Epidemiological analysis

MLST - Multi-Locus Sequence Typing

Serotype - Serotype analysis

Gene identification – Drug resistance genes

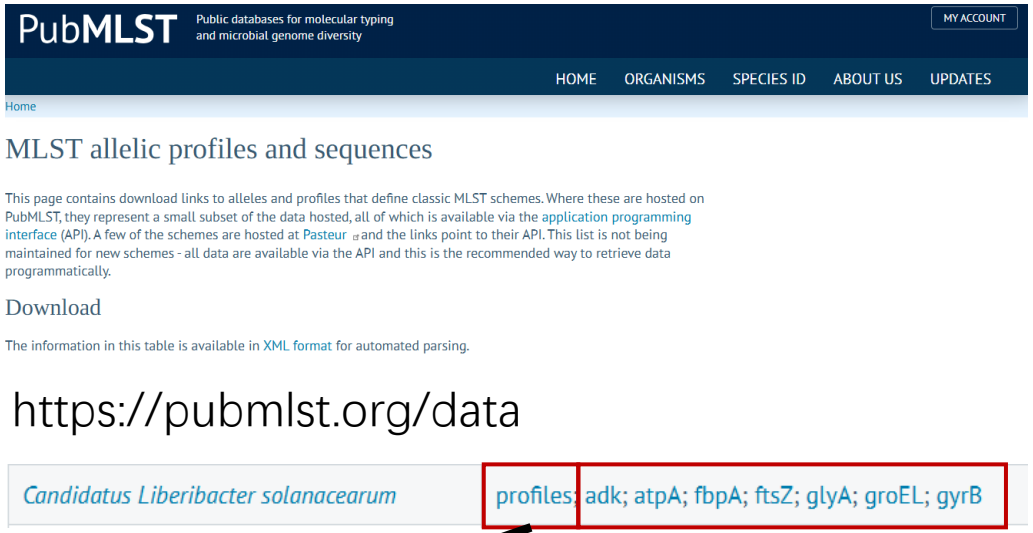
Core genome construction

Molecular clock analysis

Prokka annotation

Bayesian clustering

MLST - Multi-Locus Sequence Typing



https://pubmlst.org/data

Candidatus Liberibacter solanacearum	profiles: adk; atpA; fbpA; ftsZ; glyA; groEL; gyrB

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

```
>aroA_1
GGAGAGTCGGTCATCGTAGAAAAAGAGTTGACTCGAAACCATACAGAAGATATGATTGTC
CAGTTTGGTGGACAGTTAGAAGTCAATGGCAAGGAAATCCGCATCCAAGGTGGTCAGGAG
TTTATTGCCCAAGAGATTACAGTTCCAGGAGATATTTCAAGTGCTGCTTTTTGGTTGGTT
GCTGGCTTAATCATACCAGGTTCAAAAATTGTCCTTGAAAATGTGGGAATCAATGAAACT
CGGACTGGTATTTTAGATGTCATTAAAGCTATGGGTGGTAAAATGACTCTTTCTAACATA
GATGAACTTGCAAAATCTGCTACCATTACAGTTGAAACGTCGGAATTGAAGGCTACGGAG
ATTGCA
>aroA_2
GGAGAGTCGGTCATCGTAGAAAAAGAGTTGACTCGAAACCATACAGAAGATATGATTGTC
CAGTTTGGTGGACAGTTAGAAGTCAATGGCAAGGAAATCCGCATCCAAGGTGGTCAGGAG
TTTATTGCCCAAGAGATTACAGTTCCAGGAGATATTTCAAGTGCTGCTTTTTGGTTGGTT
GCTGGCTTAATCATACCAGGTTCAAAAATTGTCCTTGAAAATGTGGGAATCAATGAAACT
CGGACTGGTATTTTAGATGTCATTAAAGCTATGGGTGGTAAAATGACTTTTTCTAACATA
GATGAACTTGCAAAATCTGCTACCATTACAGTTGAAACGTCGGAATTGAAGGCTACGGAG
ATTGCA
>aroA_3
GGAGAGTCGGTCATCGTAGAAAAAGAGTTGACTCGAAACCATACAGAAGATATGATTGTC
CAGTTTGGTGGACAGTTAGAAGTCAATGGCAAGGAAATCCGCATCCAAGGTGGTCAGGAG
TTTATTGCCCAAGAGATTACAGTTCCAGGAGATATTTCAAGTGCTGCTTTTTGGTTGGTT
GCTGGCTTAATCATACCAGGTTCAAAAATTGTCCTTGAAAATGTGGGAATCAATGAAACT
CGGACTGGTATTTTAGATGTCATTAAAGCTATGGGTGGTAAAATGACTCTTTCTAACATA
GATGAACTTGCAAAATCGGCTACCATTACAGTTGAAACGTCGGAATTGAAGGCTACGGAG
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

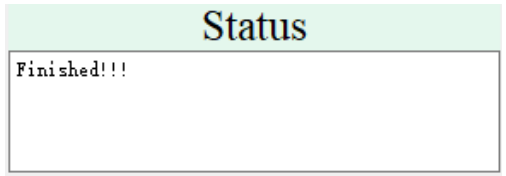
MLST - Multi-Locus Sequence Typing

Input bacterial genomes folder (contigs)

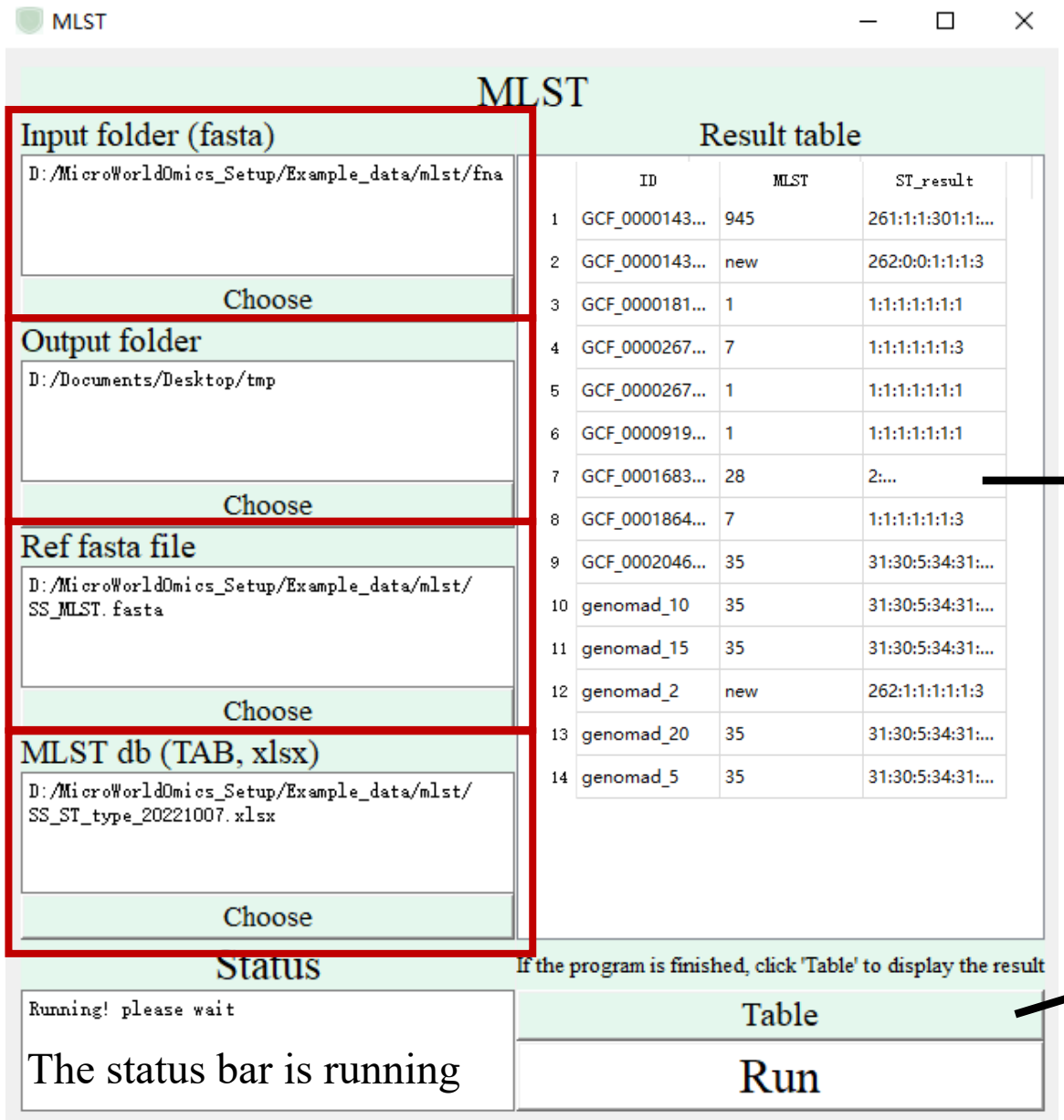
The target folder of the output file (user-defined)

Housekeeping genes associated with bacteria

Existing ST library



The status bar is completed



The status bar is running

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

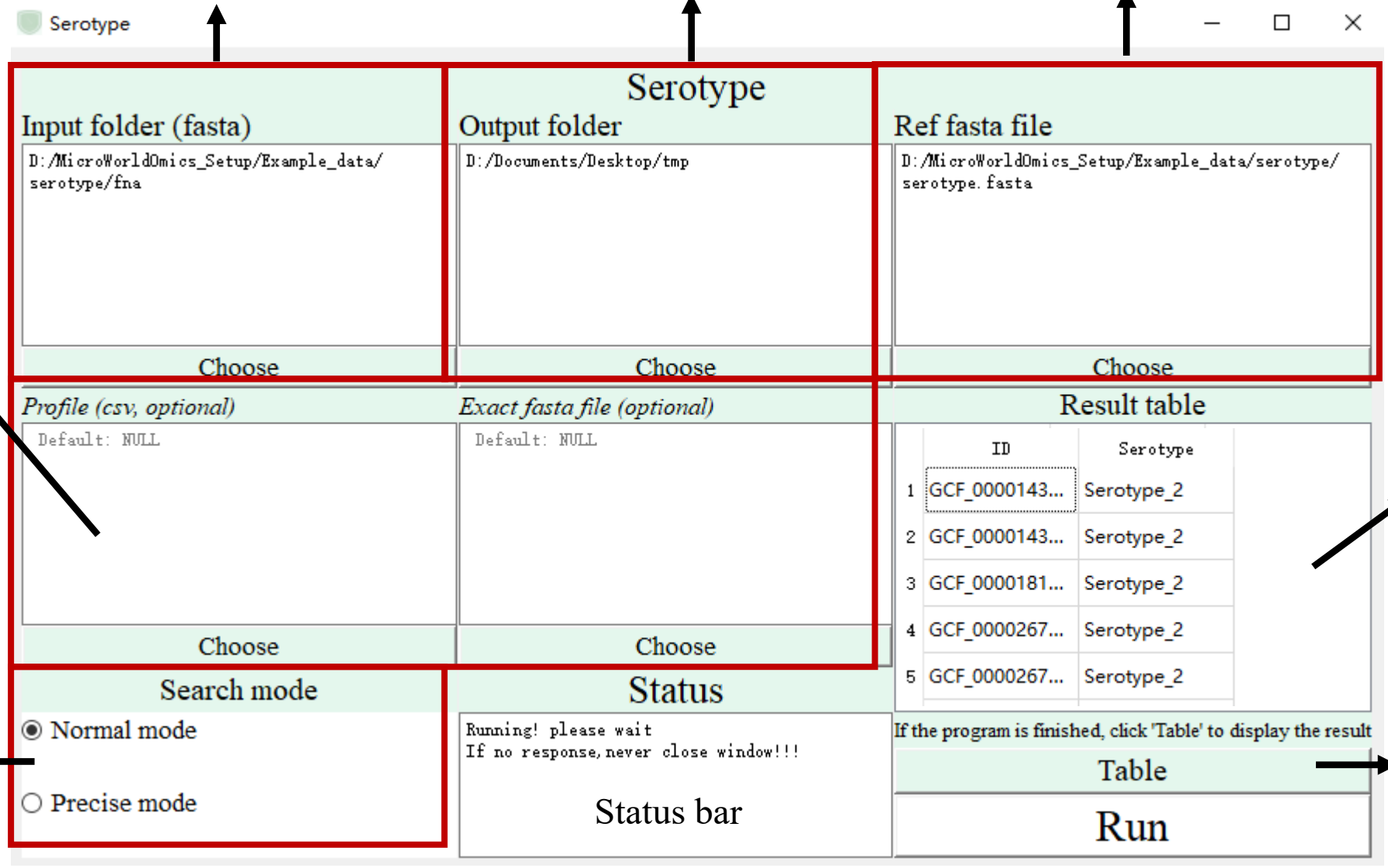
Click this button to display the table

Serotype - Serotype analysis

Input bacterial genomes folder (contigs)

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

Serotype gene, collected from the literature



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

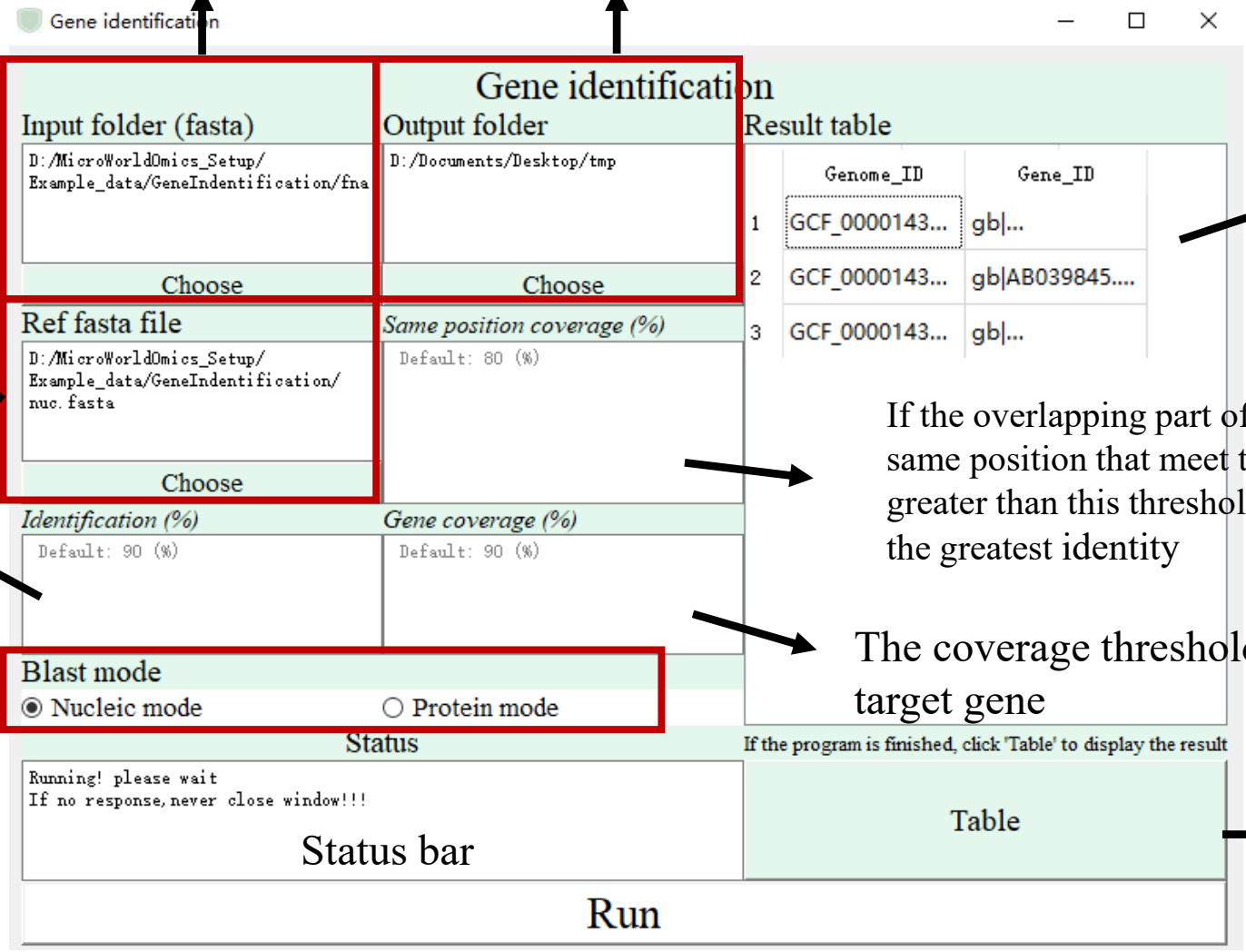
Click this button to display the table

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

The identity threshold with the target 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

Nucleic acid identification and protein identification

The coverage threshold with the target gene

Click this button to display the table

Core genome construction

Input bacterial genomes folder (contigs)

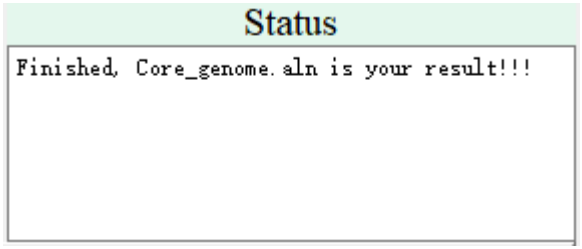
The reference strains of this category of bacteria need to be first annotated in ffn format using prokka

The identity threshold to identify core genes

The coverage threshold to identify core genes

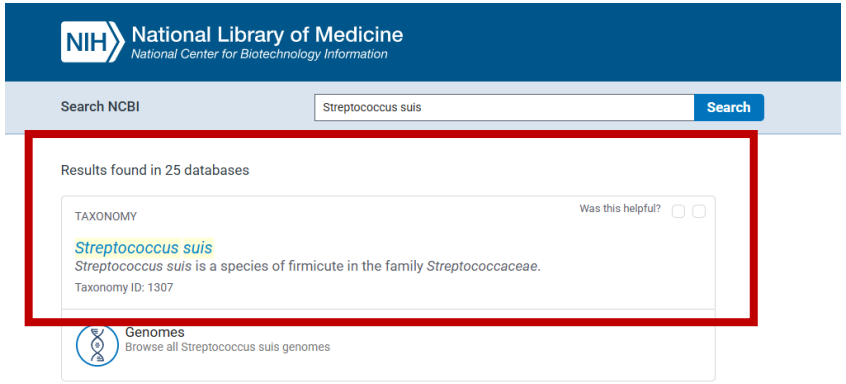
The target folder of the output file (user-defined)

The status bar is running

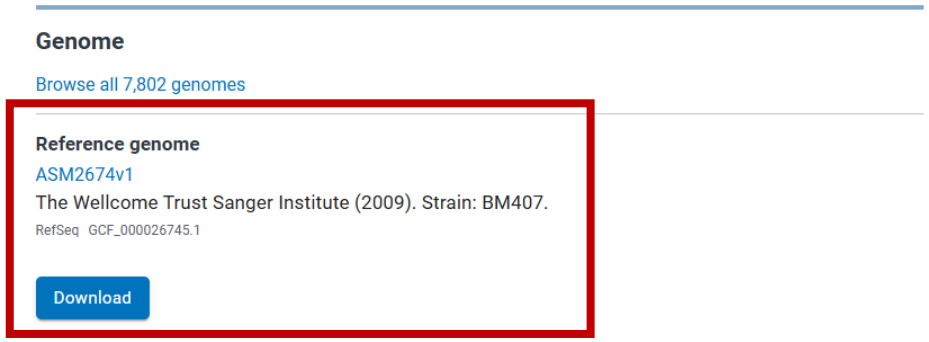


The status bar is completed

Core genome construction

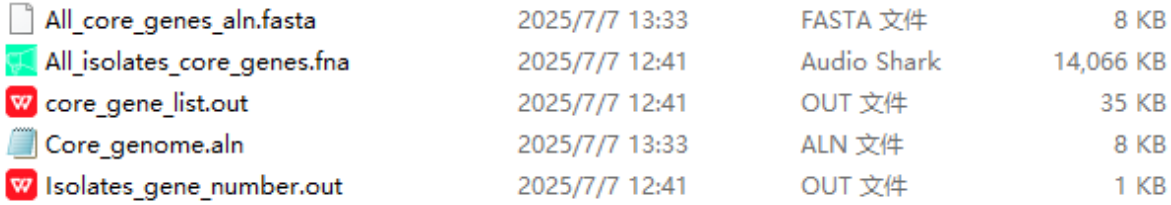


https://www.ncbi.nlm.nih.gov/datasets/taxonomy/1307/



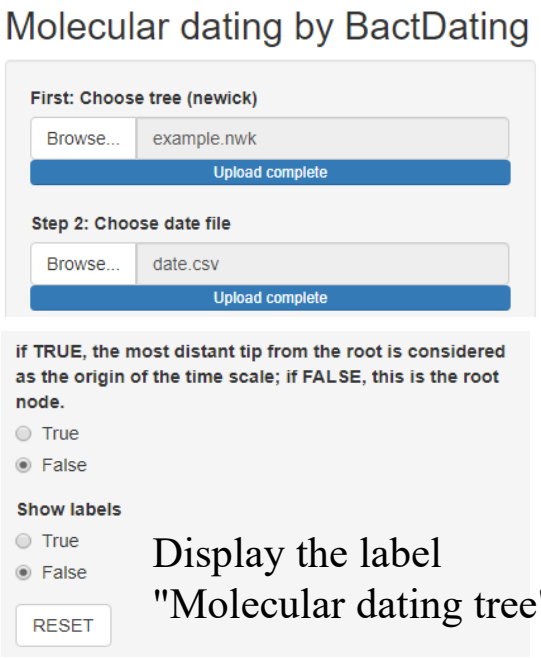
Take Streptococcus suis as an example. Download the genome of Streptococcus suis from the following website for reference, and then annotate it with prokka to obtain the ffn file

Result

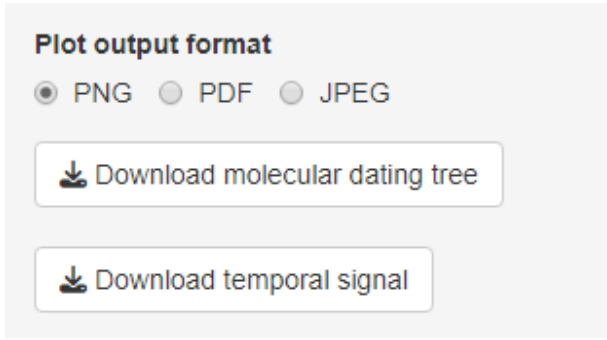


1. All_core_genes_aln.fasta: please ignore the intermediate files of the software operation
2. All_isolates_core_genes.fna: the fa file of the core gene
3. core_gene_list.out : the entry of the core gene,'::' is followed by the length of the core gene
4. Core_genome.aln: The file with all core genes 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



Display the label "Molecular dating tree"

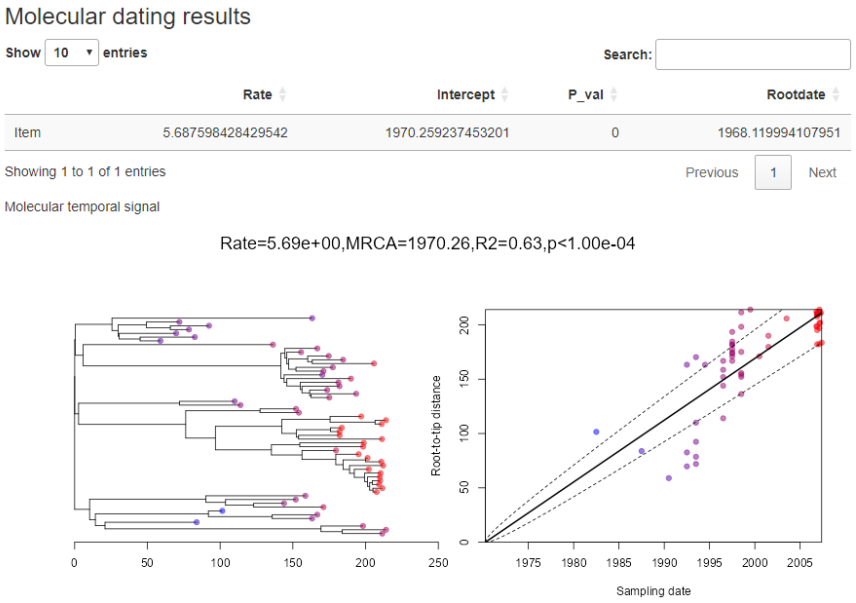


Panel

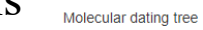
Input file 1: Tree file (newick)

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

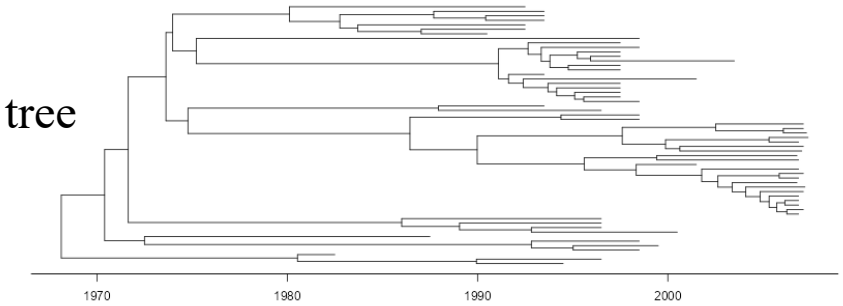
id	time
1	1998.5
2	1999.5
3	1998.5
4	1987.5
5	1994.5
6	1996.5
7	1982.5
8	2000.5
9	1996.5
10	1996.5



"Rate" represents a temporal signals



Molecular dating tree



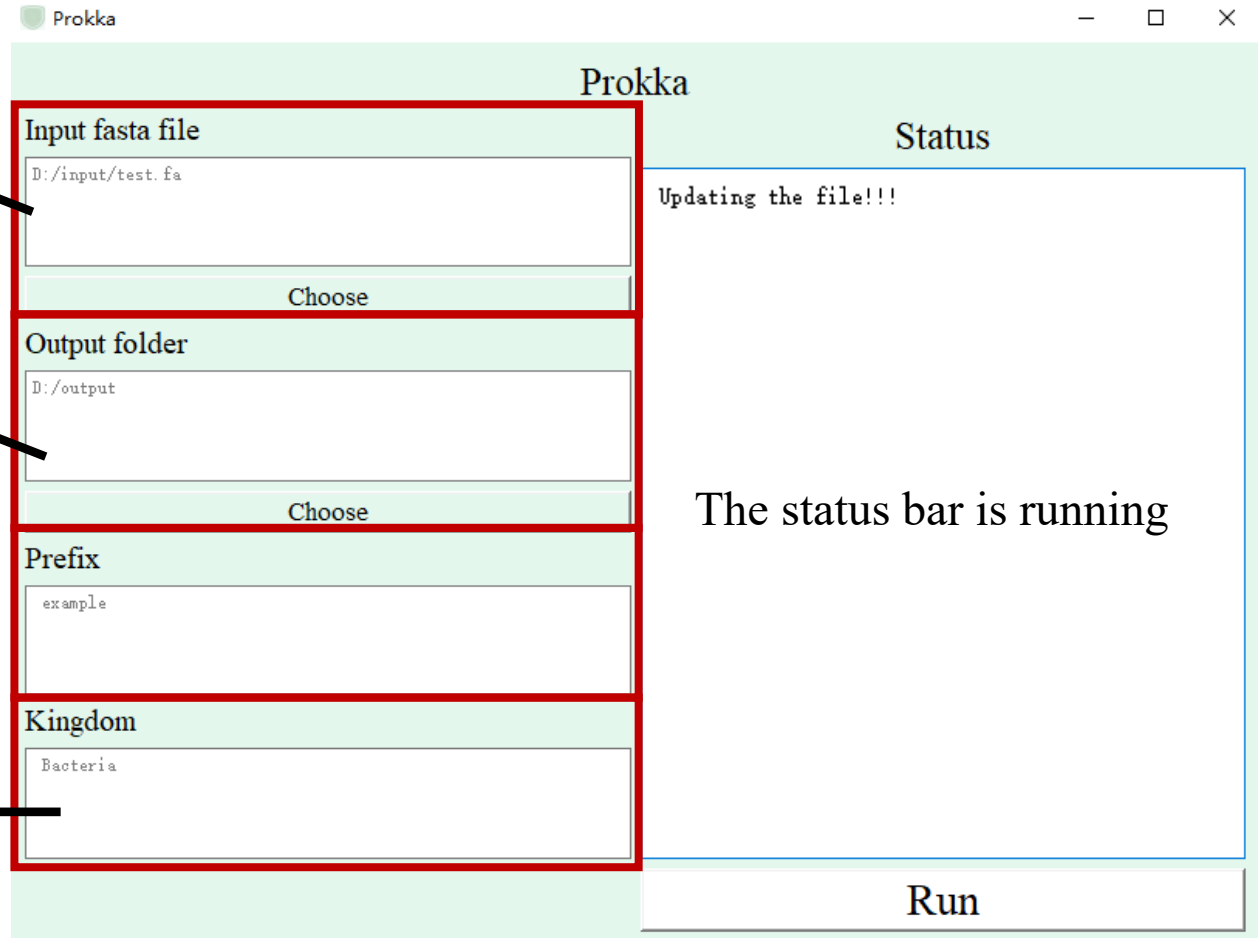
Prokka annotation

Input genome (contigs)

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

Result prefix

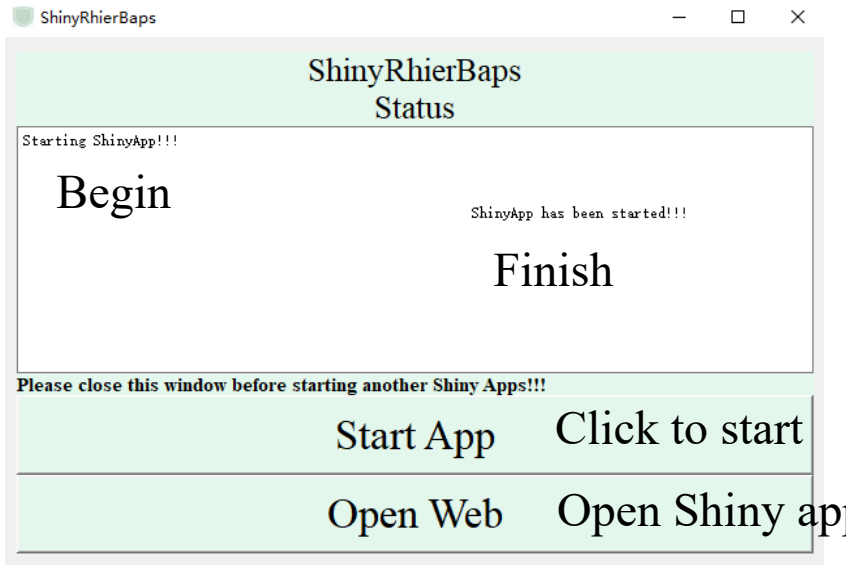
Archaea | Bacteria | Mitochondria | Viruses

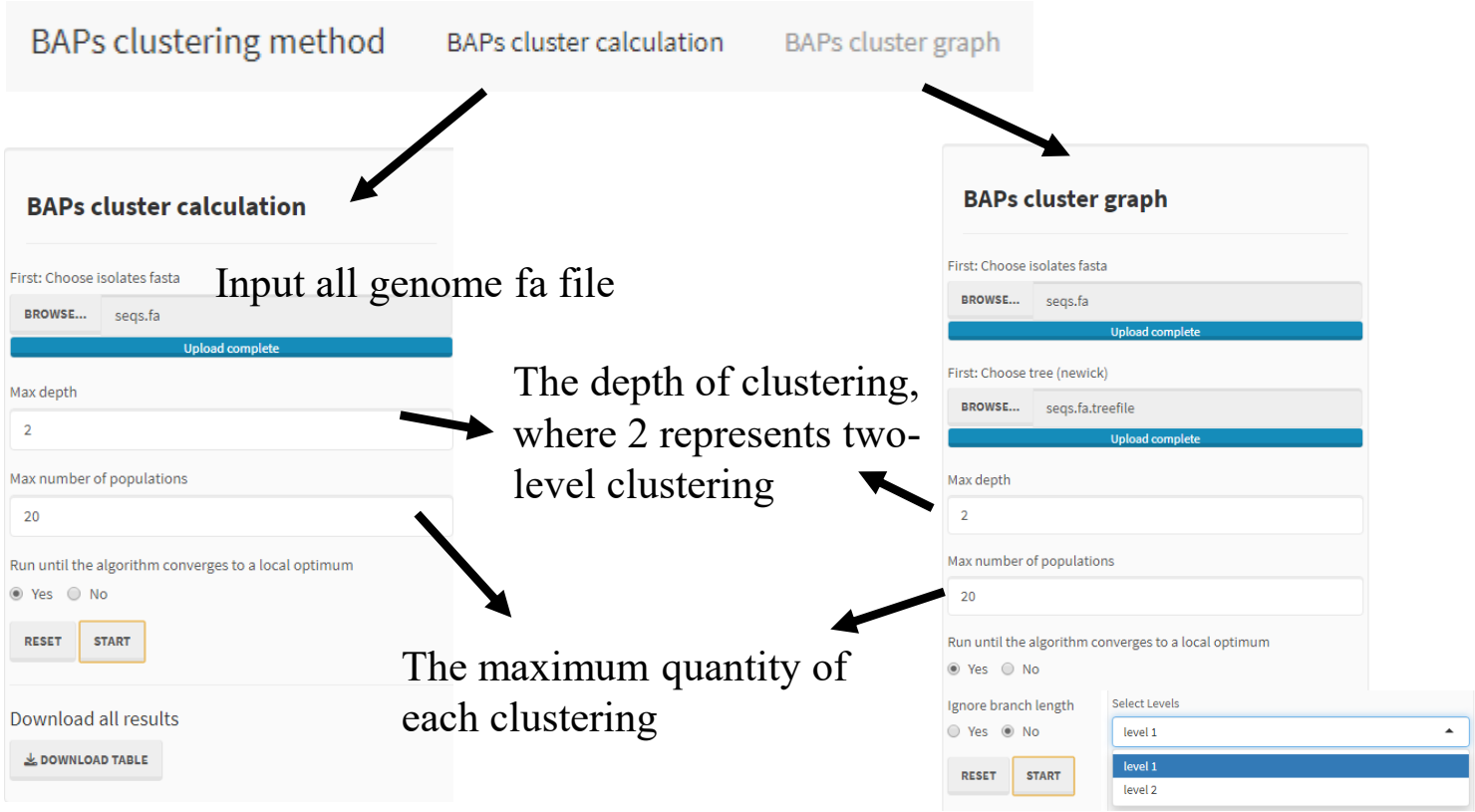


The interpretation of the results, please reference prokka:

https://github.com/tseemann/prokka

Bayesian clustering





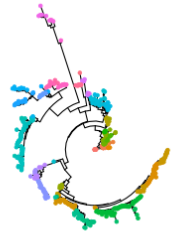
BAPs cluster table

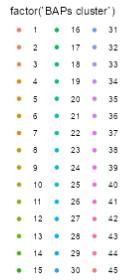
Show 10 entries Search:

	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