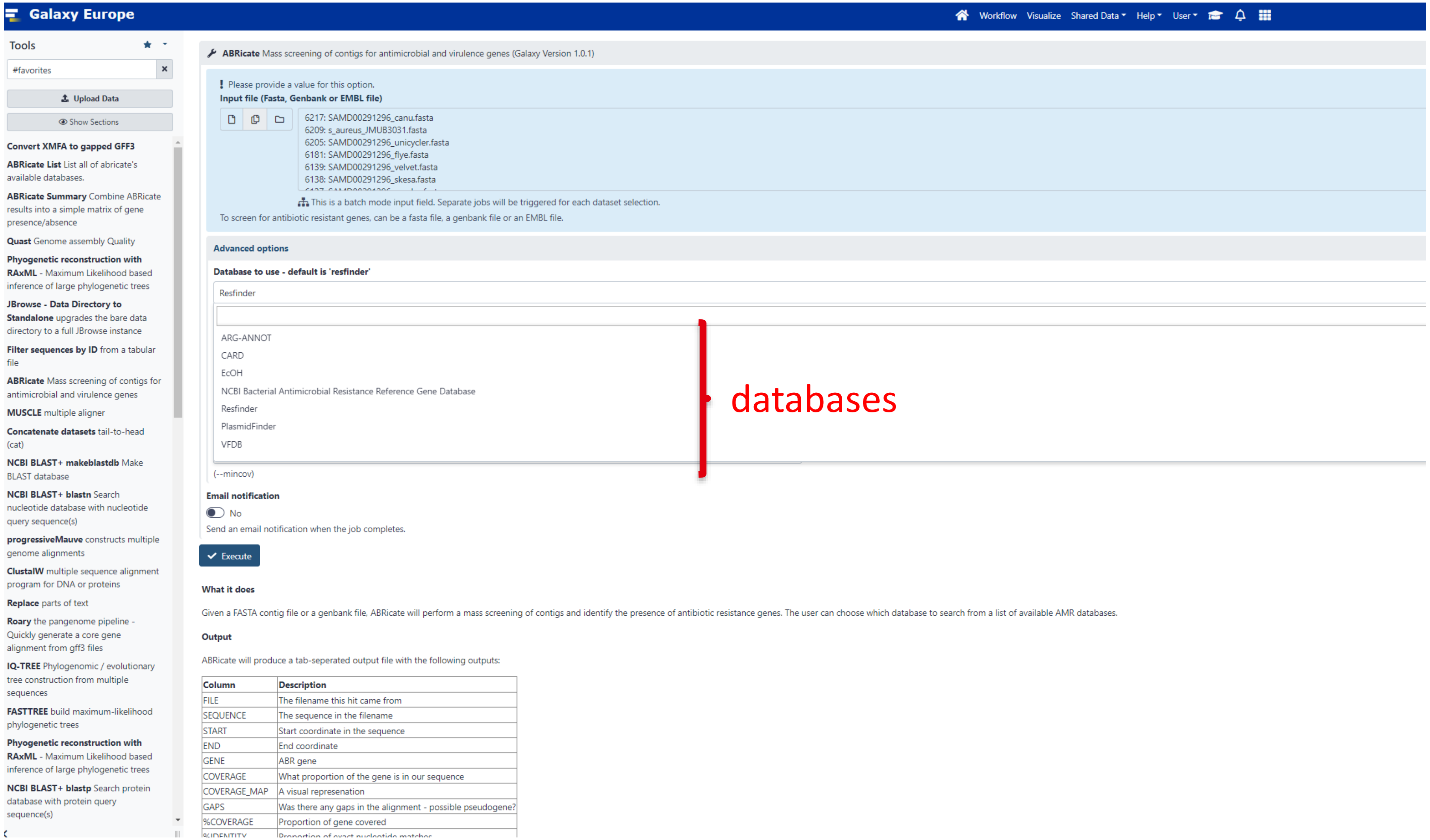
Type of input

https://cge.food.dtu.dk/services/VirulenceFinder/

No registration needed

Search for antimicrobial resistance genes and virulence factors

ABRICATE on Galaxy



Convert XMFA to gapped GFF3

ABRicate List List all of abricate's available databases.

ABRicate Summary Combine ABRicate results into a simple matrix of gene presence/absence

Quast Genome assembly Quality

Phyogenetic reconstruction with RAxML - Maximum Likelihood based inference of large phylogenetic trees

JBrowse - Data Directory to Standalone upgrades the bare data directory to a full JBrowse instance

Filter sequences by ID from a tabular file

ABRicate Mass screening of contigs for antimicrobial and virulence genes

MUSCLE multiple aligner

Concatenate datasets tail-to-head (cat)

NCBI BLAST+ makeblastdb Make BLAST database

NCBI BLAST+ blastn Search nucleotide database with nucleotide query sequence(s)

progressiveMauve constructs multiple genome alignments

ClustalW multiple sequence alignment program for DNA or proteins

Replace parts of text

Roary the pangenome pipeline - Quickly generate a core gene alignment from gff3 files

IQ-TREE Phylogenomic / evolutionary tree construction from multiple sequences

FASTTREE build maximum-likelihood phylogenetic trees

Phyogenetic reconstruction with RAxML - Maximum Likelihood based inference of large phylogenetic trees

NCBI BLAST+ blastp Search protein database with protein query sequence(s)

ABRicate Mass screening of contigs for antimicrobial and virulence genes (Galaxy Version 1.0.1)

! Please provide a value for this option.

Input file (Fasta, Genbank or EMBL file)

6217: SAMD00291296_canu.fasta
6209: s_aureus_JMUB3031.fasta
6205: SAMD00291296_unicycler.fasta
6181: SAMD00291296_flye.fasta
6139: SAMD00291296_velvet.fasta
6138: SAMD00291296_skesa.fasta

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

To screen for antibiotic resistant genes, can be a fasta file, a genbank file or an EMBL file.

Advanced options

Database to use - default is 'resfinder'

Resfinder

ARG-ANNOT
CARD
EcOH
NCBI Bacterial Antimicrobial Resistance Reference Gene Database
Resfinder
PlasmidFinder
VFDB

databases

(--mincov)

Email notification

No

Send an email notification when the job completes.

Execute

What it does

Given a FASTA contig file or a genbank file, ABRicate will perform a mass screening of contigs and identify the presence of antibiotic resistance genes. The user can choose which database to search from a list of available AMR databases.

Output

ABRicate will produce a tab-seperated output file with the following outputs:

Column	Description
FILE	The filename this hit came from
SEQUENCE	The sequence in the filename
START	Start coordinate in the sequence
END	End coordinate
GENE	ABR gene
COVERAGE	What proportion of the gene is in our sequence
COVERAGE_MAP	A visual represenation
GAPS	Was there any gaps in the alignment - possible pseudogene?
%COVERAGE	Proportion of gene covered
%IDENTITY	Proportion of exact nucleotide matches

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

anses

Thank you