



MLST - Multi-Locus Sequence Typing

Serotype - Serotype analysis

Gene identification – Drug resistance genes

Core genome construction

Molecular clock analysis

Prokka annotation

Bayesian clustering

MLST - Multi-Locus Sequence Typing



PubMLST

Public databases for molecular typing and microbial genome diversity

MY ACCOUNT

Home

HOME

ORGANISMS

SPECIES ID

ABOUT US

UPDATES

MLST allelic profiles and sequences

This page contains download links to alleles and profiles that define classic MLST schemes. Where these are hosted on PubMLST, they represent a small subset of the data hosted, all of which is available via the [application programming interface \(API\)](#). A few of the schemes are hosted at [Pasteur](#) and the links point to their API. This list is not being maintained for new schemes - all data are available via the API and this is the recommended way to retrieve data programmatically.

Download

The information in this table is available in [XML format](#) for automated parsing.

<https://pubmlst.org/data>

Candidatus Liberibacter solanacearum

profiles

adk; atpA; fbpA; ftsZ; glyA; groEL; gyrB

```
>aroA_1
GGAGAGTCGGTCATCGTAGAAAAAGAGTTGACTCGAAACCATACAGAAGATATGATTGTC
CAGTTTGGTGGACAGTTAGAAAGTCAATGGCAAGGAAATCCGCATCCAAGGTGGTCAGGAG
TTTATTGCCCAAGAGATTACAGTTCCAGGAGATATTTCAAGTGCTGCTTTTGGTTGGTT
GCTGGCTTAATCATACACAGGTTCAAAAATTGTCCTTGAAAATGTGGGAATCAATGAACT
CGGACTGGTATTTTAGATGTCATTAAAGCTATGGGTGGTAAAATGACTCTTTCTAACATA
GATGAACCTGCAAAATCTGCTACCATTACAGTTGAAACGTCGGAATTGAAGGCTACGGAG
ATTGCA
>aroA_2
GGAGAGTCGGTCATCGTAGAAAAAGAGTTGACTCGAAACCATACAGAAGATATGATTGTC
CAGTTTGGTGGACAGTTAGAAAGTCAATGGCAAGGAAATCCGCATCCAAGGTGGTCAGGAG
TTTATTGCCCAAGAGATTACAGTTCCAGGAGATATTTCAAGTGCTGCTTTTGGTTGGTT
GCTGGCTTAATCATACACAGGTTCAAAAATTGTCCTTGAAAATGTGGGAATCAATGAACT
CGGACTGGTATTTTAGATGTCATTAAAGCTATGGGTGGTAAAATGACTCTTTCTAACATA
GATGAACCTGCAAAATCTGCTACCATTACAGTTGAAACGTCGGAATTGAAGGCTACGGAG
ATTGCA
>aroA_3
GGAGAGTCGGTCATCGTAGAAAAAGAGTTGACTCGAAACCATACAGAAGATATGATTGTC
CAGTTTGGTGGACAGTTAGAAAGTCAATGGCAAGGAAATCCGCATCCAAGGTGGTCAGGAG
TTTATTGCCCAAGAGATTACAGTTCCAGGAGATATTTCAAGTGCTGCTTTTGGTTGGTT
GCTGGCTTAATCATACACAGGTTCAAAAATTGTCCTTGAAAATGTGGGAATCAATGAACT
CGGACTGGTATTTTAGATGTCATTAAAGCTATGGGTGGTAAAATGACTCTTTCTAACATA
GATGAACCTGCAAAATCGGCTACCATTACAGTTGAAACGTCGGAATTGAAGGCTACGGAG
ATTGCA
```

The housekeeping gene sequence was copied into the same fa file in order (official website order)

Housekeeping genes

名称	修改日期	类型	大小
GCF_000014305_1.fna	2023/12/13 13:35	Audio Shark	2,073 KB
GCF_000014325_1.fna	2023/12/13 13:35	Audio Shark	2,073 KB
GCF_000018185_1.fna	2023/12/13 13:35	Audio Shark	2,016 KB
GCF_000026725_1.fna	2023/12/13 13:35	Audio Shark	2,073 KB
GCF_000026745_1.fna	2023/12/13 13:35	Audio Shark	2,147 KB
GCF_000091905_1.fna	2023/12/13 13:35	Audio Shark	1,986 KB
GCF_000168355_3.fna	2023/12/13 13:35	Audio Shark	2,152 KB
GCF_000186405_1.fna	2023/12/13 13:35	Audio Shark	2,114 KB
GCF_000204625_1.fna	2023/12/13 13:35	Audio Shark	2,007 KB
genomad_2.fna	2024/2/26 19:46	Audio Shark	4,197 KB
genomad_5.fna	2024/2/26 20:57	Audio Shark	8,342 KB
genomad_10.fna	2024/2/26 20:57	Audio Shark	16,684 KB
genomad_15.fna	2024/2/26 20:58	Audio Shark	25,026 KB
genomad_20.fna	2024/2/27 22:23	Audio Shark	33,368 KB

Input folder: Put all fa files into the same folder

Existing ST library

	A	B	C	D	E	F	G	H
1	mlst	aroA	cpn60	dpr	gki	mutS	recA	thrA
2	1	1	1	1	1	1	1	1
3	2	1	1	1	2	1	1	1
4	3	1	1	1	16	1	1	1
5	4	9	1	1	1	1	1	1
6	5	1	1	1	40	1	1	1
7	6	1	1	1	1	1	1	2
8	7	1	1	1	1	1	1	3
9	8	1	1	1	1	28	1	1
10	9	1	1	1	35	1	1	1
11	10	1	10	1	1	1	1	1
12	11	3	1	1	1	1	1	1
13	12	3	1	1	1	1	2	1
14	13	1	12	1	1	6	21	21
15	14	18	1	5	12	1	10	1
16	15	8	8	5	12	1	10	4
17	16	5	17	5	12	1	10	4
18	17	8	1	5	12	1	10	1

ST type and housekeeping gene information (official website order) are retained, and redundant columns are deleted

MLST - Multi-Locus Sequence Typing



MLST

Input folder (fasta)

D:/MicroWorldOmics_Setup/Example_data/mlst/fna

Choose

Output folder

D:/Documents/Desktop/tmp

Choose

Ref fasta file

D:/MicroWorldOmics_Setup/Example_data/mlst/SS_MLST.fasta

Choose

MLST db (TAB, xlsx)

D:/MicroWorldOmics_Setup/Example_data/mlst/SS_ST_type_20221007.xlsx

Choose

Result table

	ID	MLST	ST_result
1	GCF_0000143...	945	261:1:1:301:1:...
2	GCF_0000143...	new	262:0:0:1:1:1:3
3	GCF_0000181...	1	1:1:1:1:1:1:1
4	GCF_0000267...	7	1:1:1:1:1:1:3
5	GCF_0000267...	1	1:1:1:1:1:1:1
6	GCF_0000919...	1	1:1:1:1:1:1:1
7	GCF_0001683...	28	2:...
8	GCF_0001864...	7	1:1:1:1:1:1:3
9	GCF_0002046...	35	31:30:5:34:31:...
10	genomad_10	35	31:30:5:34:31:...
11	genomad_15	35	31:30:5:34:31:...
12	genomad_2	new	262:1:1:1:1:1:3
13	genomad_20	35	31:30:5:34:31:...
14	genomad_5	35	31:30:5:34:31:...

Status

Finished!!!

Status

Running! please wait

Table

Run

If the program is finished, click 'Table' to display the result

Input bacterial genomes folder (contigs)

The target folder of the output file (user-defined)

Housekeeping genes associated with bacteria

Existing ST library

The first column represents the number of the strain, the second column represents the ST type, and the third column represents the combination of the number of housekeeping genes detected

Click this button to display the table

The status bar is completed

The status bar is running

Serotype - Serotype analysis



Input bacterial genomes folder (contigs)

The target folder of the output file (user-defined)

Serotype gene, collected from the literature

Serotype

Input folder (fasta)

D:/MicroWorldOmicSetup/Example_data/serotype/fna

Choose

Output folder

D:/Documents/Desktop/tmp

Choose

Ref fasta file

D:/MicroWorldOmicSetup/Example_data/serotype/serotype.fasta

Choose

Profile (csv, optional)

Default: NULL

Choose

Exact fasta file (optional)

Default: NULL

Choose

Search mode

☒ Normal mode

☐ Precise mode

Status

Running! please wait
If no response, never close window!!!

Status bar

Result table

	ID	Serotype
1	GCF_0000143...	Serotype_2
2	GCF_0000143...	Serotype_2
3	GCF_0000181...	Serotype_2
4	GCF_0000267...	Serotype_2
5	GCF_0000267...	Serotype_2

If the program is finished, click 'Table' to display the result

Table

Run

Designed for the differentiation of Streptococcus suis serotypes 1 and 14, as well as types 2 and 1/2

The precise mode is targeted at Streptococcus suis serotypes 1 and 14, as well as types 2 and 1/2

The first column represents the strain number, and the second column represents the serotype

Click this button to display the table

Gene identification – Drug resistance genes



Input bacterial genomes
folder (contigs)

The target folder of the
output file (user-defined)

Target gene fa
(Drug resistance
genes, If there are
multiple genes, they
should be placed in
the same fa file)

The identity
threshold with
the target gene

Nucleic acid
identification
and protein
identification

Gene identification

Gene identification	
Input folder (fasta) D:/MicroWorldOmics_Setup/ Example_data/GeneIdentification/fna Choose	Output folder D:/Documents/Desktop/tmp Choose
Ref fasta file D:/MicroWorldOmics_Setup/ Example_data/GeneIdentification/ nuc.fasta Choose	Same position coverage (%) Default: 80 (%)
Identification (%) Default: 90 (%)	Gene coverage (%) Default: 90 (%)
Blast mode <input checked="" type="radio"/> Nucleic mode <input type="radio"/> Protein mode	
Status Running! please wait If no response, never close window!!! Status bar	
Run	

Result table

	Genome_ID	Gene_ID
1	GCF_0000143...	gb ...
2	GCF_0000143...	gb AB039845....
3	GCF_0000143...	gb ...

If the program is finished, click 'Table' to display the result

Table

The first column
represents the strain
number, and the second
column represents the
identified gene

If the overlapping part of two genes at the
same position that meet the condition is
greater than this threshold, take the one with
the greatest identity

The coverage threshold with the
target gene

Click this
button to
display the
table

Core genome construction



Core genome analysis

Core genome concept

Identification (%)

Default: 80 (%)

Gene coverage (%)

Default: 80 (%)

Input folder (fasta)

D:/MicroWorldOmics_Setup/Example_data/core_genome/fna

Output folder

D:/Documents/Desktop/tmp

Ref fasta file (ffn)

D:/MicroWorldOmics_Setup/Example_data/core_genome/SC19.ffn

Status

Running! please wait
If no response, never close window!!!

Run

Input bacterial genomes folder (contigs)

The reference strains of this category of bacteria need to be first annotated in ffn format using prokka

The identity threshold to identify core genes

The coverage threshold to identify core genes

The target folder of the output file (user-defined)

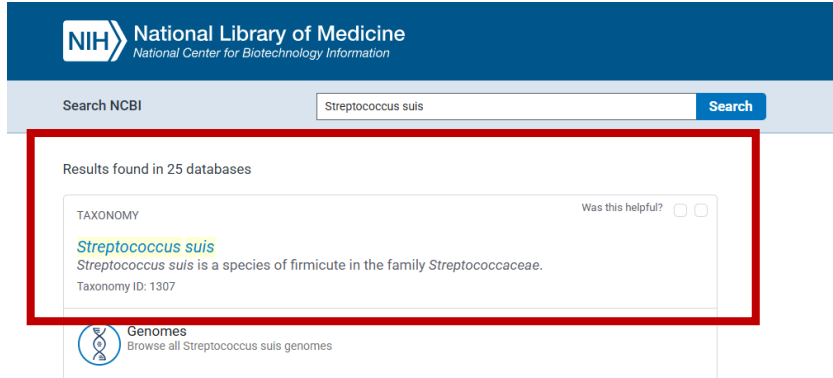
The status bar is running

The status bar is completed

Status

Finished, Core_genome.aln is your result!!!

Core genome construction



<https://www.ncbi.nlm.nih.gov/datasets/taxonomy/1307/>



Take *Streptococcus suis* as an example. Download the genome of *Streptococcus suis* from the following website for reference, and then annotate it with prokka to obtain the ffn file

Result

All_core_genes_aln.fasta	2025/7/7 13:33	FASTA 文件	8 KB
All_isolates_core_genes.fna	2025/7/7 12:41	Audio Shark	14,066 KB
core_gene_list.out	2025/7/7 12:41	OUT 文件	35 KB
Core_genome.aln	2025/7/7 13:33	ALN 文件	8 KB
Isolates_gene_number.out	2025/7/7 12:41	OUT 文件	1 KB

1. All_core_genes_aln.fasta: please ignore the intermediate files of the software operation
2. All_isolates_core_genes.fna: the fa file of the core gene
3. core_gene_list.out : the entry of the core gene, ':' is followed by the length of the core gene
4. Core_genome.aln: The file with all core genes aligned for each isolate is used to build an evolutionary tree
5. Isolates_gene_number.out: The number of genes in each isolate

Molecular clock analysis



ShinyBactDating

Status

Starting ShinyApp!!!

Begin

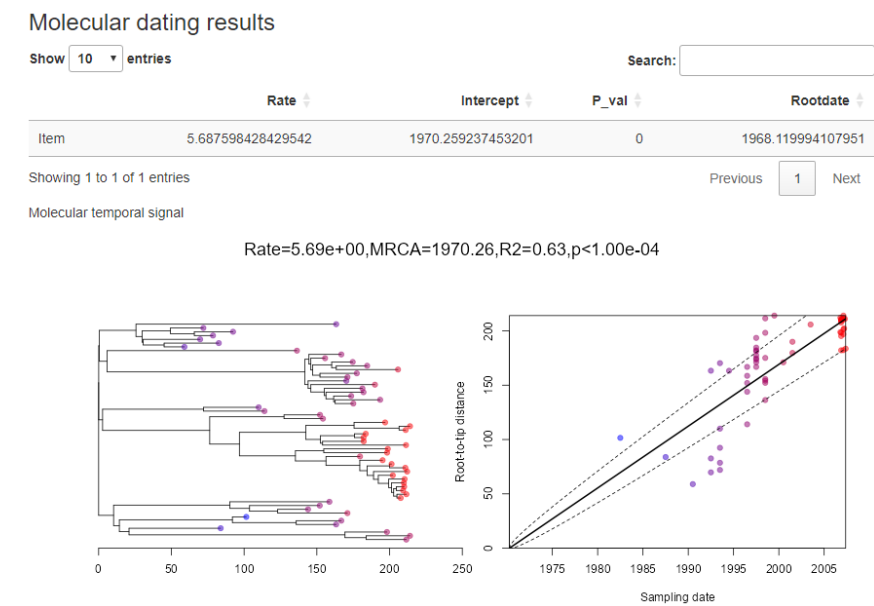
ShinyApp has been started!!!

Finish

Please close this window before starting another Shiny Apps!!!

Start App Click to start

Open Web Open Shiny app



Molecular dating by BactDating

First: Choose tree (newick)

Browse... example.nwk

Upload complete

Step 2: Choose date file

Browse... date.csv

Upload complete

if TRUE, the most distant tip from the root is considered as the origin of the time scale; if FALSE, this is the root node.

☐ True

☒ False

Show labels

☐ True

☒ False

RESET

Display the label
"Molecular dating tree"

"Rate" represents a
temporal signals

Plot output format

☒ PNG ☐ PDF ☐ JPEG

Download molecular dating tree

Download temporal signal

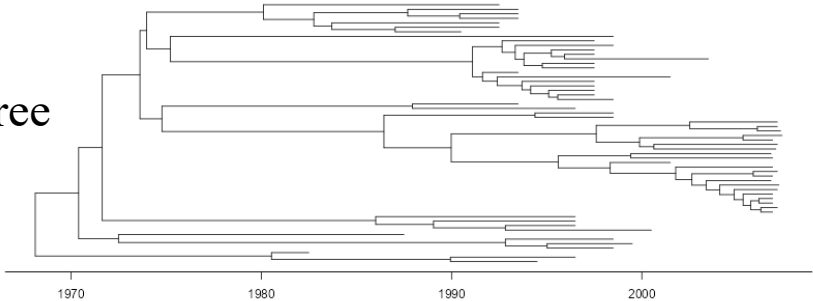
Panel

Input file 1: Tree file (newick)

Input file 2: date.csv
"id" represents the
name of the isolate

id	time
1	1998.5
2	1999.5
3	1998.5
4	1987.5
5	1994.5
6	1996.5
7	1982.5
8	2000.5
9	1996.5
10	1996.5

Molecular dating tree



Prokka annotation



Input genome (contigs)

The target folder of the output file (user-defined)

Result prefix

Archaea | Bacteria | Mitochondria | Viruses

Prokka

Input fasta file

D:/input/test.fa

Choose

Output folder

D:/output

Choose

Prefix

example

Kingdom

Bacteria

Status

Updating the file!!!

The status bar is running

Run

The interpretation of the results, please reference prokka:

<https://github.com/tseemann/prokka>

Bayesian clustering



ShinyRhierBaps

Status

Starting ShinyApp!!!

Begin

ShinyApp has been started!!!

Finish

Please close this window before starting another Shiny Apps!!!

Start App

Click to start

Open Web

Open Shiny app

BAPs cluster table

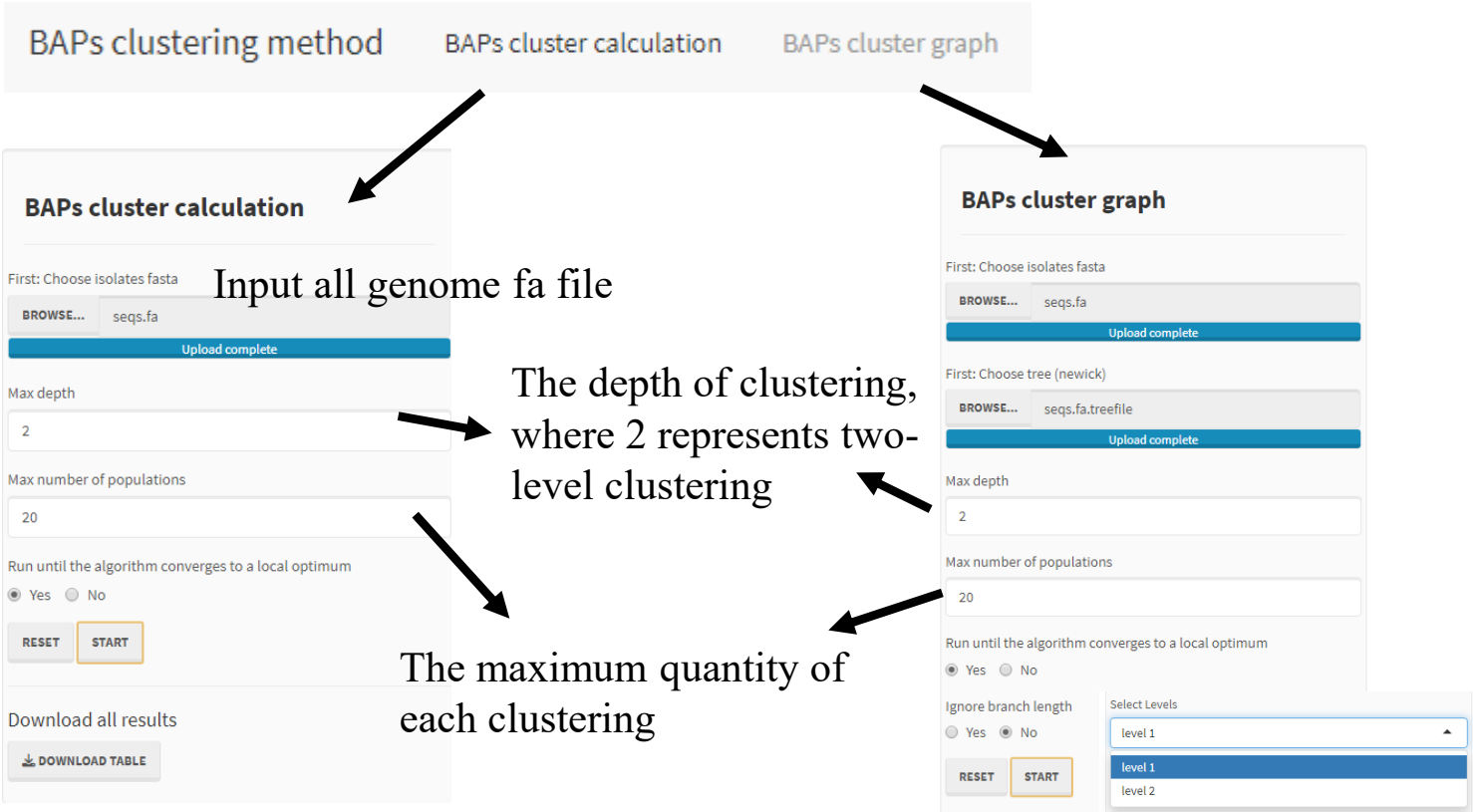
Show 10 entries

Search:

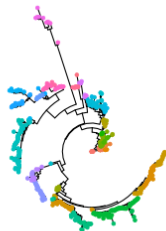
	Isolate	level 1	level 2
1	1	1	1
2	2	1	1
3	3	1	1
4	4	2	5
5	5	3	9
6	6	3	9
7	7	3	9
8	8	4	16
9	9	5	19
10	10	2	6

Showing 1 to 10 of 515 entries

Previous 1 2 3 4 5 ... 52 Next



Bayesian clustering at different levels



Select the visual clustering level