

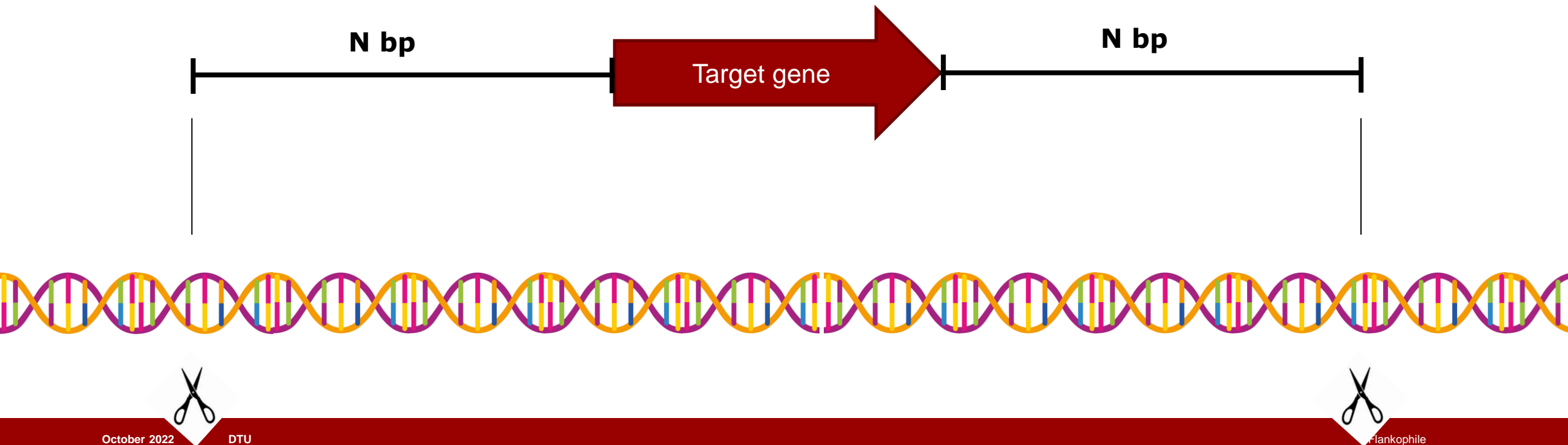
DTU



If you love flanking regions...

Flankophile

Flanking region

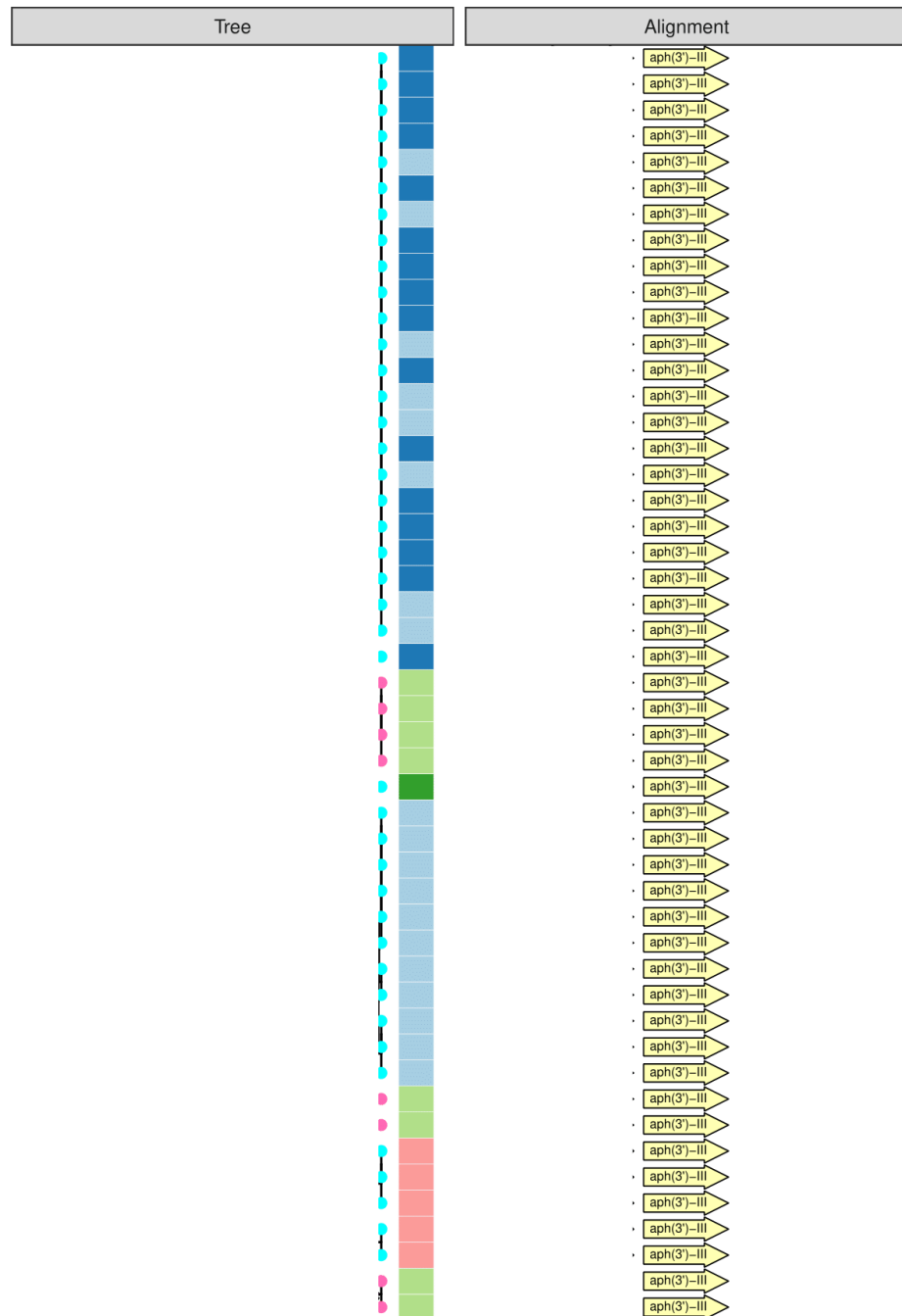


Why do we need flank analysis when studying AMR?

aph(3')-III

Aminoglycoside resistance gene.

Gene sequence is identical.



Host

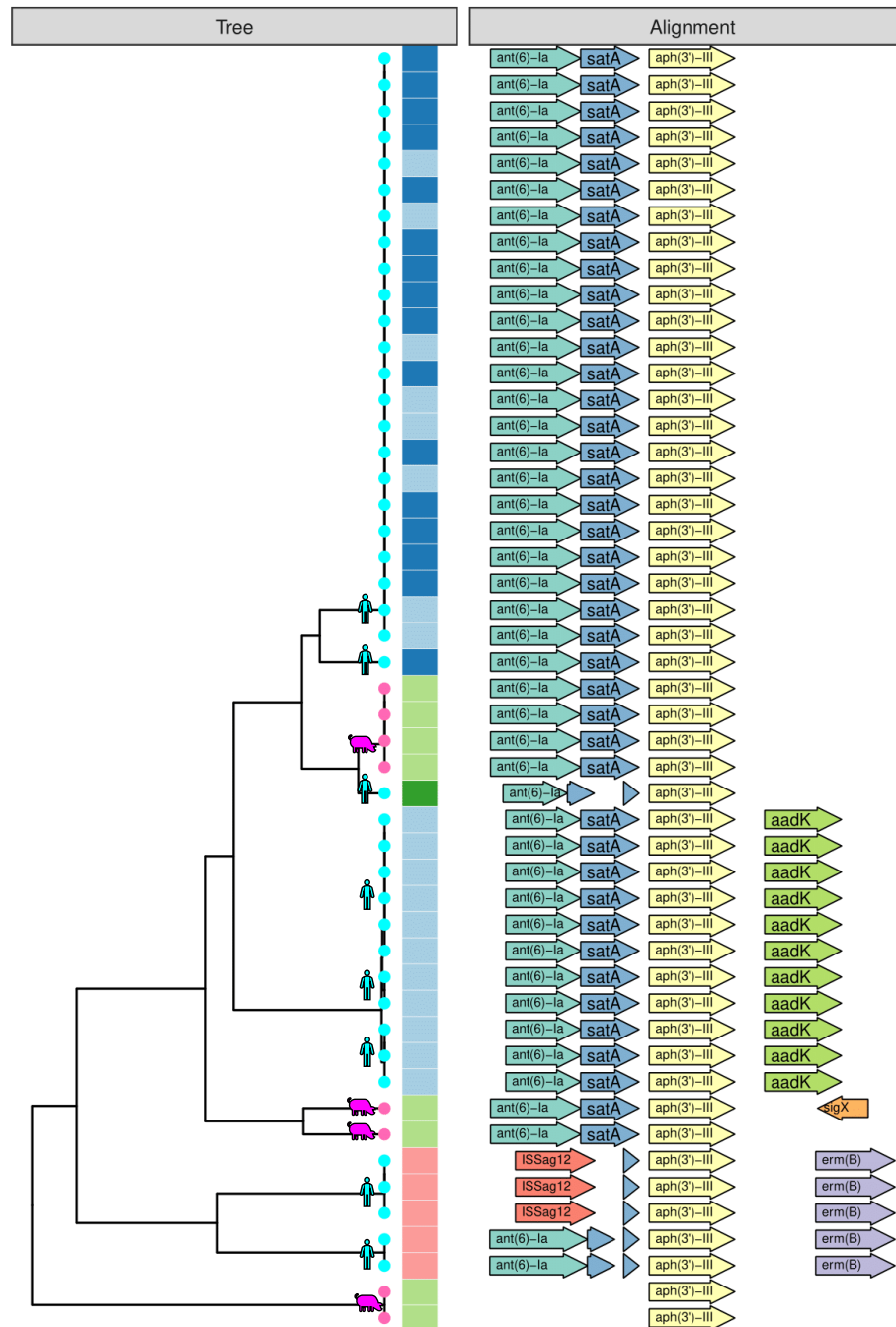
- Pig (Vetforlig II)
- Human (ODID)

Species

- Enterococcus faecalis
- Enterococcus faecium
- Metagenome
- Staphylococcus epidermidis
- Streptococcus agalactiae

aph(3')-III

12 unique flank sequences.
Cluster by host.



Host

- Pig (Vetforlig II)
- Human (ODID)

Species

- Enterococcus faecalis
- Enterococcus faecium
- Metagenome
- Staphylococcus epidermidis
- Streptococcus agalactiae



Patrick Munk



Flankophile



Genomic Epidemiology / CGE

flankophile

Here's where you'll find this repository's source files. To give your users an idea of what they'll find here, [add a description to your repository](#).

master Files Filter files

Name	Size	Last commit	Message
/			
R		3 days ago	new_plots
bin		2022-07-18	v. 0.0.5
example_output		2022-09-21	new logo
input		2022-07-18	edit README
README.md	18.45 KB	4 days ago	plot
Snakefile	31.98 KB	2022-08-15	more suited for bigger datasets
config.yaml	1.69 KB	2022-08-15	more suited for bigger datasets
environment.yaml	350 B	2022-07-18	v. 0.0.5
quick_start.md	3.16 KB	2022-09-21	quick

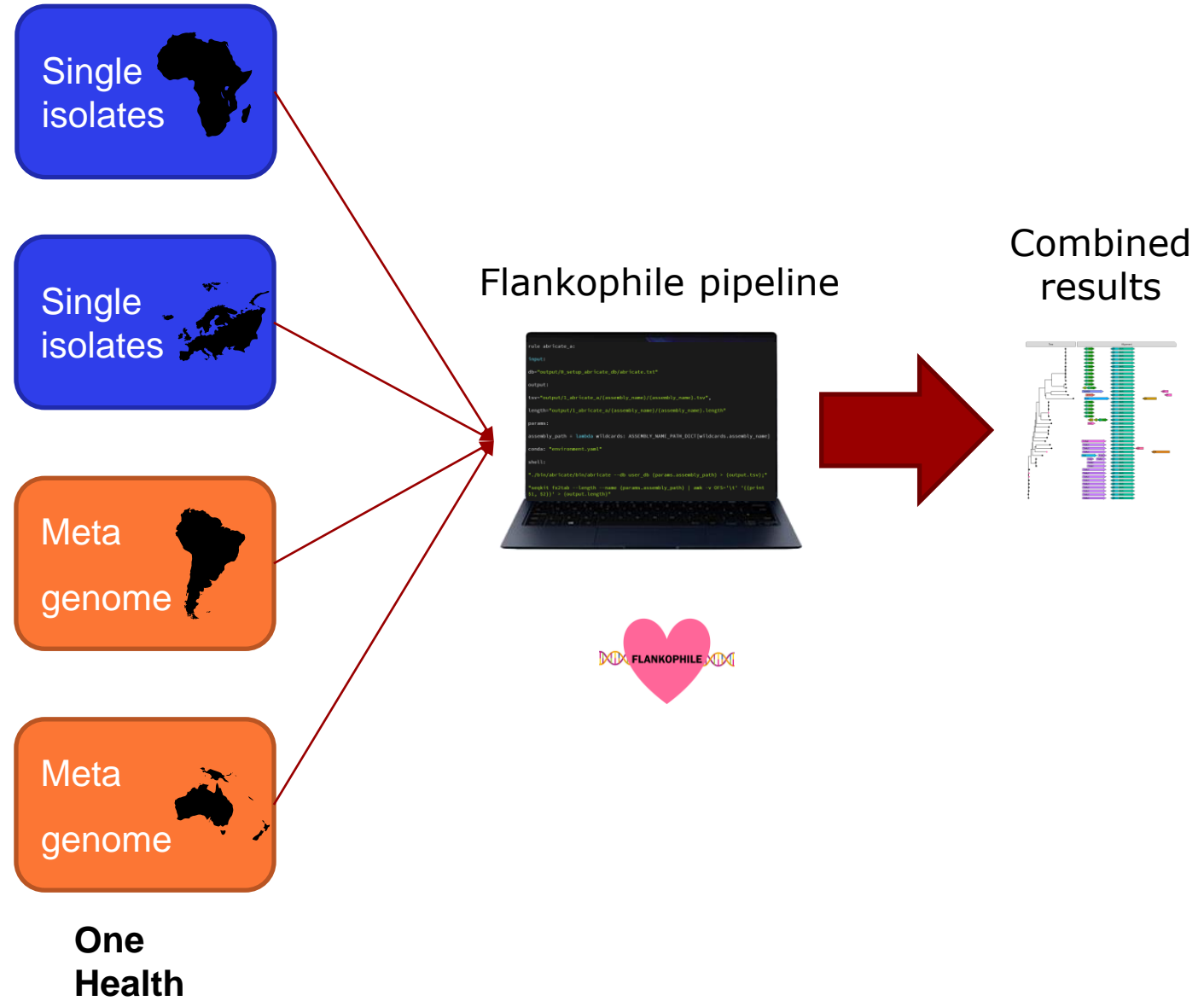
How to run Flankophile

Computerome user? - quick guide for running Flankophile

Input

- Assemblies, genomes or individual contigs in fasta format

- A reference database in fasta format.
For example Resfinder DB



What you get

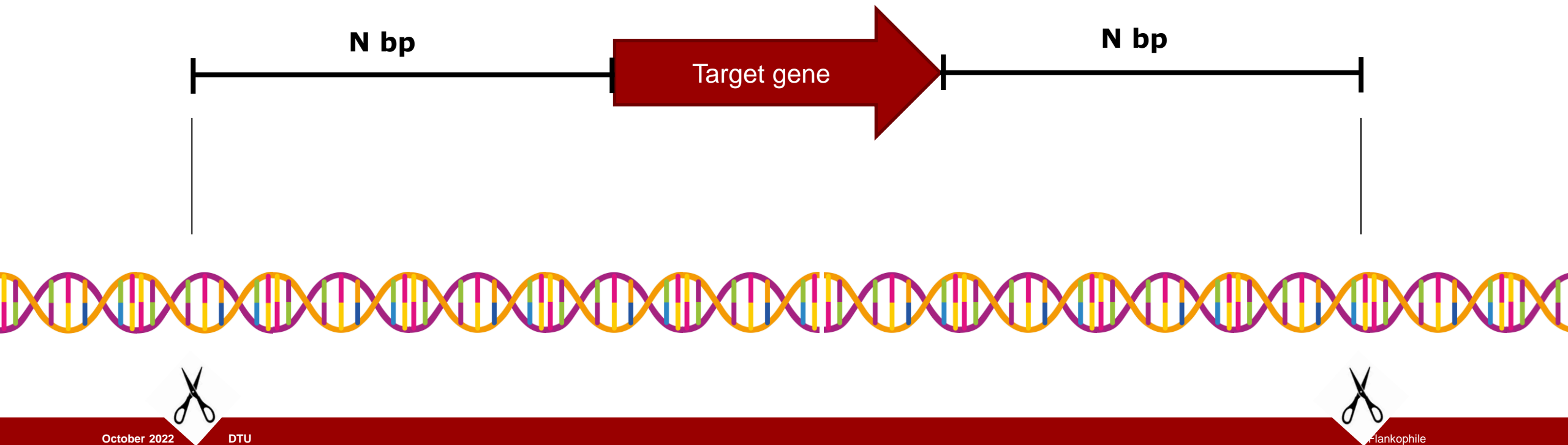
Tsv gene observations

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
#FILE	SEQUENCE	START	END	STRAND	GENE	COVERAGE	COVERAGE	GAPS	%COVERAGE	%IDENTITY	DATABASE	ACCESSION	Assembly	Gene	observation_ID	
2	/home/pr	NODE_199	3692	4687	+	cfxA6_1_c1-1996/996	=====	0/0	100.00	99.60	user_db	cfxA6_1_c	effort_MC11			
3	/home/pr	NODE_281	5121	6605	+	lsa[E]_1_j1-1485/14	=====	0/0	100.00	99.93	user_db	lsa[E]_1_j	effort_MC12			
4	/home/pr	NODE_104	2224	3390	+	tet[X]_2_f1-1167/11	=====	0/0	100.00	99.83	user_db	tet[X]_2	effort_MC13			
5	/home/pr	NODE_104	3603	4403	-	erm[F]_3_1-801/801	=====	0/0	100.00	100.00	user_db	erm[F]_3	effort_MC14			
6	/home/pr	NODE_291	2823	3626	-	lnu(B)_2_1-804/804	=====	0/0	100.00	100.00	user_db	lnu(B)_2	effort_MC15			
7	/home/pr	NODE_291	3680	5164	-	lsa[E]_1_j1-1485/14	=====	0/0	100.00	100.00	user_db	lsa[E]_1_j	effort_MC16			
8	/home/pr	NODE_241	1854	3020	+	tet[X]_2_f1-1167/11	=====	0/0	100.00	99.83	user_db	tet[X]_2_f	effort_MC17			
9	/home/pr	NODE_241	3233	4033	-	erm[F]_3_1-801/801	=====	0/0	100.00	100.00	user_db	erm[F]_3	effort_MC18			
10	/home/pr	NODE_68	1571	2386	+	su12_2_AY1-816/816	=====	0/0	100.00	100.00	user_db	su12_2_A	effort_S1_19			
11	/home/pr	NODE_68	2447	3250	+	aph[3]-it-1-804/804	=====	0/0	100.00	100.00	user_db	aph[3]-it	effort_S1_110			
12	/home/pr	NODE_68	3250	4086	+	aph[6]-id-1-837/837	=====	0/0	100.00	99.88	user_db	aph[6]-id	effort_S1_111			
13	/home/pr	NODE_44	4380	5182	+	aph[3]-it-2-804/804	=====	0/0	99.88	100.00	user_db	aph[3]-it	effort_S1_112			
14	/home/pr	NODE_44	5182	6018	+	aph[6]-id-1-837/837	=====	0/0	100.00	100.00	user_db	aph[6]-id	effort_S1_113			
15	/home/pr	NODE_23	20155	20957	-	aph[3]-it-2-804/804	=====	0/0	99.88	100.00	user_db	aph[3]-it	effort_S1_114			
16	/home/pr	NODE_32	12934	13407	+	dfrA1_10_1-474/474	=====	0/0	100.00	99.79	user_db	dfrA1_10	effort_S1_115			
17	/home/pr	NODE_32	15973	16764	-	su13_2_AJ1-792/792	=====	0/0	100.00	100.00	user_db	su13_2_AJ	effort_S1_116			
18	/home/pr	NODE_66	1988	2803	+	su12_3_HC1-816/816	=====	0/0	100.00	100.00	user_db	su12_3_HC	effort_S1_117			
19	/home/pr	NODE_66	2864	3667	+	aph[3]-it-1-804/804	=====	0/0	100.00	100.00	user_db	aph[3]-it	effort_S1_118			
20	/home/pr	NODE_66	3667	4503	+	aph[6]-id-1-837/837	=====	0/0	100.00	100.00	user_db	aph[6]-id	effort_S1_119			
21	/home/pr	NODE_66	5790	6223	+	dfrA1_10_1-474/474	=====	0/0	100.00	99.79	user_db	dfrA1_10	effort_S1_120			
22	/home/pr	NODE_66	7585	8451	+	su11_5_EU1-867/867	=====	0/0	100.00	99.89	user_db	su11_5_EU	effort_S1_121			
23	/home/pr	NODE_57	5187	6053	-	su11_5_EU1-867/867	=====	0/0	100.00	99.89	user_db	su11_5_EU	effort_S1_122			
24	/home/pr	NODE_57	7415	7888	-	dfrA1_10_1-474/474	=====	0/0	100.00	99.79	user_db	dfrA1_10	effort_S1_123			
25	/home/pr	NODE_57	9135	9971	-	aph[6]-id-1-837/837	=====	0/0	100.00	100.00	user_db	aph[6]-id	effort_S1_124			

Cluster

0_blaSHV_187_1_LN515533	16-08-2022 12:45	File folder
1_blaTEM_57_1_FJ405211	16-08-2022 12:46	File folder
2_blaOXA_486_1_AY597426	16-08-2022 12:47	File folder
3_blaCMY_109_1_AJ746169	16-08-2022 12:45	File folder
4_blaOXY_2_2_2_AJ871867	16-08-2022 12:46	File folder
5_blaLEN25_1_HQ709169	16-08-2022 12:46	File folder
6_cepA_6_FR688022	16-08-2022 12:45	File folder
7_tet_M_4_X75073	16-08-2022 12:47	File folder
8_blaOKP_B_3_1_AM051152	16-08-2022 12:47	File folder
9_blaCTX_M_1_1_DQ915955	16-08-2022 12:48	File folder
10_blaPAO_3_FJ666073	16-08-2022 12:49	File folder
11_tet_O_2_M20925	16-08-2022 12:47	File folder
12_mecA_6_BX571856	16-08-2022 12:46	File folder
13_blaDHA_17_1_KM087850	16-08-2022 12:46	File folder
14_aph_3__la_3_EF015636	16-08-2022 12:45	File folder

Flanking region

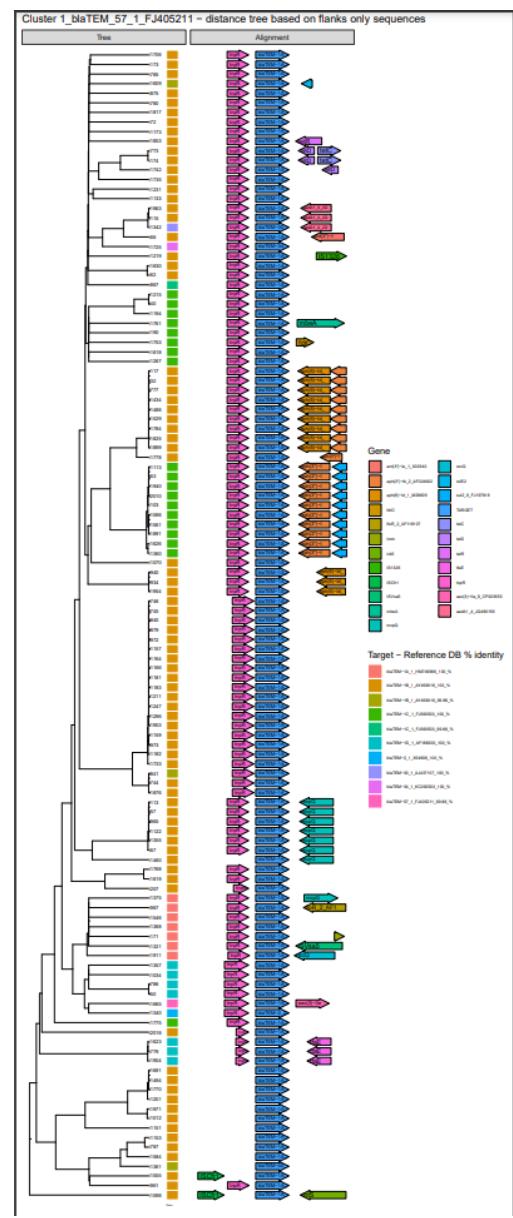


What you get

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
#FILE	SEQUENCE	START	END	STRAND	GENE	COVERAG	COVERAG	GAPS	%COV	%IDENT	DATABAS	ACCESSIO	Assembly	Gene	observation_ID	
2	/home/pr NODE_19	3692	4687	+	cfxA6_1_c1-196/996	=====	0/0		100.00	99.60	user_db	cfxA6_1	effort_Mc11			
3	/home/pr NODE_28	5121	6605	+	lsa(E)_1_j_1-1485/14	=====	0/0		100.00	99.93	user_db	lsa(E)_1_j	effort_Mc12			
4	/home/pr NODE_104	2224	3390	+	tet(X)_2_f1-1167/11	=====	0/0		100.00	99.83	user_db	tet(X)_2	effort_Mc13			
5	/home/pr NODE_104	3603	4403	-	erm(F)_3_1-801/801	=====	0/0		100.00	100.00	user_db	erm(F)_3	effort_Mc14			
6	/home/pr NODE_29	2823	3626	-	lnu(B)_2_1-804/804	=====	0/0		100.00	100.00	user_db	lnu(B)_2	effort_Mc15			
7	/home/pr NODE_29	3680	5164	-	lsa(E)_1_j_1-1485/14	=====	0/0		100.00	100.00	user_db	lsa(E)_1_j	effort_Mc16			
8	/home/pr NODE_241	1854	3020	+	tet(X)_2_f1-1167/11	=====	0/0		100.00	99.83	user_db	tet(X)_2	effort_Mc17			
9	/home/pr NODE_241	3233	4033	-	erm(F)_3_1-801/801	=====	0/0		100.00	100.00	user_db	erm(F)_3	effort_Mc18			
10	/home/pr NODE_68	1571	2386	+	su12_2_AY3-816/816	=====	0/0		100.00	100.00	user_db	su12_2_Ay	effort_St_19			
11	/home/pr NODE_68	2447	3250	+	aph(3)*-Hc-1-804/804	=====	0/0		100.00	100.00	user_db	aph(3)*-Hc	effort_St_110			
12	/home/pr NODE_68	3250	4086	+	aph(6)-id-1-837/837	=====	0/0		100.00	99.88	user_db	aph(6)-id	effort_St_111			
13	/home/pr NODE_44	4380	5182	+	aph(3)*-Hc-2-804/804	=====	0/0		100.00	99.88	100.00	user_db	aph(3)*-Hc	effort_St_112		
14	/home/pr NODE_44	5182	6018	+	aph(6)-id-1-837/837	=====	0/0		100.00	100.00	100.00	user_db	aph(6)-id	effort_St_113		
15	/home/pr NODE_23	20155	20957	-	aph(3)*-Hc-2-804/804	=====	0/0		99.88	100.00	user_db	aph(3)*-Hc	effort_St_114			
16	/home/pr NODE_52	12934	13407	+	dfrA1_10_1-474/474	=====	0/0		100.00	99.79	user_db	dfrA1_10	effort_St_115			
17	/home/pr NODE_52	15973	16764	-	su13_2_AJ1-792/792	=====	0/0		100.00	100.00	user_db	su13_2_AJ	effort_St_116			
18	/home/pr NODE_66	1988	2803	+	su12_3_HC-816/816	=====	0/0		100.00	100.00	user_db	su12_3_HC	effort_St_117			
19	/home/pr NODE_66	2864	3667	+	aph(3)*-Hc-1-804/804	=====	0/0		100.00	100.00	user_db	aph(3)*-Hc	effort_St_118			
20	/home/pr NODE_66	3667	4503	+	aph(6)-id-1-837/837	=====	0/0		100.00	100.00	100.00	user_db	aph(6)-id	effort_St_119		
21	/home/pr NODE_66	5750	6223	+	dfrA1_10_1-474/474	=====	0/0		100.00	99.79	user_db	dfrA1_10	effort_St_120			
22	/home/pr NODE_66	7585	8451	+	su11_5_EU1-867/867	=====	0/0		100.00	99.89	user_db	su11_5_EU	effort_St_121			
23	/home/pr NODE_57	5187	6053	-	su11_5_EU1-867/867	=====	0/0		100.00	99.89	user_db	su11_5_EU	effort_St_122			
24	/home/pr NODE_57	7415	7888	-	dfrA1_10_1-474/474	=====	0/0		100.00	99.79	user_db	dfrA1_10	effort_St_123			
25	/home/pr NODE_57	9135	9971	-	aph(6)-id-1-837/837	=====	0/0		100.00	100.00	user_db	aph(6)-id	effort_St_124			

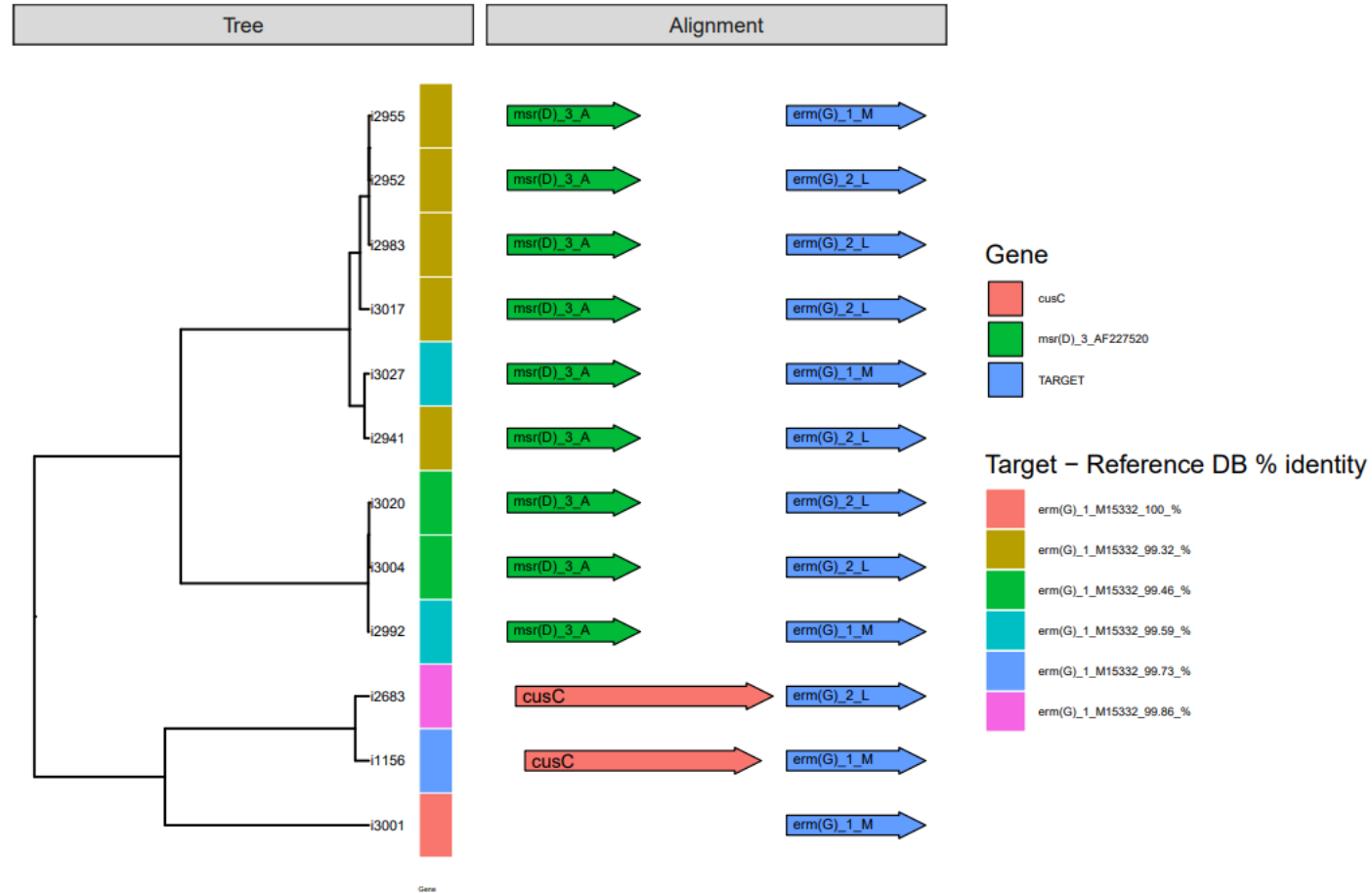
- prokka
- 1_blaTEM_57_1_FJ405211.flanks_with_gene_dist
- 1_blaTEM_57_1_FJ405211.flanks_with_gene_fast
- 1_blaTEM_57_1_FJ405211.flanks_with_gene_tree
- 1_blaTEM_57_1_FJ405211.ggenes
- 1_blaTEM_57_1_FJ405211.just_gene_dist
- 1_blaTEM_57_1_FJ405211.just_gene_fasta
- 1_blaTEM_57_1_FJ405211.just_gene_tree
- 1_blaTEM_57_1_FJ405211.masked_gene_dist
- 1_blaTEM_57_1_FJ405211.masked_gene_fasta
- 1_blaTEM_57_1_FJ405211.masked_gene_tree
- 1_blaTEM_57_1_FJ405211.tsv

- 0_blaSHV_187_1_LN515533
- 1_blaTEM_57_1_FJ405211
- 2_blaOXA_486_1_AY597426
- 3_blaCMY_109_1_AJ746169
- 4_blaOXY_2_2_2_AJ871867
- 5_blaLEN25_1_HQ709169
- 6_cepA_6_FR688022
- 7_tet_M_4_X75073
- 8_blaOKP_8_3_1_AM051152
- 9_blaCTX_M_1_1_DQ915955
- 10_blaPAO_3_FJ666073
- 11_tet_O_2_M20925
- 12_mecA_6_BX571856
- 13_blaDHA_17_1_KM087850
- 14_aph_3_la_3_EF015636



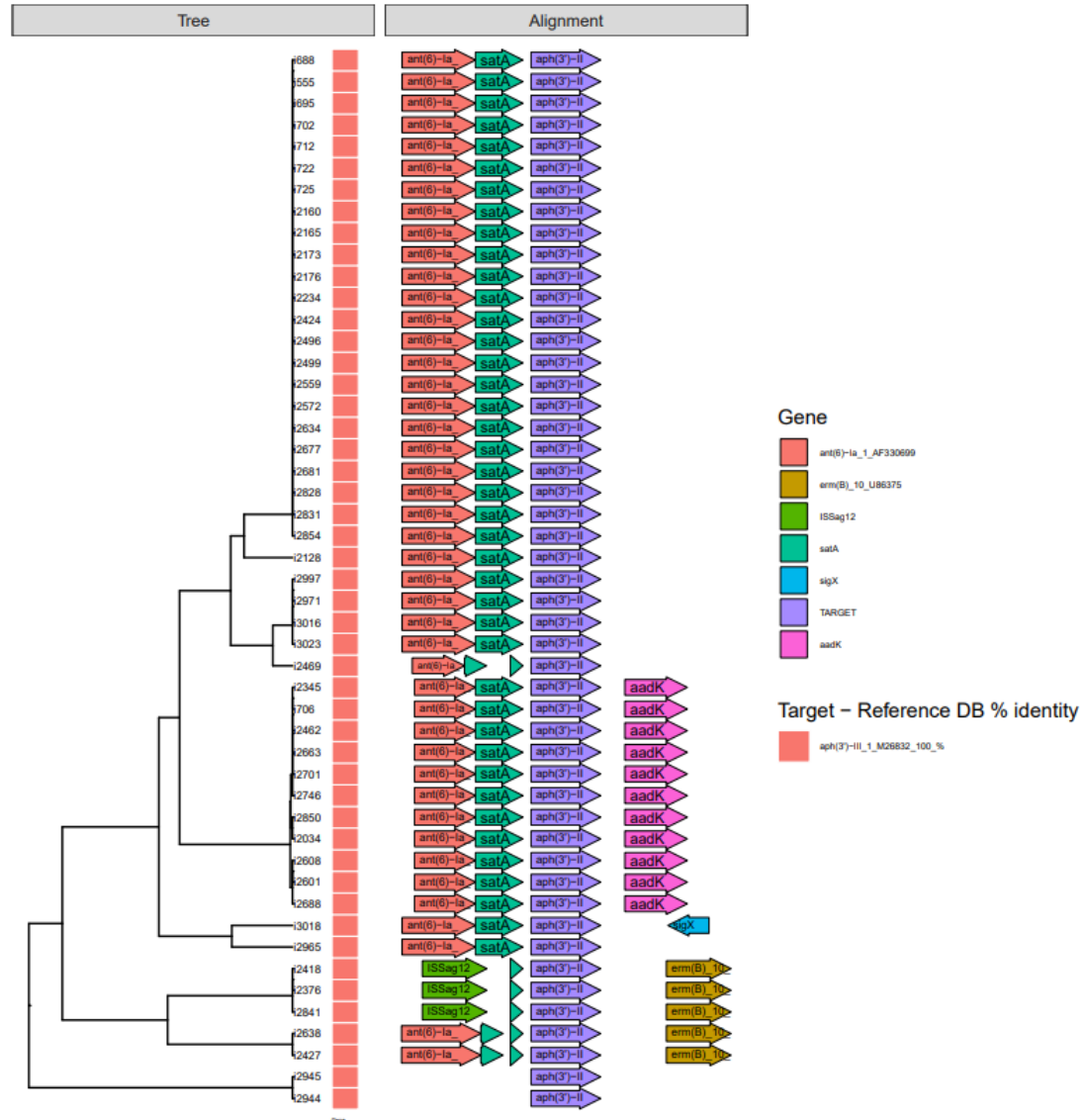
Ready made plot from Flankophile

Cluster 45_erm_G__1_M15332 – distance tree based on flanks only sequences



Cluster 53_aph_3__III_1_M26832 – distance tree based on flanks only sequences

Gene annotation



Gene annotation with

1. Your reference database
2. Uniprot
3. Isfinder Database

Config file – no flags

```
## 0 ##
database: "input/example_input_files/Resfinder_08_02_2022_dub_rem.fa" # a multifasta file

## 1 ##
#input_format: "contigs"
input_format: "assemblies" # "assemblies" for assembly list. "contigs" for contigs list.

input_list: "input/example_input_files/test_input_1000_assemblies_list.tsv" # Path to input_list.

## 2 ##
flank_length_upstreams: "2000"
flank_length_downstreams: "2000"
min_coverage_abricate: "95" # percentage
min_identity_abricate: "95" # percentage

## 4 ##
cluster_identity_cd_hit: "0.95" # -c
cluster_wordsize_cd_hit: "9" # -n
cluster_length_dif_cd_hit: "0.9" # -s

## 5 ##
Kmersize_kma: "16" # Kmersize used by kma index. Try with "16" if in doubt.

distance_measure: "1" # Choose measurement for making distance matrix

# Distance calculation methods:
# 1 k-mer hamming distance
# 64 Jaccard distance
# 256 Cosine distance
# 4096 Chi-square distance
```

Logos



I am looking for

- Feedback on documentation
- Suggestions for paper

Take home message

- I would love to help you get started with Flankophile

Thank you