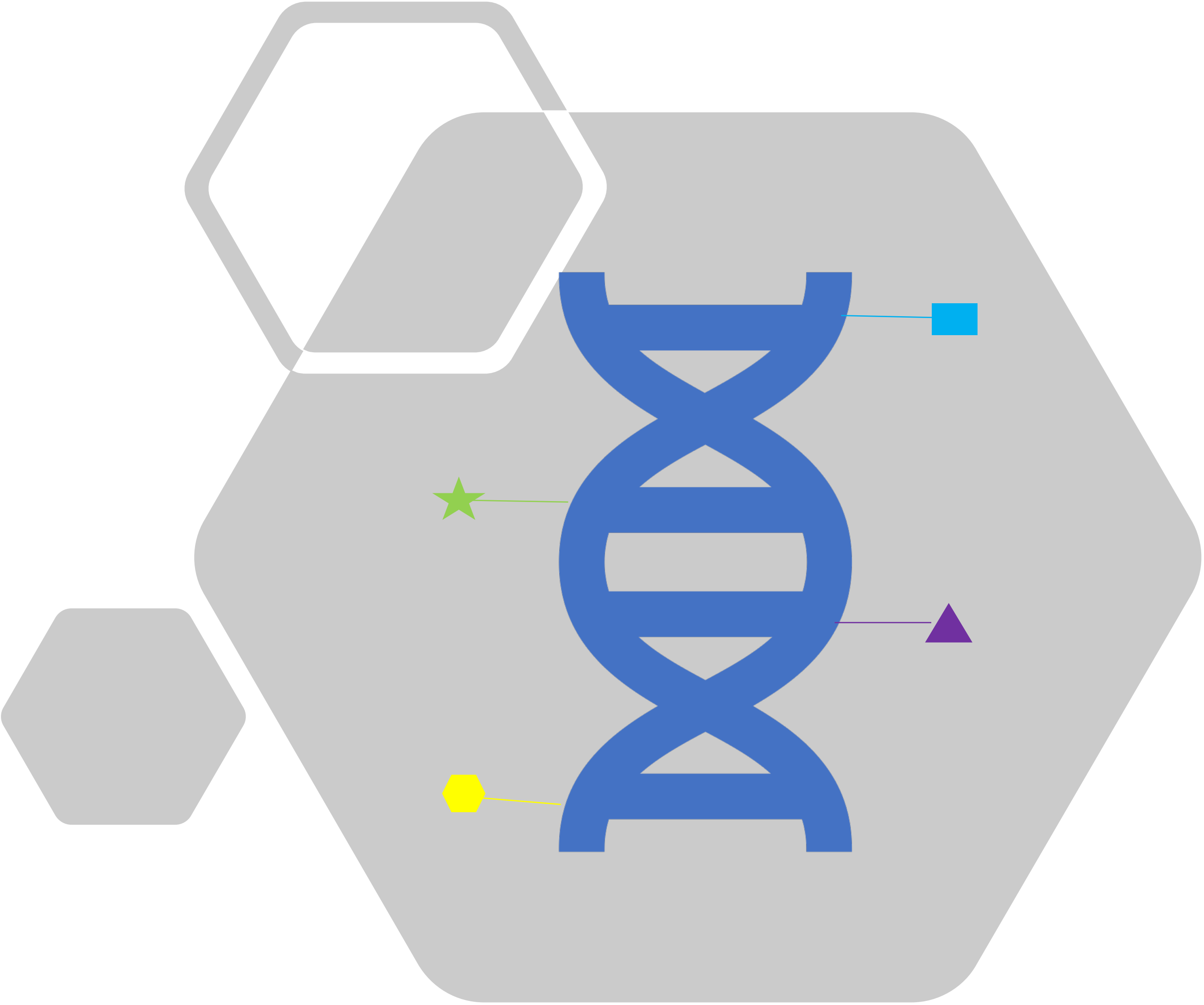
Functional Annotation Final Results

Team 2

Danielle Temples, Courtney Astore, Ujani Hazra, Sooyoun Oh, Rhiya Sharma

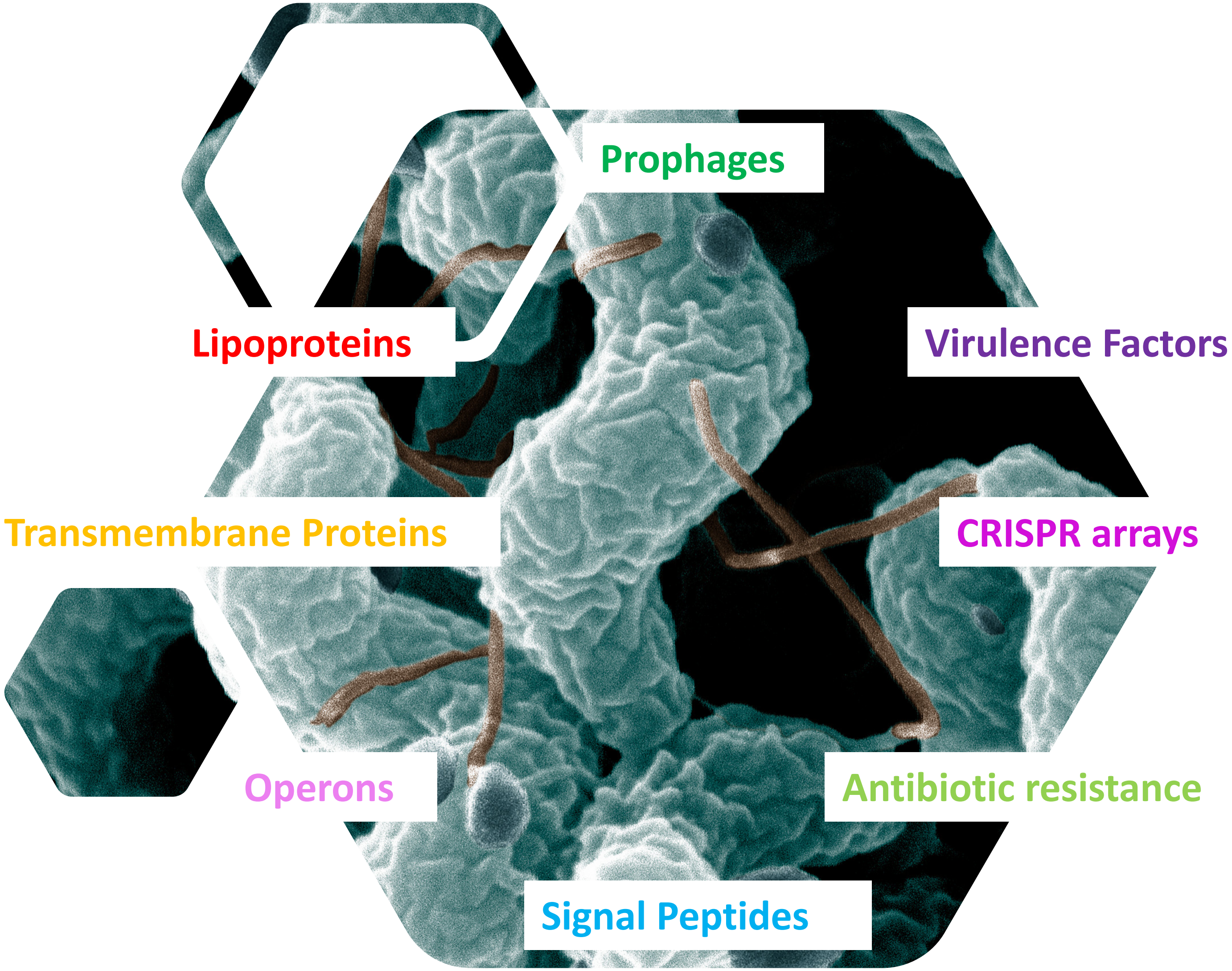
Outline

- *C. jejuni* & functional annotation background
- Overview and objective
- Final Pipeline
- Homology vs. *ab initio* techniques
- Results
  - Clustering
  - Homology
  - *Ab initio*
- Concluding remarks
- Deliverables



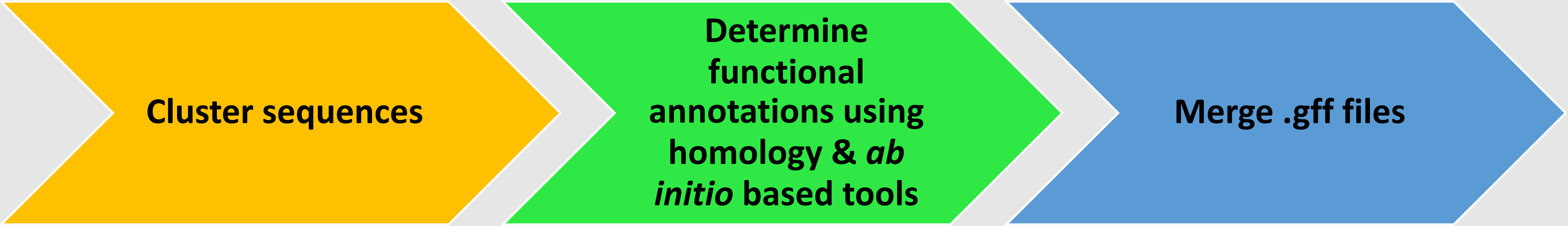
Campylobacter jejuni

- Campylobacter species are the major cause of human bacterial gastroenteritis
- *C. jejuni* and *C. coli* together account for over 95% of Campylobacter infections in humans
- Certain strains are linked with the development of the neurological disorder Guillain-Barre syndrome (GBS)
- Gram negative bacteria
- Outer membrane is highly related to antibiotic resistance



Overview

- Perform a full functional annotation on the genes and proteins determined by the Gene Prediction group that is relevant to *C. jejuni*
- General approach:



Cluster sequences

Determine functional annotations using homology & *ab initio* based tools

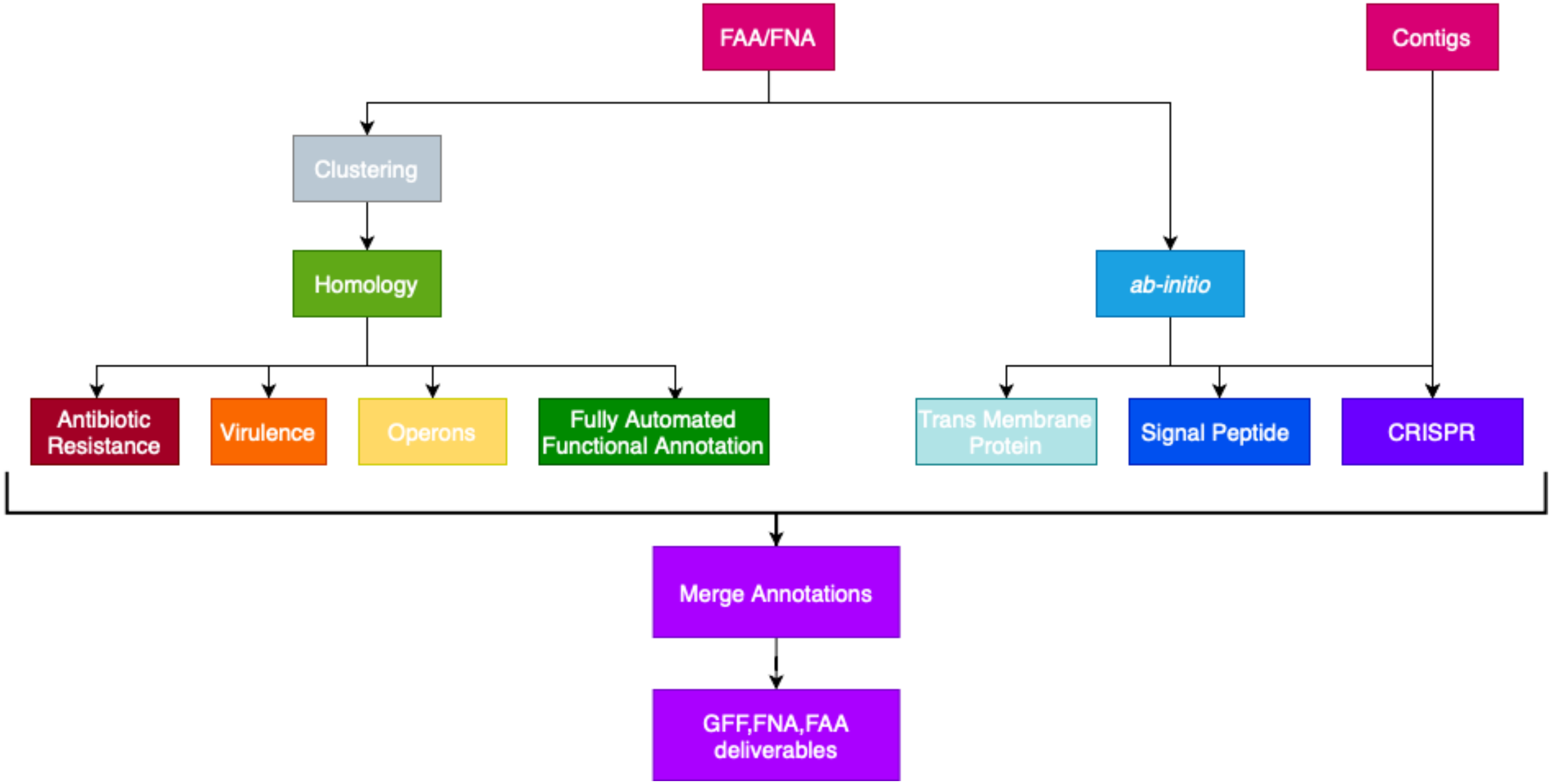
Merge .gff files

- Deliverables:
    - Merged .gff files for each genome
    - FASTA file header annotations

Final Pipeline

Software/database Legend

- CD-HIT
- CARD
- VFDB
- MicrobesOnline
- eggNog-mapper
- TMHMM2
- SignalP v5
- PilerCR



FAA/FNA

Contigs

Clustering

Homology

ab-initio

Antibiotic Resistance

Virulence

Operons

Fully Automated Functional Annotation

Trans Membrane Protein

Signal Peptide

CRISPR

Merge Annotations

GFF,FNA,FAA deliverables

Homology vs. *ab initio*-Based Techniques

Homology

- Determine function via sequence similarity to already functionally annotated sequences
- This is limited by what we already know

ab-initio

- Determine function via predictive model without comparing to existing sequences
- This is based on laws of nature
- Difficult to verify without experiments

CD-HIT: Clustering sequences

- **Command:**
  - `./cd-hit -i <input_file> -o <output_file_name>`
- **Runtime:** 26.29 + 5.36 = 31.65 seconds
- **Basis for choosing this software:**
  - Widely used program for clustering both gene and protein sequences
  - Fast runtime
  - Able to process and handle large databases

Clustering Results

Nucleotide sequences

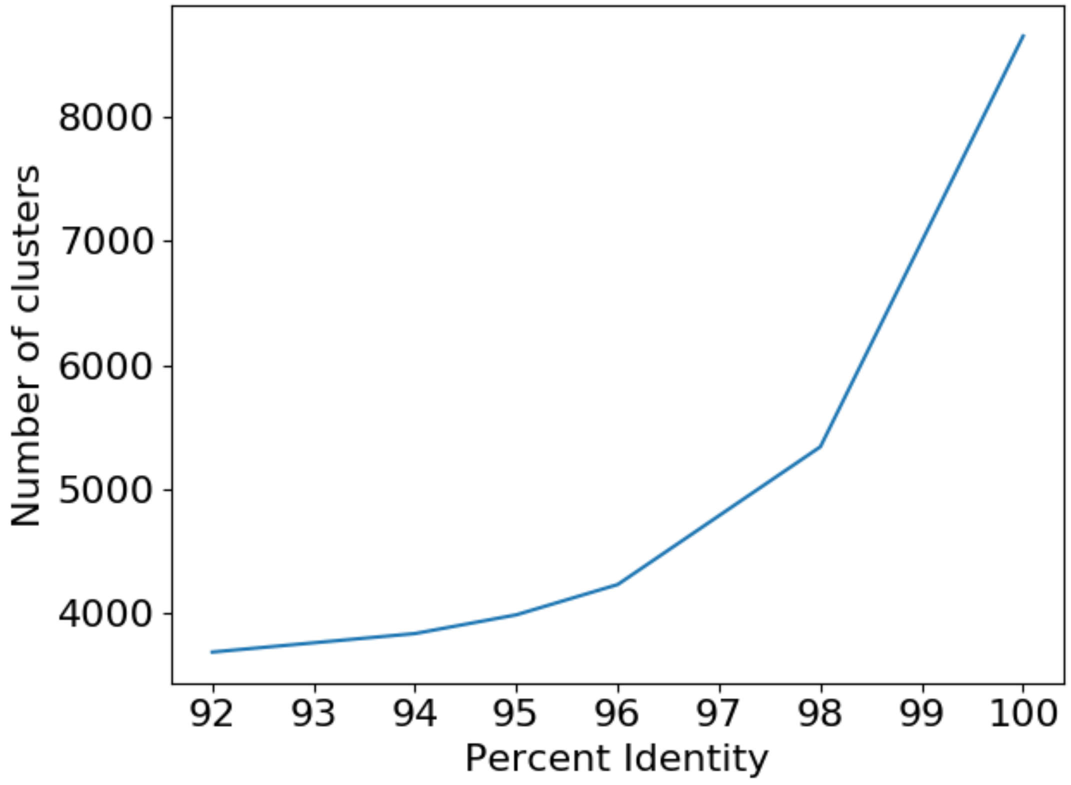
Percent Identity (%)	Number of clusters	Runtime
92	3,689	27.45
94	3,838	28.15
95	**3,989**	**26.29**
96	4,233	25.93
98	5,343	28.66
100	8,652	25.22

Protein sequences

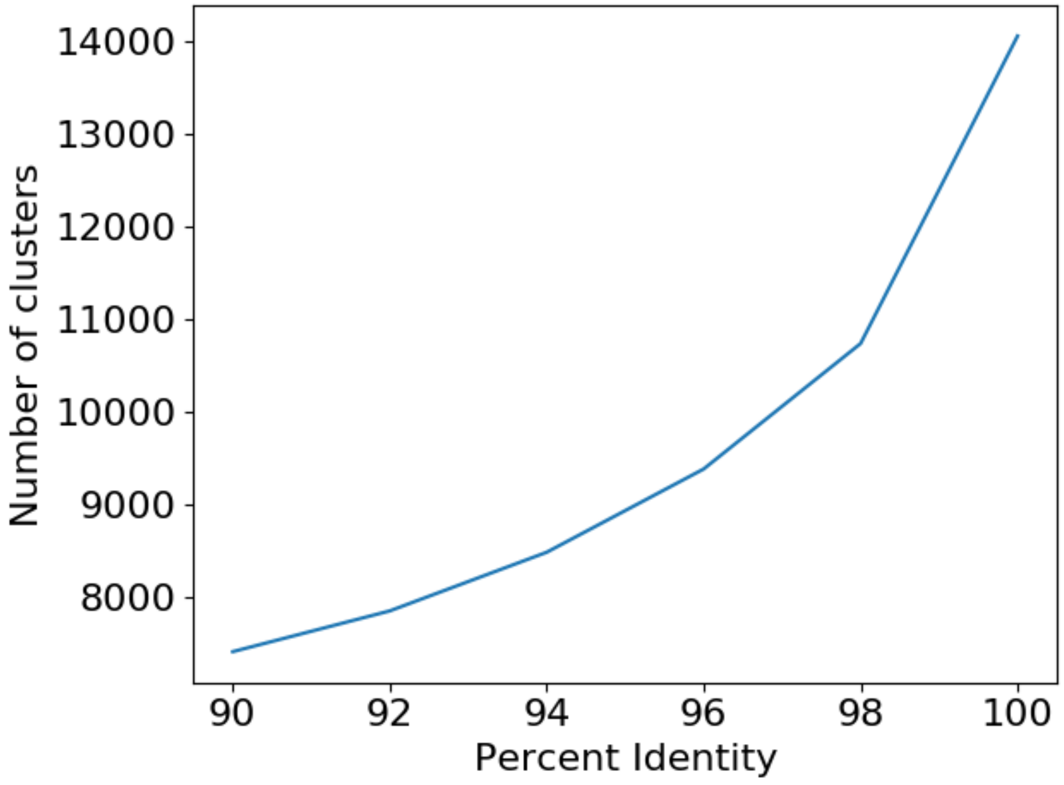
Percent Identity (%)	Number of clusters	Runtime
90	**7,414**	**5.36**
92	7,853	5.25
94	8,485	5.17
96	9,382	5.09
98	10,734	4.58
100	14,048	3.83

Clustering Results

Nucleotide sequences



Protein sequences



N/A: Plasmids

- **Basis for picking no software**
  - PlasmidSeeker:
    - Originally our choice tool for plasmids
    - No hits for individual fna files, one hit for the combined fna files (concatenated version of all fna files)
  - PlasmidFinder:
    - Originally our validation tool for plasmids
    - No hits for the combined fna file

Homology Categories

Prophage:

- Play an important role in the evolution of bacterial genomes and their pathogenicity
- Can change or knock out gene functions; alter gene expression

Virulence:

- A pathogen's ability to infect or damage a host
- Ex: toxins, surface coats that inhibit phagocytosis, surface receptors that bind to host cells

Fully Automated Functional Annotation Tools:

- Tools that annotate a spectrum of features related to the function

Antibiotic Resistance:

- When bacteria develop the ability to defeat the drugs designed to kill them
- Leads to higher medical costs, prolonged hospital stays, and increased mortality

Operons:

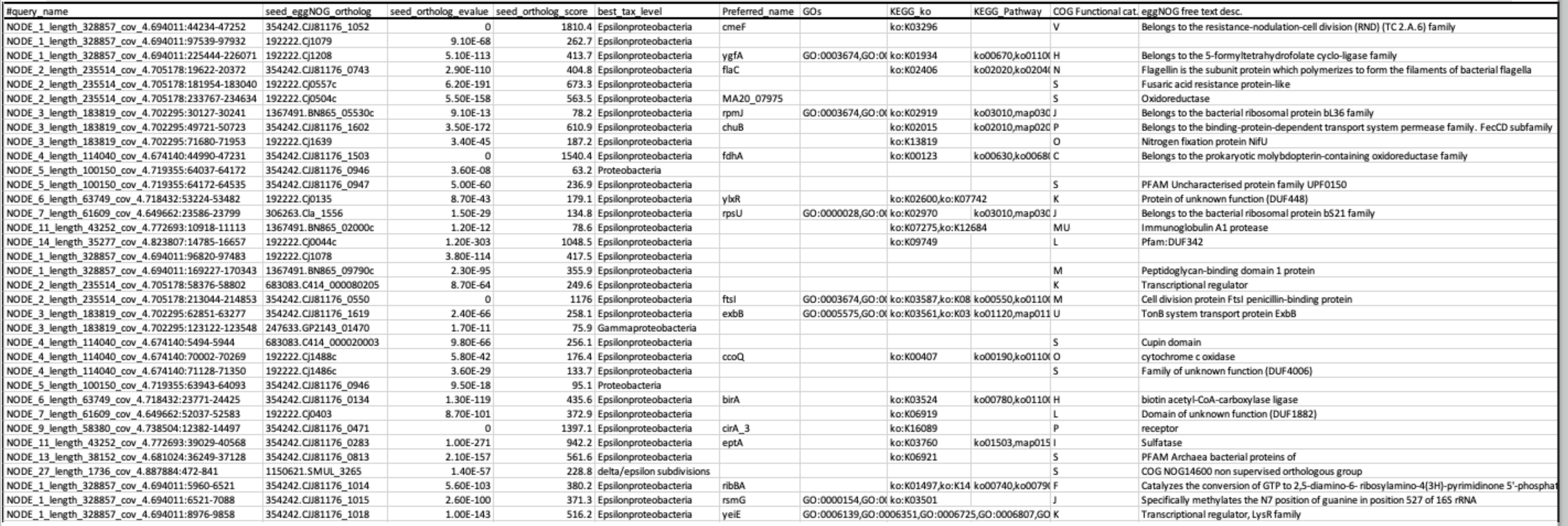
- A functional unit of transcription and genetic regulation
- Identifying these may enhance our knowledge of gene regulation & function which is a key addition to genome annotation

eggNog-mapper: Fully Automated Functional Annotation

- **Commands:**

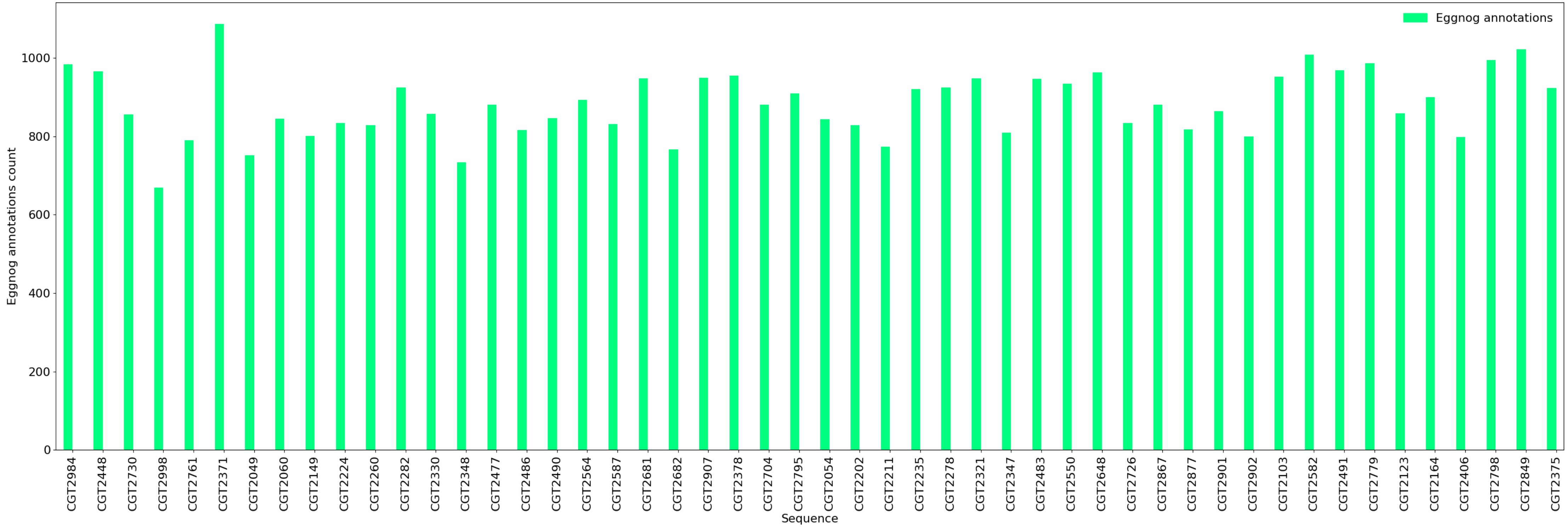
  -
    ```
    ./emapper.py  -i <cluster> --output <result> -d bact -m diamond
    ```
- **Basis for choosing this software:**
  - PANNZER2 could not generate database (errors in script)
  - 7% lower false positive rate and 15x faster than BLAST
  - 5,090 representative genomes (4445 bacteria)
  - Provides important annotations such as *seed eggNOG ortholog, seed ortholog evalue/score, predicted taxonomic group, predicted protein name, GO terms, KEGG ko, KEGG Pathways, eggnog OG, COG Functional Category, free text descriptions*

eggNog-mapper Output



#query_name	seed_eggNOG_ortholog	seed_ortholog_evalue	seed_ortholog_score	best_tax_level	Preferred_name	GOs	KEGG_ko	KEGG_Pathway	COG Functional cat.	eggNOG free text desc.
NODE_1_length_328857_cov_4.694011:44234-47252	354242.CJJ81176_1052	0	1810.4	Epsilonproteobacteria	cmeF		ko:K03296		V	Belongs to the resistance-nodulation-cell division (RND) (TC 2.A.6) family
NODE_1_length_328857_cov_4.694011:97539-97932	192222.Cj1079	9.10E-68	262.7	Epsilonproteobacteria						
NODE_1_length_328857_cov_4.694011:225444-226071	192222.Cj1208	5.10E-113	413.7	Epsilonproteobacteria	ygfA	GO:0003674,GO:0	ko:K01934	ko00670,ko0110	H	Belongs to the 5-formyltetrahydrofolate cyclo-ligase family
NODE_2_length_235514_cov_4.705178:19622-20372	354242.CJJ81176_0743	2.90E-110	404.8	Epsilonproteobacteria	flaC		ko:K02406	ko02020,ko0204	N	Flagellin is the subunit protein which polymerizes to form the filaments of bacterial flagella
NODE_2_length_235514_cov_4.705178:181954-183040	192222.Cj0557c	6.20E-191	673.3	Epsilonproteobacteria					S	Fusaric acid resistance protein-like
NODE_2_length_235514_cov_4.705178:233767-234634	192222.Cj0504c	5.50E-158	563.5	Epsilonproteobacteria	MA20_07975				S	Oxidoreductase
NODE_3_length_183819_cov_4.702295:30127-30241	1367491.BN865_05530c	9.10E-13	78.2	Epsilonproteobacteria	rpmJ	GO:0003674,GO:0	ko:K02919	ko03010,map030	J	Belongs to the bacterial ribosomal protein bL36 family
NODE_3_length_183819_cov_4.702295:49721-50723	354242.CJJ81176_1602	3.50E-172	610.9	Epsilonproteobacteria	chuB		ko:K02015	ko02010,map020	P	Belongs to the binding-protein-dependent transport system permease family. FecCD subfamily
NODE_3_length_183819_cov_4.702295:71680-71953	192222.Cj1639	3.40E-45	187.2	Epsilonproteobacteria			ko:K13819		O	Nitrogen fixation protein NifU
NODE_4_length_114040_cov_4.674140:44990-47231	354242.CJJ81176_1503	0	1540.4	Epsilonproteobacteria	fdhA		ko:K00123	ko00630,ko0068	C	Belongs to the prokaryotic molybdopterin-containing oxidoreductase family
NODE_5_length_100150_cov_4.719355:64037-64172	354242.CJJ81176_0946	3.60E-08	63.2	Proteobacteria						
NODE_5_length_100150_cov_4.719355:64172-64535	354242.CJJ81176_0947	5.00E-60	236.9	Epsilonproteobacteria					S	PFAM Uncharacterised protein family UPF0150
NODE_6_length_63749_cov_4.718432:53224-53482	192222.Cj0135	8.70E-43	179.1	Epsilonproteobacteria	ylxR		ko:K02600,ko:K07742		K	Protein of unknown function (DUF448)
NODE_7_length_61609_cov_4.649662:23586-23799	306263.Cla_1556	1.50E-29	134.8	Epsilonproteobacteria	rpsU	GO:0000028,GO:0	ko:K02970	ko03010,map030	J	Belongs to the bacterial ribosomal protein bS21 family
NODE_11_length_43252_cov_4.772693:10918-11113	1367491.BN865_02000c	1.20E-12	78.6	Epsilonproteobacteria			ko:K07275,ko:K12684		MU	Immunoglobulin A1 protease
NODE_14_length_35277_cov_4.823807:14785-16657	192222.Cj0044c	1.20E-303	1048.5	Epsilonproteobacteria			ko:K09749		L	Pfam:DUF342
NODE_1_length_328857_cov_4.694011:96820-97483	192222.Cj1078	3.80E-114	417.5	Epsilonproteobacteria						
NODE_1_length_328857_cov_4.694011:169227-170343	1367491.BN865_09790c	2.30E-95	355.9	Epsilonproteobacteria					M	Peptidoglycan-binding domain 1 protein
NODE_2_length_235514_cov_4.705178:58376-58802	683083.C414_000080205	8.70E-64	249.6	Epsilonproteobacteria					K	Transcriptional regulator
NODE_2_length_235514_cov_4.705178:213044-214853	354242.CJJ81176_0550	0	1176	Epsilonproteobacteria	ftsI	GO:0003674,GO:0	ko:K03587,ko:K08	ko00550,ko0110	M	Cell division protein FtsI penicillin-binding protein
NODE_3_length_183819_cov_4.702295:62851-63277	354242.CJJ81176_1619	2.40E-66	258.1	Epsilonproteobacteria	exbB	GO:0005575,GO:0	ko:K03561,ko:K03	ko01120,map011	U	TonB system transport protein ExbB
NODE_3_length_183819_cov_4.702295:123122-123548	247633.GP2143_01470	1.70E-11	75.9	Gammaproteobacteria						
NODE_4_length_114040_cov_4.674140:5494-5944	683083.C414_000020003	9.80E-66	256.1	Epsilonproteobacteria					S	Cupin domain
NODE_4_length_114040_cov_4.674140:70002-70269	192222.Cj1488c	5.80E-42	176.4	Epsilonproteobacteria	ccoQ		ko:K00407	ko00190,ko0110	O	cytochrome c oxidase
NODE_4_length_114040_cov_4.674140:71128-71350	192222.Cj1486c	3.60E-29	133.7	Epsilonproteobacteria					S	Family of unknown function (DUF4006)
NODE_5_length_100150_cov_4.719355:63943-64093	354242.CJJ81176_0946	9.50E-18	95.1	Proteobacteria						
NODE_6_length_63749_cov_4.718432:23771-24425	354242.CJJ81176_0134	1.30E-119	435.6	Epsilonproteobacteria	birA		ko:K03524	ko00780,ko0110	H	biotin acetyl-CoA-carboxylase ligase
NODE_7_length_61609_cov_4.649662:52037-52583	192222.Cj0403	8.70E-101	372.9	Epsilonproteobacteria			ko:K06919		L	Domain of unknown function (DUF1882)
NODE_9_length_58380_cov_4.738504:12382-14497	354242.CJJ81176_0471	0	1397.1	Epsilonproteobacteria	cirA_3		ko:K16089		P	receptor
NODE_11_length_43252_cov_4.772693:39029-40568	354242.CJJ81176_0283	1.00E-271	942.2	Epsilonproteobacteria	eptA		ko:K03760	ko01503,map015	I	Sulfatase
NODE_13_length_38152_cov_4.681024:36249-37128	354242.CJJ81176_0813	2.10E-157	561.6	Epsilonproteobacteria			ko:K06921		S	PFAM Archaea bacterial proteins of
NODE_27_length_1736_cov_4.887884:472-841	1150621.SMUL_3265	1.40E-57	228.8	delta/epsilon subdivisions					S	COG NOG14600 non supervised orthologous group
NODE_1_length_328857_cov_4.694011:5960-6521	354242.CJJ81176_1014	5.60E-103	380.2	Epsilonproteobacteria	ribBA		ko:K01497,ko:K14	ko00740,ko0079	F	Catalyzes the conversion of GTP to 2,5-diamino-6- ribosylamino-4(3H)-pyrimidinone 5'-phospha
NODE_1_length_328857_cov_4.694011:6521-7088	354242.CJJ81176_1015	2.60E-100	371.3	Epsilonproteobacteria	rsmG	GO:0000154,GO:0	ko:K03501		J	Specifically methylates the N7 position of guanine in position 527 of 16S rRNA
NODE_1_length_328857_cov_4.694011:8976-9858	354242.CJJ81176_1018	1.00E-143	516.2	Epsilonproteobacteria	yeiE	GO:0006139,GO:0006351,GO:0006725,GO:0006807,GO			K	Transcriptional regulator, LysR family

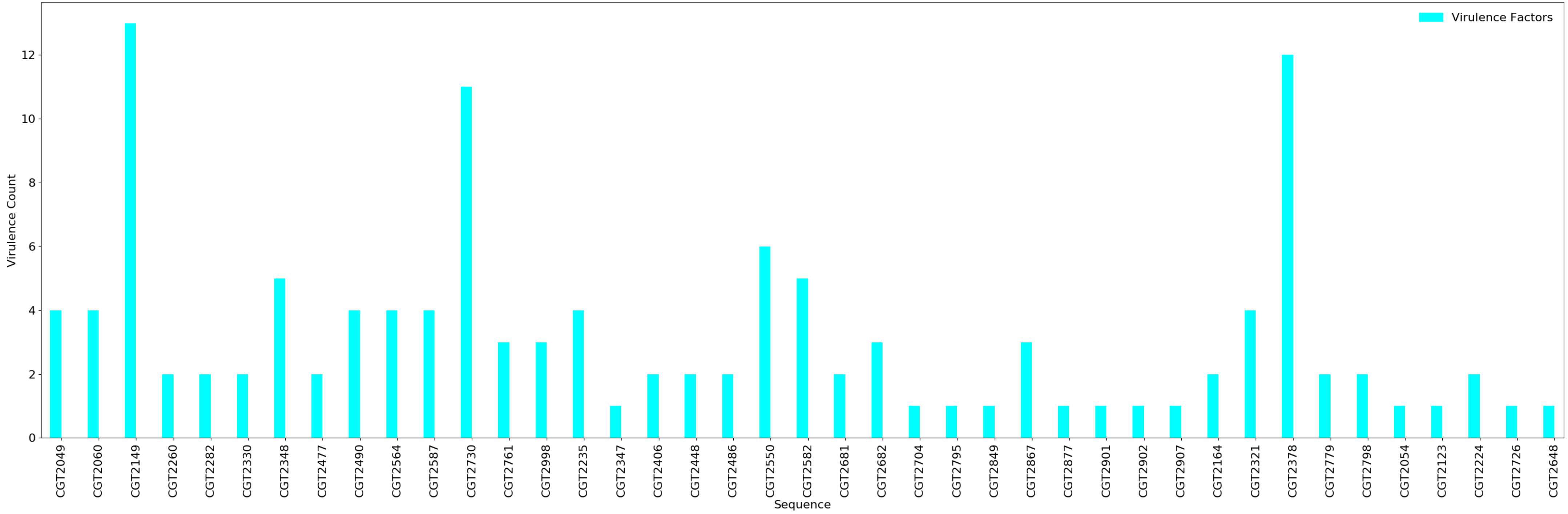
eggNog-mapper Results



VFDB: Virulence Factors

- **Commands:**
  ```
  makeblastdb -in VFDB_db -dbtype 'nucl' -out <db_name>
  ```
  ```
  blastn -db <db_name> -query <cluster> -out <result> -max_hsps 1 -max_target_seqs 1 -outfmt 6 -perc_identity 100 -num_threads 5
  ```
- **Runtime:** < 5 minutes
- **Basis for choosing this software:**
  - Very well known/reliant source for virulence factors
  - “broadest and most comprehensive” and “had the highest quality with its curated dataset and virulence-guided classification system”

VFDB Results



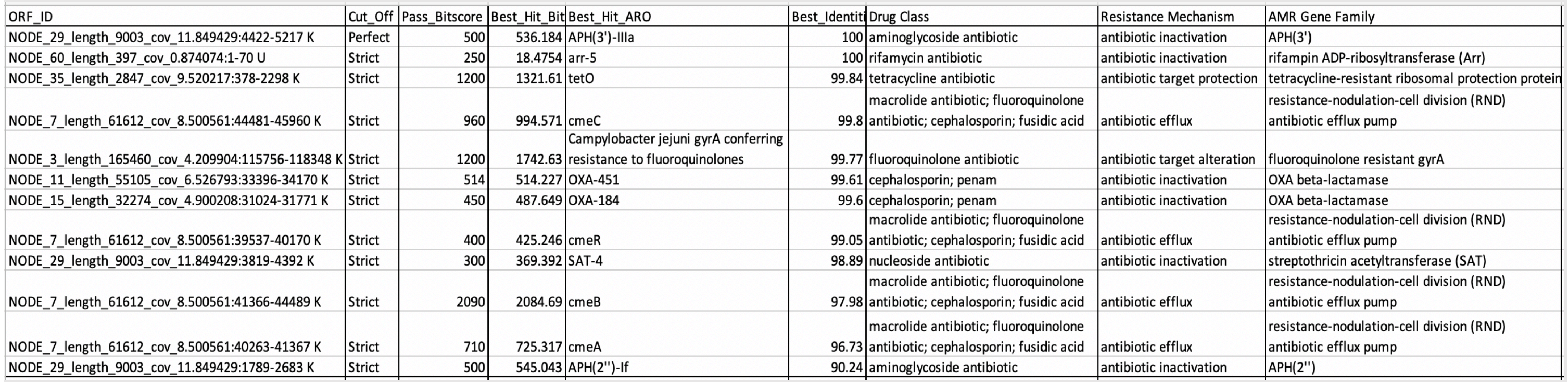
CARD: Antibiotic Resistance

- **Commands:**

```
wget https://card.mcmaster.ca/latest/data
tar -xvf data ./card.json
rgi load --card_json <path to card.json> --local
rgi main -i <path to cluster.faa> -o <output_file_name> -t protein –local
rgi tab -i <path to output_file_name.json>
```

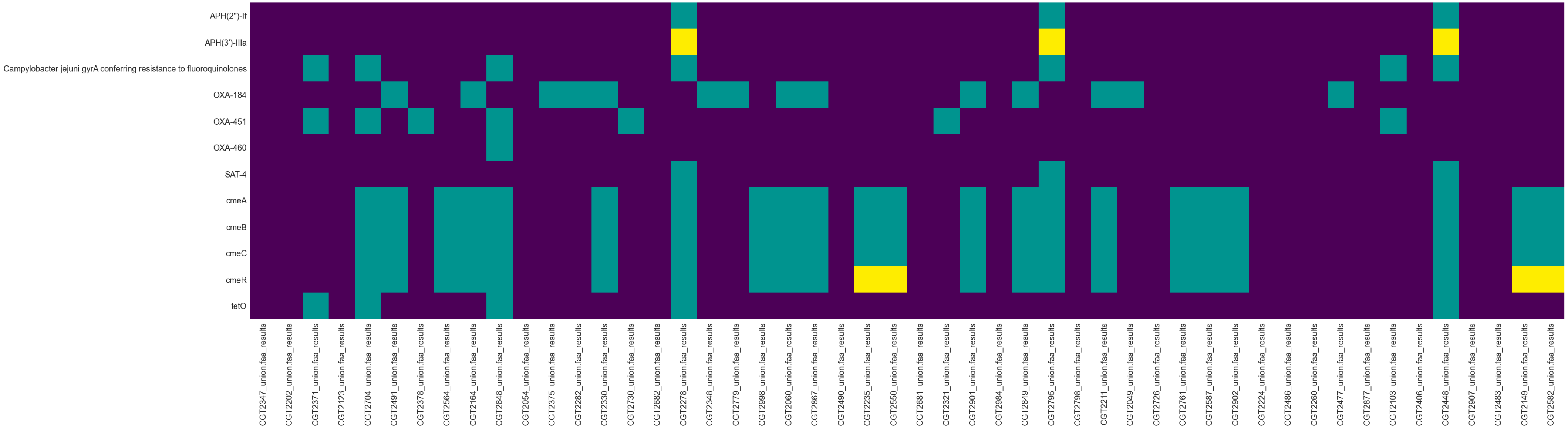
- **Runtime:** 1 minute
- **Basis for choosing this software:** Produced results while ResFinder did not produce any results.

CARD: Output



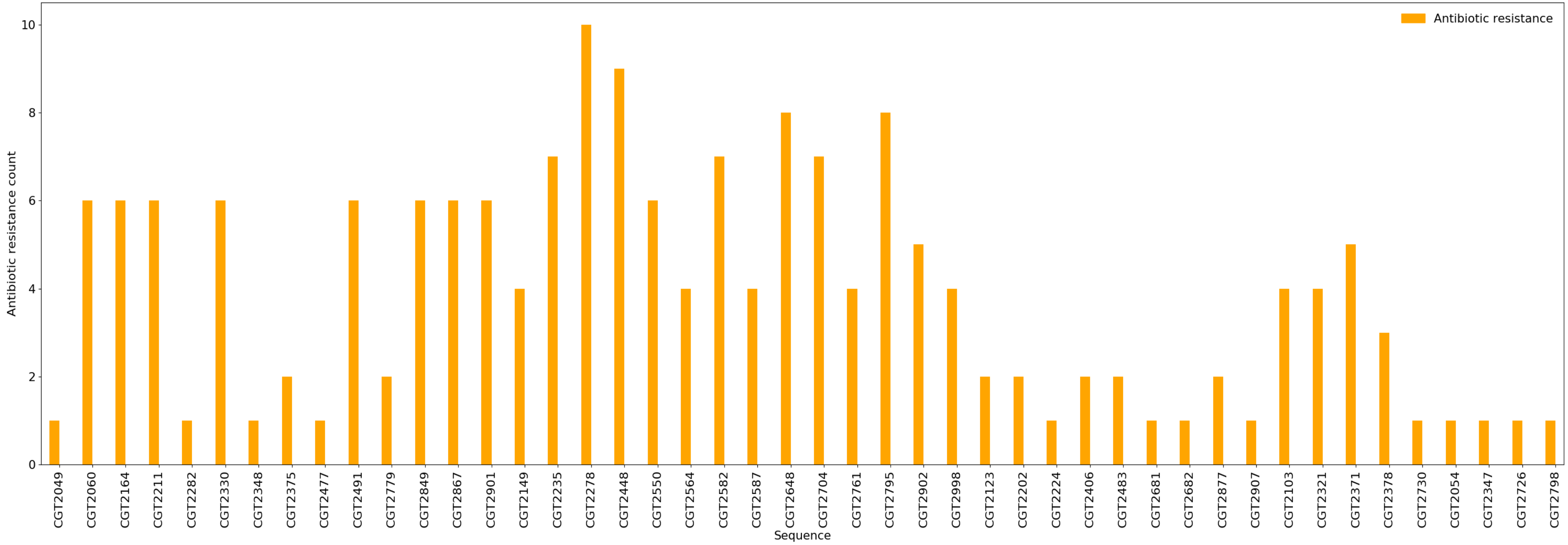
ORF_ID	Cut_Off	Pass_Bitscore	Best_Hit_Bit	Best_Hit_ARO	Best_Identiti	Drug Class	Resistance Mechanism	AMR Gene Family
NODE_29_length_9003_cov_11.849429:4422-5217 K	Perfect	500	536.184	APH(3')-IIIa	100	aminoglycoside antibiotic	antibiotic inactivation	APH(3')
NODE_60_length_397_cov_0.874074:1-70 U	Strict	250	18.4754	arr-5	100	rifamycin antibiotic	antibiotic inactivation	rifampin ADP-ribosyltransferase (Arr)
NODE_35_length_2847_cov_9.520217:378-2298 K	Strict	1200	1321.61	tetO	99.84	tetracycline antibiotic	antibiotic target protection	tetracycline-resistant ribosomal protection protein
NODE_7_length_61612_cov_8.500561:44481-45960 K	Strict	960	994.571	cmeC	99.8	macrolide antibiotic; fluoroquinolone antibiotic; cephalosporin; fusidic acid	antibiotic efflux	resistance-nodulation-cell division (RND) antibiotic efflux pump
NODE_3_length_165460_cov_4.209904:115756-118348 K	Strict	1200	1742.63	Campylobacter jejuni gyrA conferring resistance to fluoroquinolones	99.77	fluoroquinolone antibiotic	antibiotic target alteration	fluoroquinolone resistant gyrA
NODE_11_length_55105_cov_6.526793:33396-34170 K	Strict	514	514.227	OXA-451	99.61	cephalosporin; penam	antibiotic inactivation	OXA beta-lactamase
NODE_15_length_32274_cov_4.900208:31024-31771 K	Strict	450	487.649	OXA-184	99.6	cephalosporin; penam	antibiotic inactivation	OXA beta-lactamase
NODE_7_length_61612_cov_8.500561:39537-40170 K	Strict	400	425.246	cmeR	99.05	macrolide antibiotic; fluoroquinolone antibiotic; cephalosporin; fusidic acid	antibiotic efflux	resistance-nodulation-cell division (RND) antibiotic efflux pump
NODE_29_length_9003_cov_11.849429:3819-4392 K	Strict	300	369.392	SAT-4	98.89	nucleoside antibiotic	antibiotic inactivation	streptothricin acetyltransferase (SAT)
NODE_7_length_61612_cov_8.500561:41366-44489 K	Strict	2090	2084.69	cmeB	97.98	macrolide antibiotic; fluoroquinolone antibiotic; cephalosporin; fusidic acid	antibiotic efflux	resistance-nodulation-cell division (RND) antibiotic efflux pump
NODE_7_length_61612_cov_8.500561:40263-41367 K	Strict	710	725.317	cmeA	96.73	macrolide antibiotic; fluoroquinolone antibiotic; cephalosporin; fusidic acid	antibiotic efflux	resistance-nodulation-cell division (RND) antibiotic efflux pump
NODE_29_length_9003_cov_11.849429:1789-2683 K	Strict	500	545.043	APH(2'')-If	90.24	aminoglycoside antibiotic	antibiotic inactivation	APH(2'')

CARD: Heatmap



- Heatmap generated using the command: rgi heatmap –i <CARD output file directory>
- CARD annotations from unclustered protein samples.
- Yellow represents a perfect hit, teal represents a strict hit, purple represents no hit.
- Perfect Algorithm: Detects perfect matches to the curated reference sequences and mutations in the CARD
- Strict Algorithm: Detects previously unknown variants of known AMR genes using detection models with CARD's curated similarity cut-offs to detect variant is likely a functional AMR gene.

CARD Results



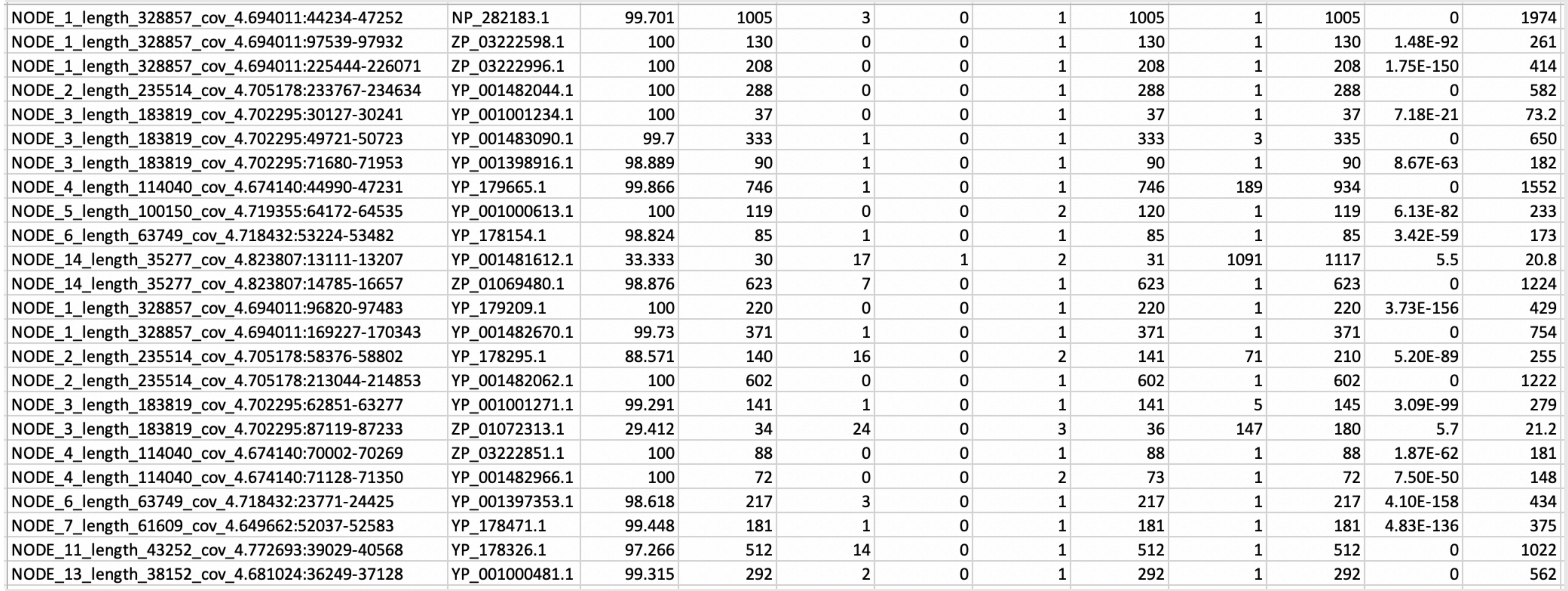
MicrobesOnline: Operons

- Downloaded all the operon tables for Campylobacter jejuni.
- Use GID to retrieve the protein reference sequences.
- Created a database using BLAST.
- Query the clustered sequences with the with the reference protein database.
- **Commands:**

  -
    ```
    makeblastdb -in <fasta file > -dbtype prot -out <database>
    ```

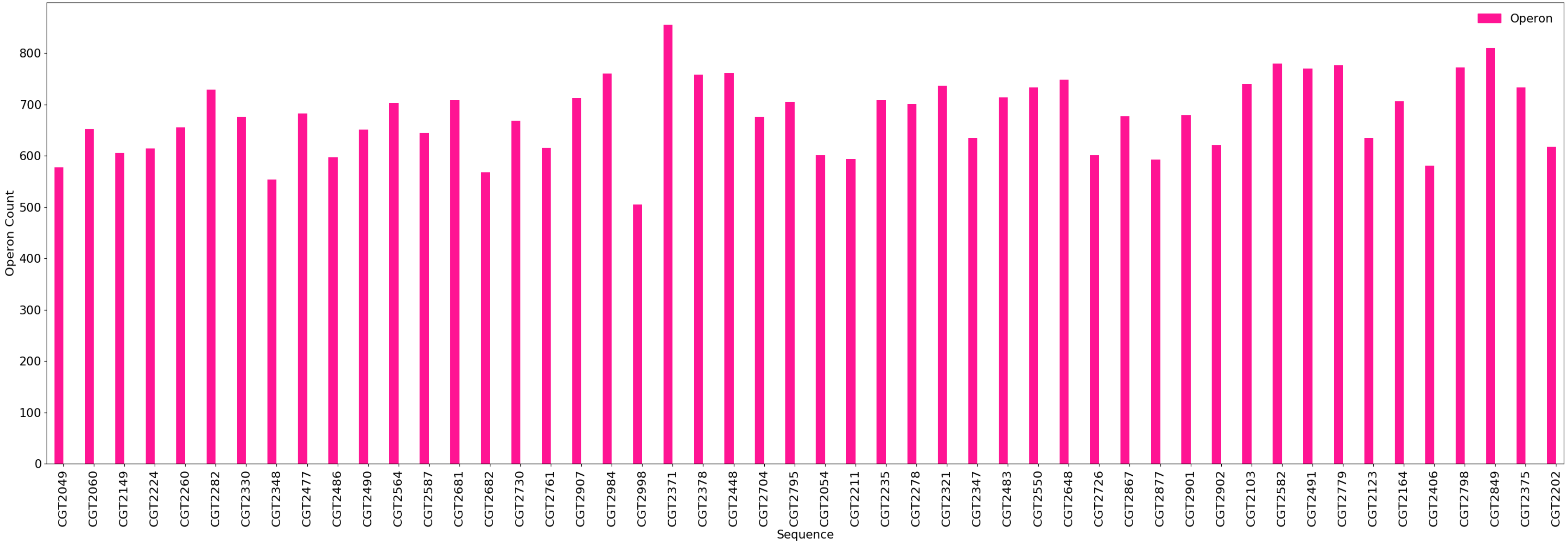
  -
    ```
    blastp -query cdhit/faa_rep_seq.faa -db tmp/db_operon -evalue 0.01 -max_target_seqs 1 -max_hsps 1 -outfmt 6 -out tmp/hits_0.01.txt -num_threads 5
    ```
- **Runtime:** 30 seconds

MicrobesOnline: Output



NODE_1_length_328857_cov_4.694011:44234-47252	NP_282183.1	99.701	1005	3	0	1	1005	1	1005	0	1974
NODE_1_length_328857_cov_4.694011:97539-97932	ZP_03222598.1	100	130	0	0	1	130	1	130	1.48E-92	261
NODE_1_length_328857_cov_4.694011:225444-226071	ZP_03222996.1	100	208	0	0	1	208	1	208	1.75E-150	414
NODE_2_length_235514_cov_4.705178:233767-234634	YP_001482044.1	100	288	0	0	1	288	1	288	0	582
NODE_3_length_183819_cov_4.702295:30127-30241	YP_001001234.1	100	37	0	0	1	37	1	37	7.18E-21	73.2
NODE_3_length_183819_cov_4.702295:49721-50723	YP_001483090.1	99.7	333	1	0	1	333	3	335	0	650
NODE_3_length_183819_cov_4.702295:71680-71953	YP_001398916.1	98.889	90	1	0	1	90	1	90	8.67E-63	182
NODE_4_length_114040_cov_4.674140:44990-47231	YP_179665.1	99.866	746	1	0	1	746	189	934	0	1552
NODE_5_length_100150_cov_4.719355:64172-64535	YP_001000613.1	100	119	0	0	2	120	1	119	6.13E-82	233
NODE_6_length_63749_cov_4.718432:53224-53482	YP_178154.1	98.824	85	1	0	1	85	1	85	3.42E-59	173
NODE_14_length_35277_cov_4.823807:13111-13207	YP_001481612.1	33.333	30	17	1	2	31	1091	1117	5.5	20.8
NODE_14_length_35277_cov_4.823807:14785-16657	ZP_01069480.1	98.876	623	7	0	1	623	1	623	0	1224
NODE_1_length_328857_cov_4.694011:96820-97483	YP_179209.1	100	220	0	0	1	220	1	220	3.73E-156	429
NODE_1_length_328857_cov_4.694011:169227-170343	YP_001482670.1	99.73	371	1	0	1	371	1	371	0	754
NODE_2_length_235514_cov_4.705178:58376-58802	YP_178295.1	88.571	140	16	0	2	141	71	210	5.20E-89	255
NODE_2_length_235514_cov_4.705178:213044-214853	YP_001482062.1	100	602	0	0	1	602	1	602	0	1222
NODE_3_length_183819_cov_4.702295:62851-63277	YP_001001271.1	99.291	141	1	0	1	141	5	145	3.09E-99	279
NODE_3_length_183819_cov_4.702295:87119-87233	ZP_01072313.1	29.412	34	24	0	3	36	147	180	5.7	21.2
NODE_4_length_114040_cov_4.674140:70002-70269	ZP_03222851.1	100	88	0	0	1	88	1	88	1.87E-62	181
NODE_4_length_114040_cov_4.674140:71128-71350	YP_001482966.1	100	72	0	0	2	73	1	72	7.50E-50	148
NODE_6_length_63749_cov_4.718432:23771-24425	YP_001397353.1	98.618	217	3	0	1	217	1	217	4.10E-158	434
NODE_7_length_61609_cov_4.649662:52037-52583	YP_178471.1	99.448	181	1	0	1	181	1	181	4.83E-136	375
NODE_11_length_43252_cov_4.772693:39029-40568	YP_178326.1	97.266	512	14	0	1	512	1	512	0	1022
NODE_13_length_38152_cov_4.681024:36249-37128	YP_001000481.1	99.315	292	2	0	1	292	1	292	0	562

Operon Results



Homology Results Overview

Functional Annotation	Software / Database	Min	Max	Median	Mean
Fully Automated Functional Annotation	eggnog-mapper	669	1087	880	881.9
Virulence Factors	VFDB	1	13	2	3.1
Antibiotic Resistance	CARD	1	10	4	3.8
Operons	MicrobesOnline	505	856	678	678.1

ab-initio Categories

Transmembrane Proteins :

- Bacteria have the ability to export effector proteins in membranes of eukaryotic host
- Integral membrane protein that function as gates or docking sites that allow or prevent the entry or exit of materials across the cell membrane

Signal Peptides:

- Guide secretory proteins to find their correct locations outside the cell membrane for signal transduction

CRISPR:

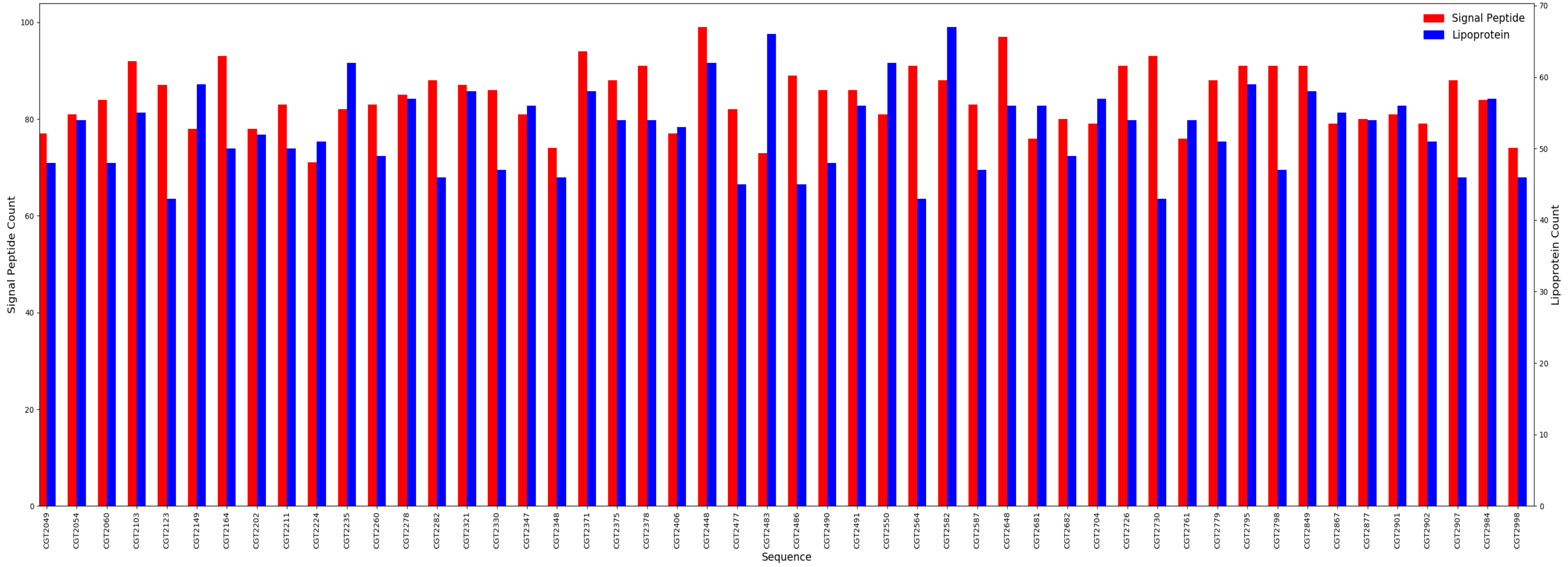
- Provides immunity to the bacteria against Bacteriophages
- Contributes to the Virulence and Pathogenicity of the bacteria

SignalP: Signal Peptide & Lipoproteins

- **Command:**

  -
    ```
    signalp -fasta <input_sequence_file> -org gram- -format short -gff3
    ```
- **Runtime:** T ~ 5 min/file
- **Basis for choosing this software:**
  - Fast runtime
  - .gff output type- easier for merging
  - Multiple versions of SignalP have been developed and compared to previous versions

Signal Peptide & Lipoprotein Results



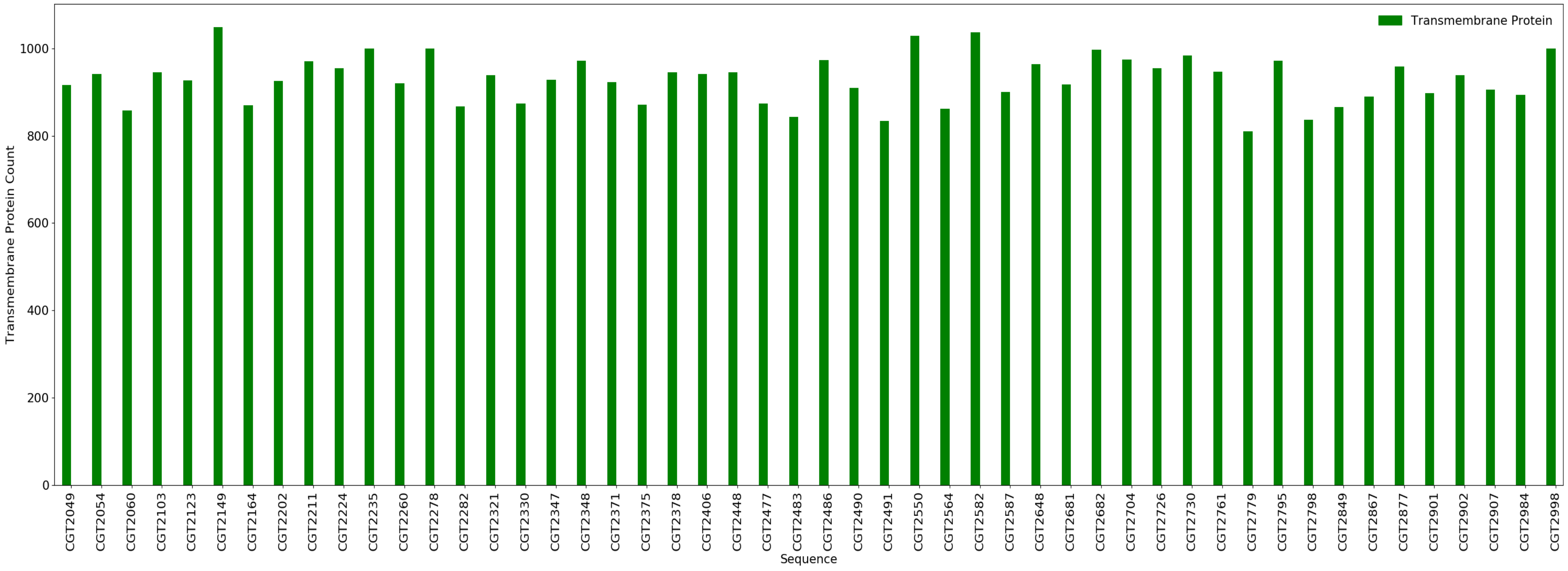
TMHMM2: Transmembrane Proteins

- **Command:**

  ```
  tmhmm <input multifasta file> > <output_file>
  ```

- **Runtime:** T ~ 2 min/file
- **Basis for choosing this software:**
  - Better at differentiating cytoplasmic proteins from transmembrane proteins
  - Alternate tool HMMTOP2 also reports cytoplasmic proteins as transmembrane proteins. Post processing needs to be done, in order to remove the cytoplasmic proteins based on their Topology
- **Sample Output:**
  - **Predicted Helices:** 12
  - **Topology:** i12-34o319-338i345-367o372-394i415-437o452-474i508-530o829-851i856-878o883-905i931-953o963-985i

Transmembrane Protein Results

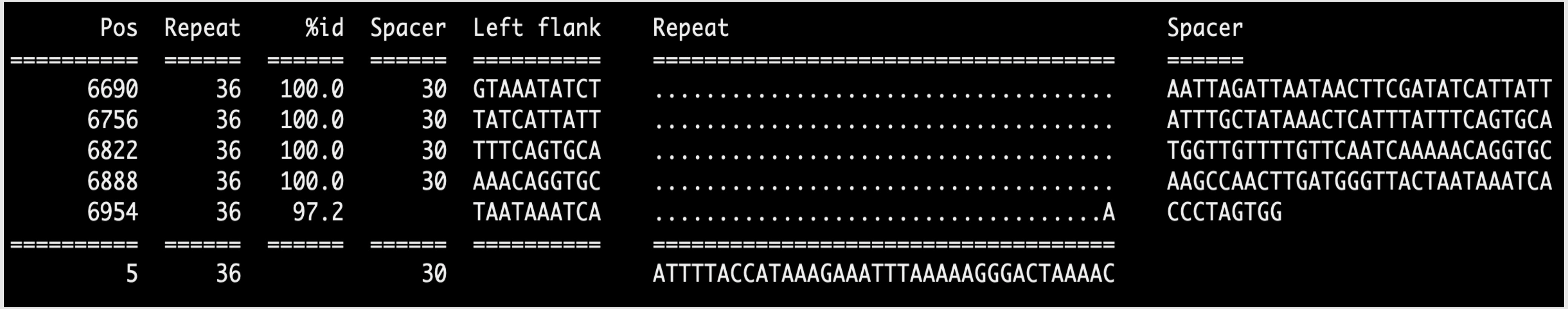


PilerCR: CRISPR Arrays

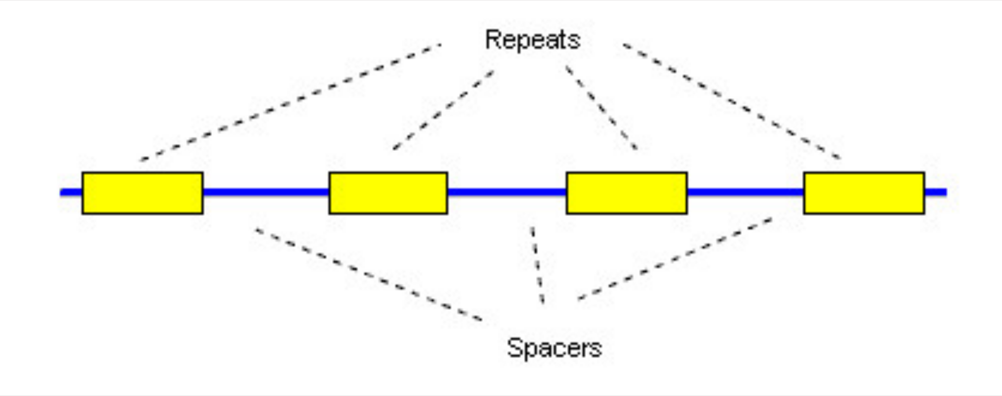
- **Command:**

  -
    ```
    pilercr –in <input multifasta file> -out <output file> -noinfo –quiet
    ```
- **Runtime:** T ~ 1.2 sec/file
- **Basis for choosing this software:**
  - High sensitivity and specificity
  - Fast runtime

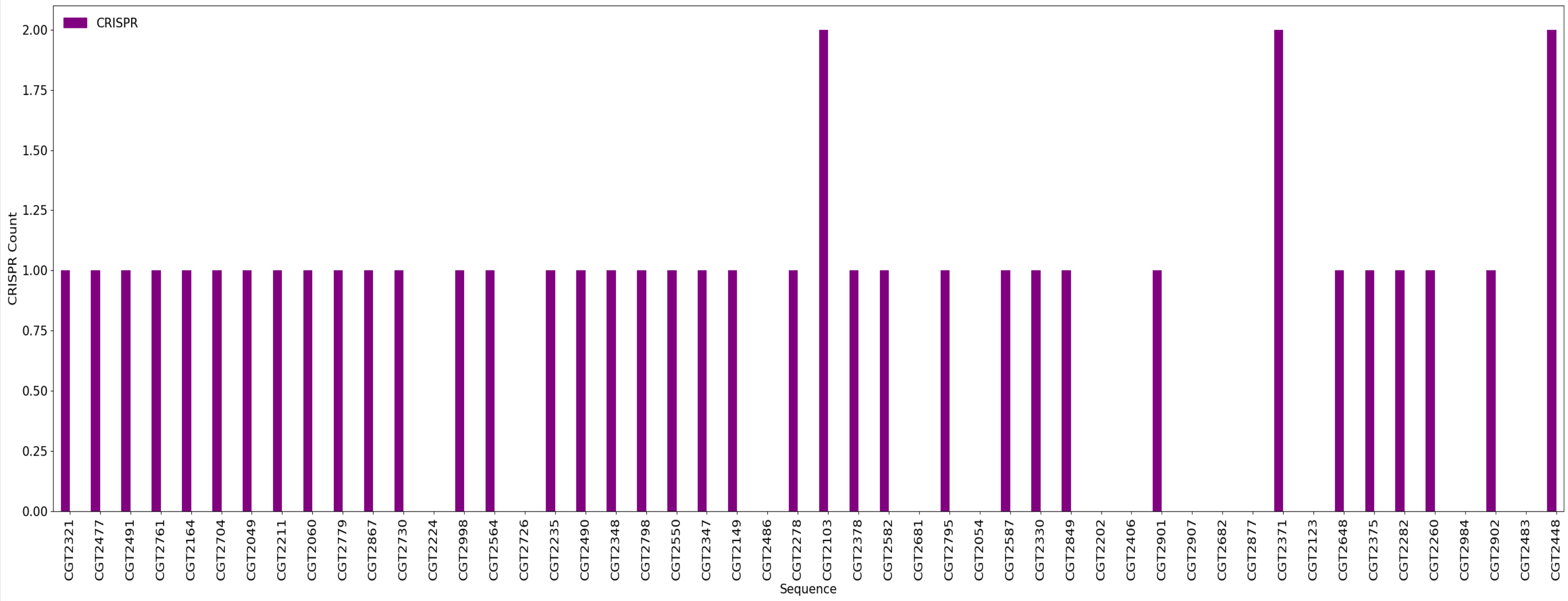
PilerCR: Sample output



```
       Pos  Repeat     %id  Spacer  Left flank    Repeat                                  Spacer
==========  ======  ======  ======  ==========    ====================================    ======
      6690      36   100.0      30  GTAAATATCT    ....................................    AATTAGATTAATAACTTCGATATCATTATT
      6756      36   100.0      30  TATCATTATT    ....................................    ATTTGCTATAAACTCATTTATTTCAGTGCA
      6822      36   100.0      30  TTTCAGTGCA    ....................................    TGGTTGTTTTGTTCAATCAAAAACAGGTGC
      6888      36   100.0      30  AAACAGGTGC    ....................................    AAGCCAACTTGATGGGTTACTAATAAATCA
      6954      36    97.2          TAATAAATCA    ...................................A    CCCTAGTGG
==========  ======  ======  ======  ==========    ====================================
         5      36              30                ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC
```



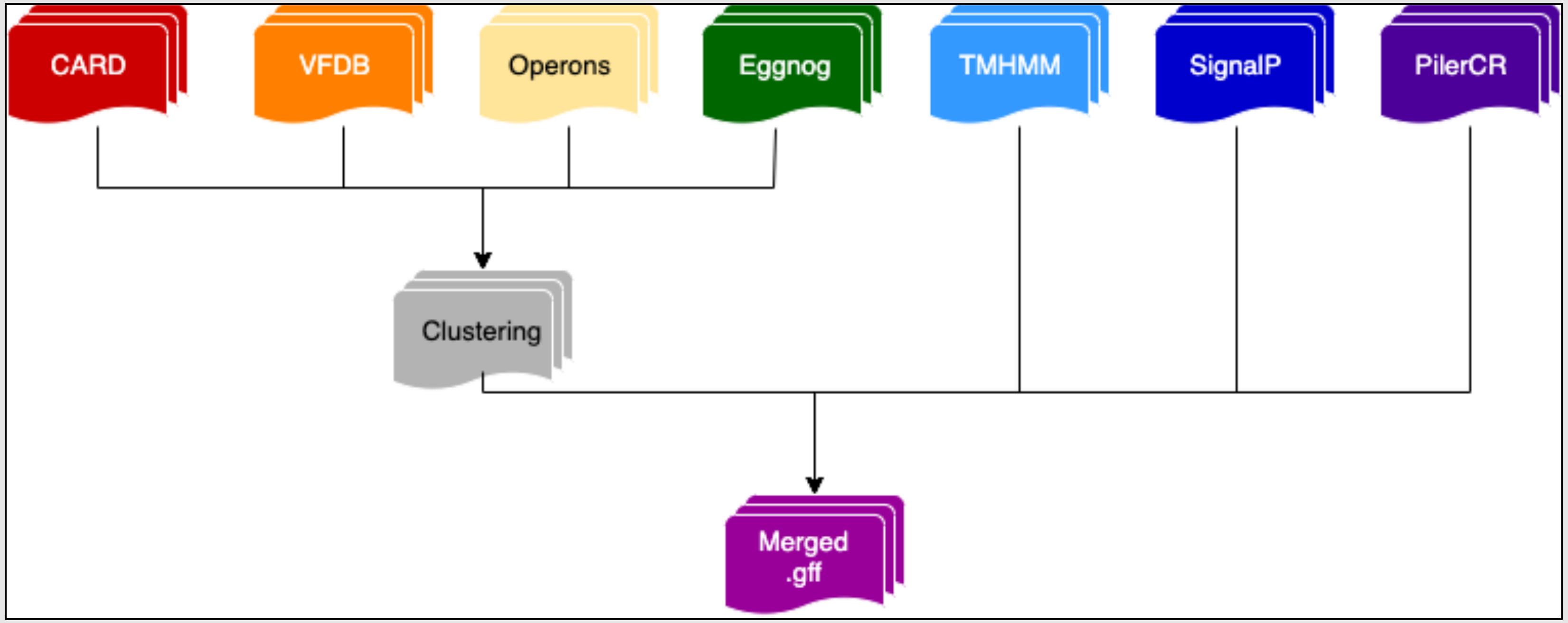
PilerCR Results



ab initio Results Overview

Functional Annotation	Software	Min/Sample	Max/Sample	Median	Mean
Transmembrane Protein	TMHMM	810	1,049	933.5	928.8
Signal Peptide	SignalP	71	99	84	84.3
Lipoprotein	SignalP	43	67	54	52.8
CRISPR	PilerCR	0	2	1	0.8

Deliverables: Merged .gff

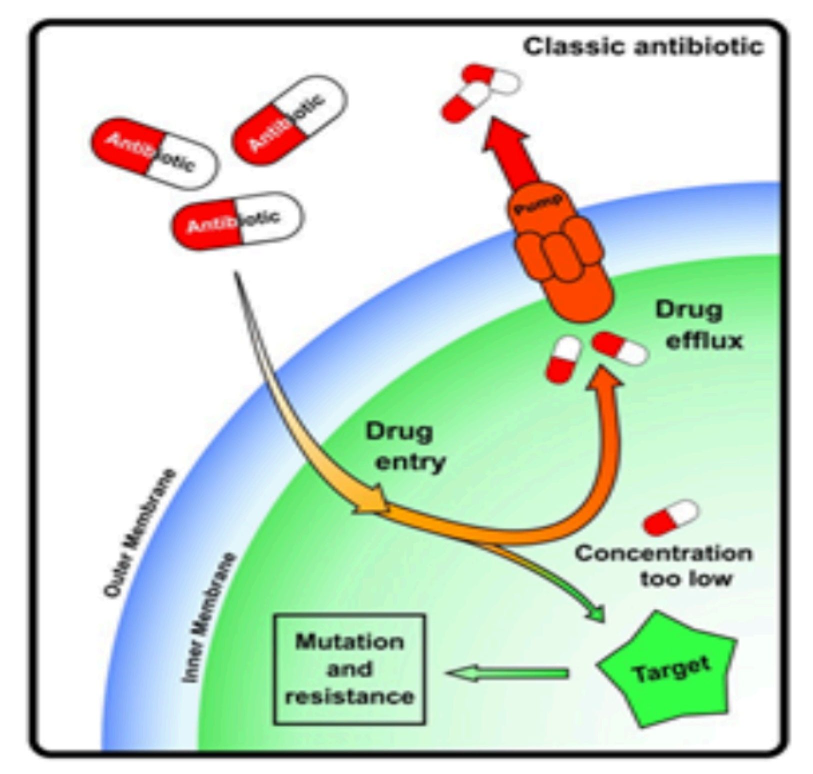


.gff files were created from outputs of each tool, and merged into a single .gff file. Total 50 final .gff files are produced for 50 samples.

Annotations of Interest

Gram-negative bacteria and drug efflux:

- Drug efflux is a key mechanism of resistance in Gram-negative bacteria
- Efflux pumps allow the microorganisms to regulate their internal environment by removing toxic substances like antimicrobial agents
- These systems pump solutes out of the cell, indicating a role of transmembrane protein in the mechanism
- Thus transmembrane proteins involved in the efflux pumps also have antibiotic-resistance properties.



Annotations of Interest

Gram-negative bacteria and drug efflux:

- We found annotations, where predicted Transmembrane Proteins were also identified to be involved in antibiotic resistance
- The mechanism of the resistance is reported as "antibiotic efflux pump", which validates the drug efflux theory
- The proteins are reported to be resistant towards broad-spectrum antibiotics of class fluoroquinolone

Reported Annotation

- Transmembrane Protein(TMHMM2), with 12 predicted α-helices

- Antibiotic Resistance(CARD)
  to fluoroquinolone antibiotics
  with antibiotic efflux pump mechanism

Example from sample CGT2202.
Sequence: NODE_5_length_54021_cov_9.675326
Coordinates: 11462-14586

References

Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1899501/
Genomic Characterization of *Campylobacter jejuni* Strain M1 . https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2928727/
A Genome-Wide Association Study to Identify Diagnostic Markers for Human Pathogenic *Campylobacter jejuni* Strains https://www.frontiersin.org/articles/10.3389/fmicb.2017.01224/full
Campylobacter jejuni transcriptional and genetic adaptation during human infection https://centerforimmunizationresearch.org/wp-content/uploads/2018/08/Campylobacter-jejuni-transcriptional-and-genetic-adaptation-during-human-infection.pdf
A proteome-wide protein interaction map for *Campylobacter jejuni* https://link.springer.com/article/10.1186/gb-2007-8-7-r130
Li, Weizhong & Fu, Limin & Wu, Sitao & Wooley, John. (2012). Ultrafast clustering algorithms for metagenomic sequence analysis. *Briefings in bioinformatics*. 13. 10.1093/bib/bbs035.
Combining multiple functional annotation tools increases coverage of metabolic annotation . https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-5221-9
https://vbn.aau.dk/ws/portalfiles/portal/306201236/Hsu_et_al._2019_Comparative_genomics_and_genome_biology_of_Campylobacter_showae.pdf
https://services.healthtech.dtu.dk
https://link.springer.com/protocol/10.1007%2F978-1-4939-1115-8_22
https://www.nature.com/articles/s41467-018-05969-w
Xavier, Basil Britto et al. "Consolidating and Exploring Antibiotic Resistance Gene Data Resources." *Journal of clinical microbiology* vol. 54,4 (2016): 851-9. doi:10.1128/JCM.02717-15
Huerta-Cepas, Jaime & Forslund, Sofia & Coelho, Luis Pedro & Szklarczyk, Damian & Jensen, Lars & von Mering, Christian & Bork, Peer. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Molecular Biology and Evolution. 34. 10.1093/molbev/msx148.
Alcock, Brian P et al. "CARD 2020: Antibiotic Resistome Surveillance With The Comprehensive Antibiotic Resistance Database". *Nucleic Acids Research*, 2019. *Oxford University Press (OUP)*, doi:10.1093/nar/gkz935
M. N. Price, E. J. Alm, and A. P. Arkin (2005). Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication.. Nucleic Acids Research **33**:3224-3234.
Marasini, Daya et al. "Phylogenetic Relatedness Among Plasmids Harbored by *Campylobacter jejuni* and *Campylobacter coli* Isolated From Retail Meats." *Frontiers in microbiology* vol. 9 2167. 12 Sep. 2018, doi:10.3389/fmicb.2018.02167
Carattoli, Alessandra et al. "In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing." *Antimicrobial agents and chemotherapy* vol. 58,7 (2014): 3895-903. doi:10.1128/AAC.02412-14
Marasini, Daya et al. "Phylogenetic Relatedness Among Plasmids Harbored by *Campylobacter jejuni* and *Campylobacter coli* Isolated From Retail Meats." *Frontiers in microbiology* vol. 9 2167. 12 Sep. 2018, doi:10.3389/fmicb.2018.02167
Carattoli, Alessandra et al. "In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing." *Antimicrobial agents and chemotherapy* vol. 58,7 (2014): 3895-903. doi:10.1128/AAC.02412-14
Roosaare M, Puustusmaa M, Möls M, Vaher M, Remm M. 2018. PlasmidSeeker: identification of known plasmids from bacterial whole genome sequencing reads. *PeerJ*6:e4588 https://doi.org/10.7717/peerj.4588
Weizhong Li, Lukasz Jaroszewski, Adam Godzik, Clustering of highly homologous sequences to reduce the size of large protein databases , *Bioinformatics*, Volume 17, Issue 3, March 2001, Pages 282–283, https://doi.org/10.1093/bioinformatics/17.3.282
Steinegger, M., Söding, J. Clustering huge protein sequence sets in linear time. *Nat Commun* 9, 2542 (2018). https://doi.org/10.1038/s41467-018-04964-5
Benjamin T James, Brian B Luczak, Hani Z Girgis, MeShClust: an intelligent tool for clustering DNA sequences, *Nucleic Acids Research*, Volume 46, Issue 14, 21 August 2018, Page e83, https://doi.org/10.1093/nar/gky315
Robert C. Edgar, Search and clustering orders of magnitude faster than BLAST, *Bioinformatics*, Volume 26, Issue 19, 1 October 2010, Pages 2460–2461, https://doi.org/10.1093/bioinformatics/btq461
Soto SM. Role of efflux pumps in the antibiotic resistance of bacteria embedded in a biofilm. *Virulence*. 2013;4(3):223–229. doi:10.4161/viru.23724

Thank you!

Questions?