MAG5

Metagenomic Assembly Analysis

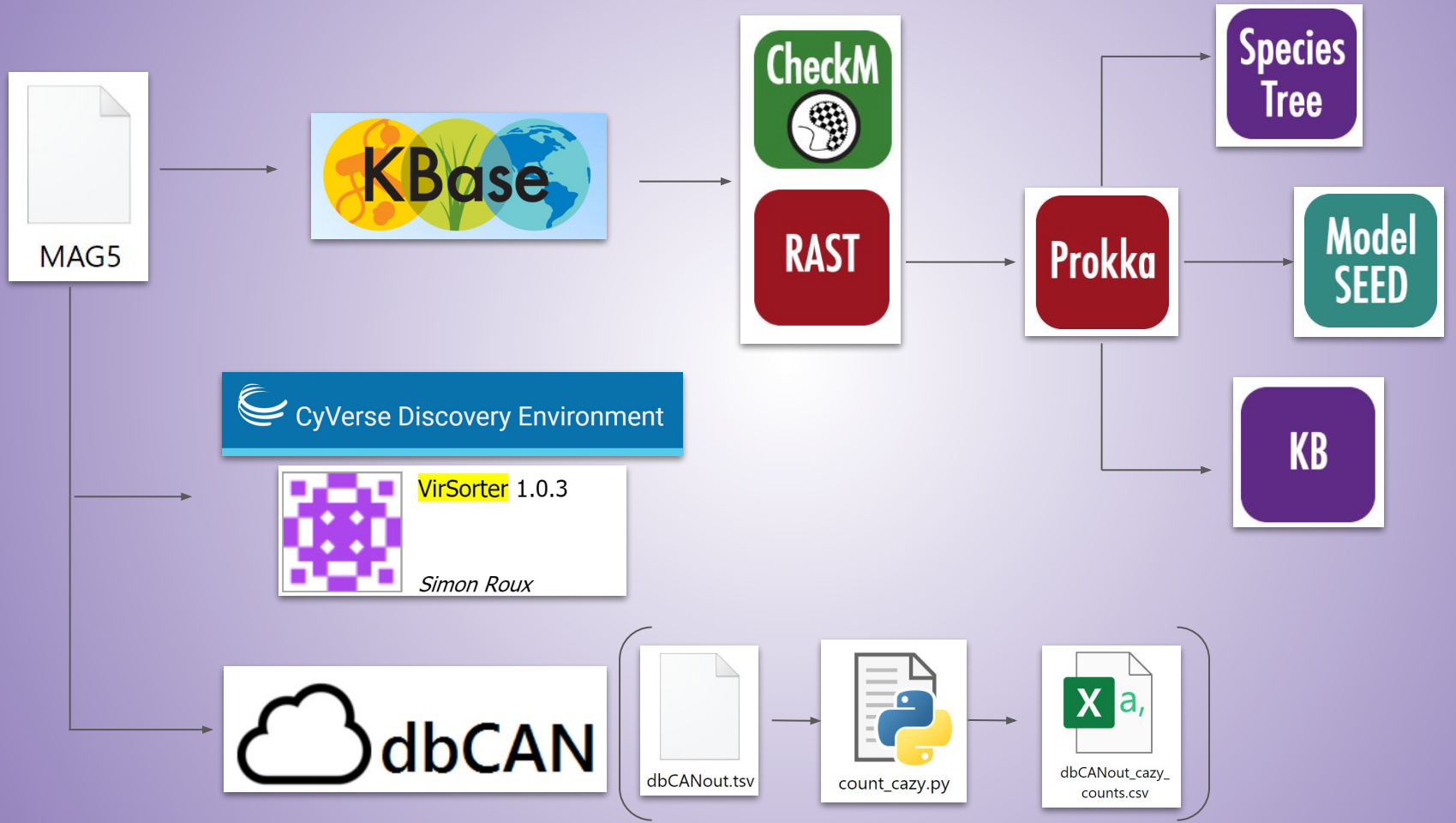
By David Podorefsky
Na Liu
Ketaki Malaviya
Nikola Malinov
Julian Elmurib
Robert Davison

Introduction

- Possible untapped ***biotech potential*** in goat rumen
  - Biofuels, Antibiotics, etc.
- Identify and research potential applications of MAG5

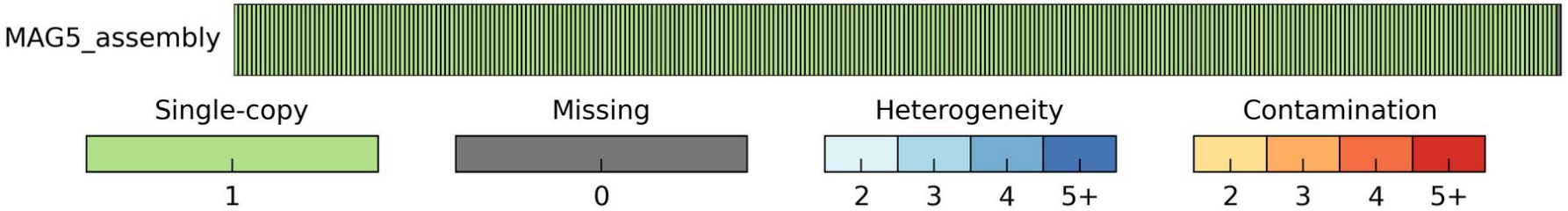
San Clemente Island Goats

Methods


MAG5
KBase
CheckM
RAST
Prokka
Species Tree
Model SEED
KB
CyVerse Discovery Environment
VirSorter 1.0.3
Simon Roux
dbCAN
dbCANout.tsv
count_cazy.py
dbCANout_cazy_counts.csv

Genome Quality

Figure 1.
Genome Quality Plot



Note: Plot of phylogenetic marker genes. Green bars represent a single copy of a marker, while a grey bar indicates a missing clade-specific marker.

Table 1.
Quality Check

Bin Name	Marker Lineage	# Genomes	# Markers	# Marker Sets	0	1	2	3	4	5+	Completeness	Contamination
MAG5_assembly	p__Firmicutes	100	295	158	1	294	0	0	0	0	99.367	0.0

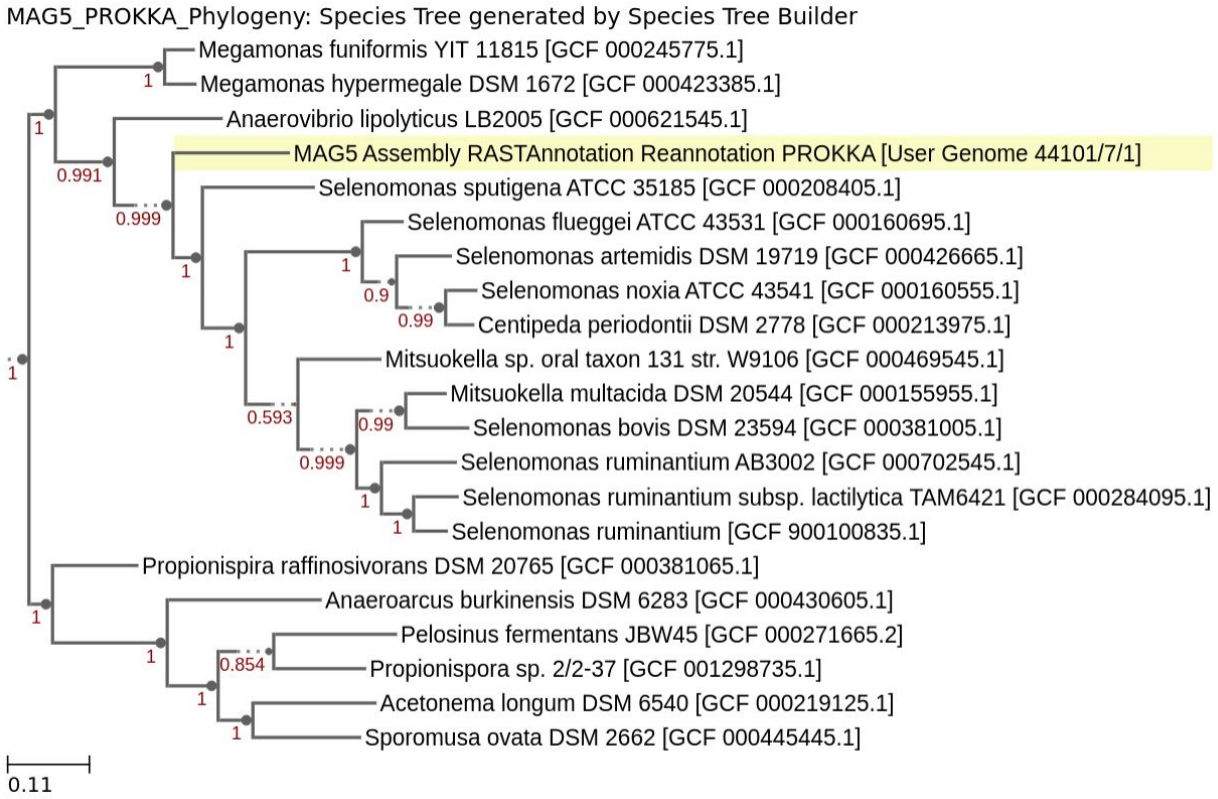
Note: The KBase app Check M v1.0.8 was used to assess the quality of the MAG5 genome.

Clusters of Orthologous Genes (COGs)

- The **Insert Genome Into Species Tree 2.1.10** App uses **49** COGs when building the species tree
- These COGs are predominantly ***ribosomal subunits***
- Ribosomes are used for ***phylogenetic analysis*** they're present in all organisms
- The COGs used for ***Energy acquisition*** are much broader and cannot be compared from species to species to produce the phylogenetic tree

COG0012	COG0012	Predicted GTPase, probable translation factor [Translation, ribosomal structure and biogenesis].
COG0013	AlaS	Alanyl-tRNA synthetase [Translation, ribosomal structure and biogenesis].
COG0016	PheS	Phenylalanyl-tRNA synthetase alpha subunit [Translation, ribosomal structure and biogenesis].
COG0018	ArgS	Arginyl-tRNA synthetase [Translation, ribosomal structure and biogenesis].
COG0030	KsgA	Dimethyladenosine transferase (rRNA methylation) [Translation, ribosomal structure and biogenesis].

Phylogenetic Tree



Phylogeny

- RNAmmer used to identify ribosomal RNA
  - 5s sub-unit identified
- The obtained sequence was blasted

```
>rRNA_NODE_12_length_459105_cov_16.432523_254-369_DIR+ /molecule=5s_rRNA /score=72.6
CGGTGGTGATGCCTACGTGGTTCCACCTGTTCCCATTCCGAACACAGTAGTTAAGCACGT
AAAGGCCGAAAGTACTTGGCTGGAGACGGCCCGGGAGGATAGGAAGCTGCCGGTTA
MAG5_rRNA_seq.fa (END)
```

Sequences producing significant alignments:

Select: All None Selected:0

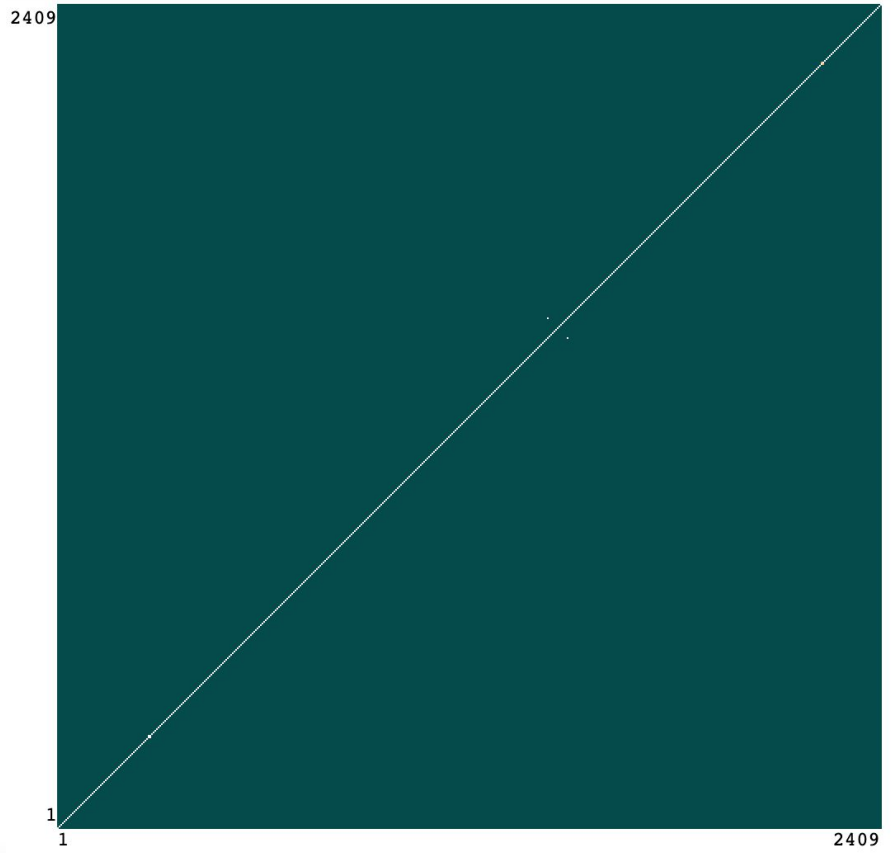
Alignments Download GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Veillonella rodentium strain NCTC12018 genome assembly, chromosome: 1	137	551	89%	6e-29	90.38%	LT906470.1
Christensenella sp. Marseille-P2438 genome assembly, chromosome: contig00001	137	137	97%	6e-29	88.50%	LT700187.1
Uncultured bacterium BD_contig02516 genomic sequence	106	106	77%	2e-19	87.91%	KJ799706.1
Selenomonas sp. oral taxon 126 strain W7667 genome	102	371	96%	2e-18	83.04%	CP016201.1
Selenomonas sp. oral taxon 920 strain W5150 chromosome, complete genome	97.1	383	96%	1e-16	82.14%	CP017042.1
Selenomonas sp. oral taxon 136 strain F0591, complete genome	97.1	269	96%	1e-16	82.14%	CP014239.1
Eubacterium siraeum V10Sc8a strain V10Sc8a 5S ribosomal RNA, complete sequence	97.1	97.1	81%	1e-16		

Questions/comm

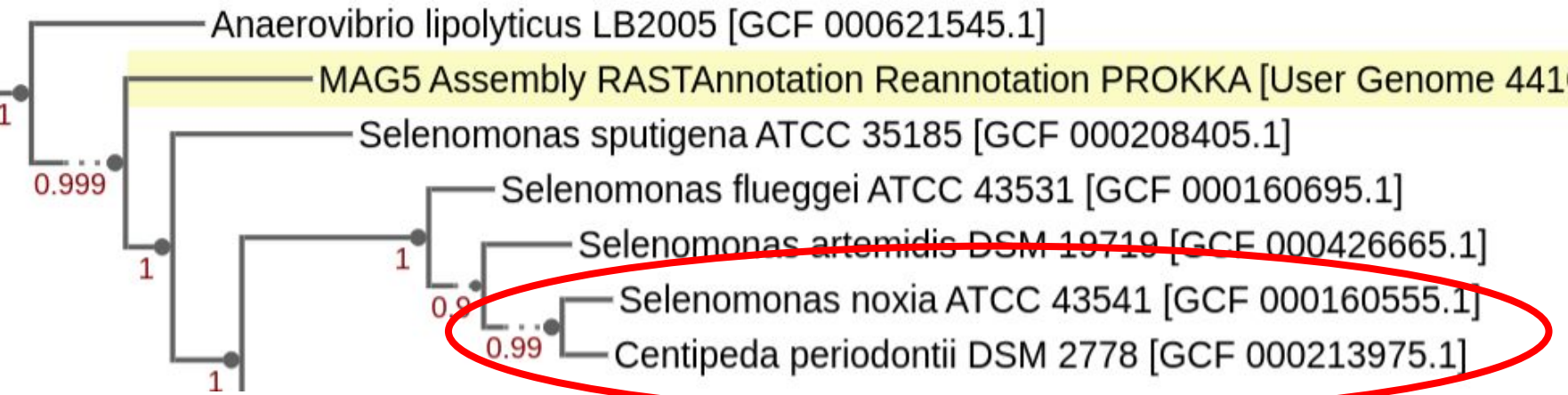
Synteny Plot Comparison

An identical match would look like this:



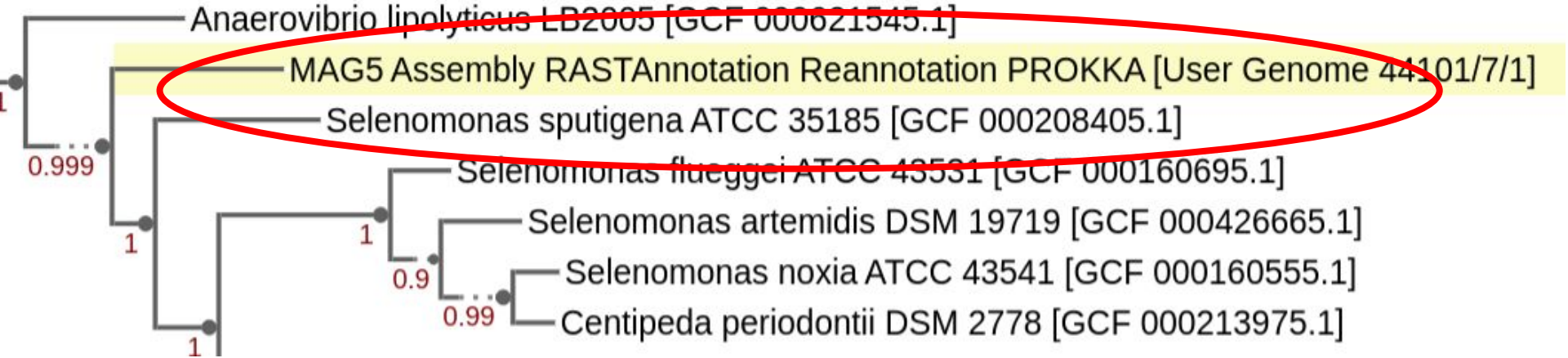
MAG5 compared to MAG5

Example of two known close relatives



Selenomonas Noxia ATCC

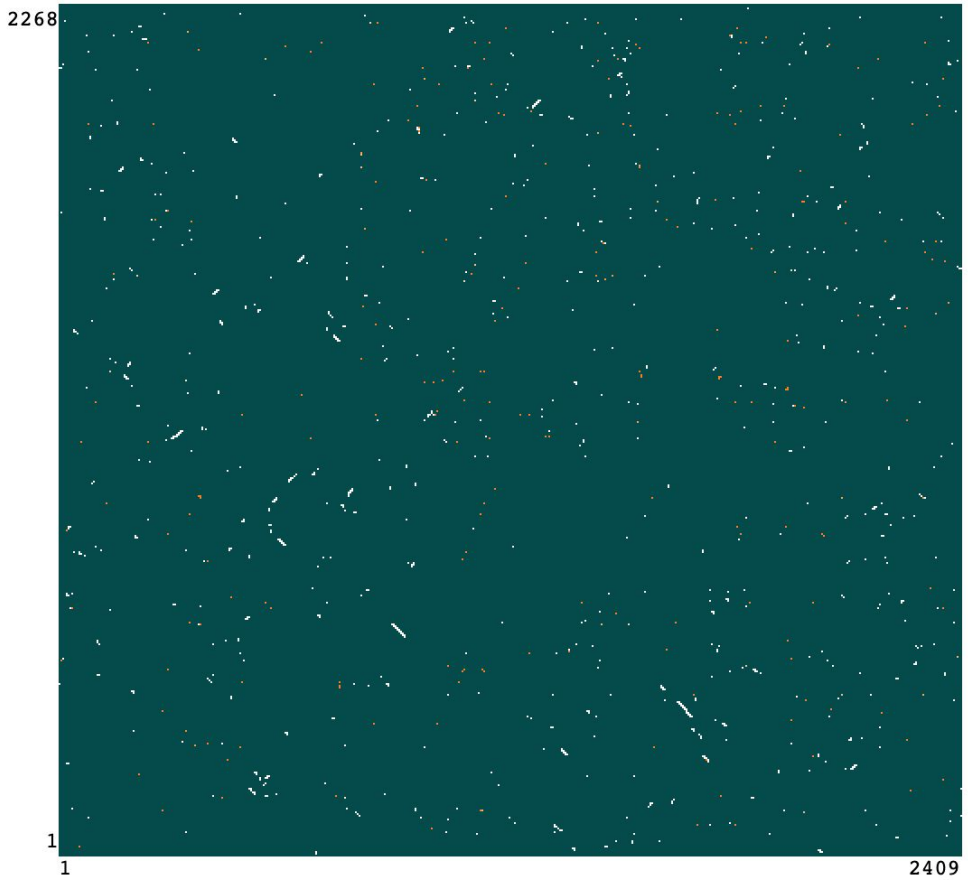
Centipeda Periodontii DSM

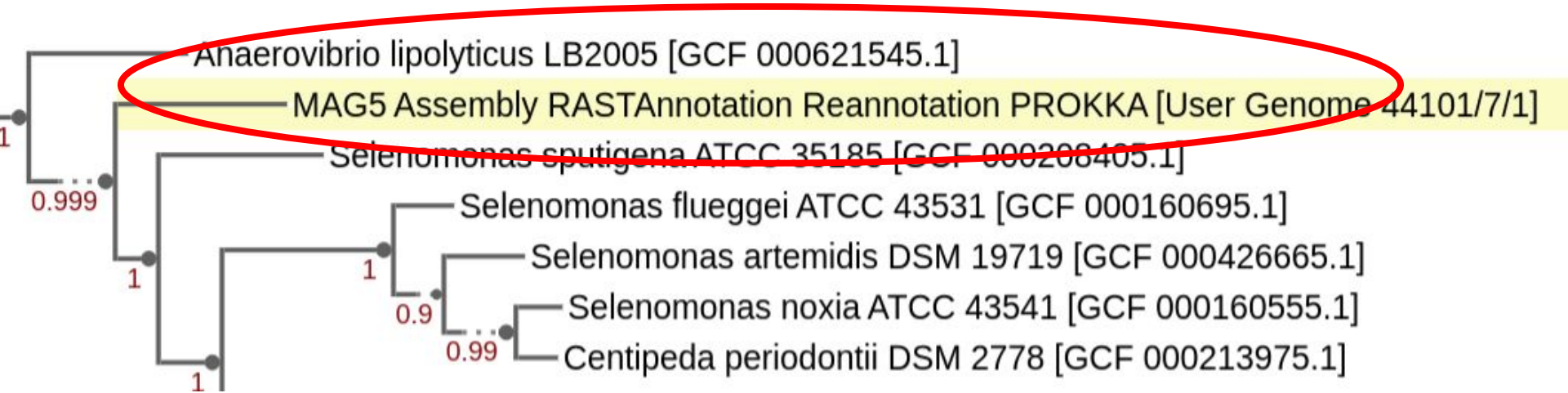

Anaerovibrio lipolyticus LB2005 [GCF 000621545.1]
MAG5 Assembly RASTAnnotation Reannotation PROKKA [User Genome 44101/7/1]
Selenomonas sputigena ATCC 35185 [GCF 000208405.1]
Selenomonas flueggei ATCC 43531 [GCF 000160695.1]
Selenomonas artemidis DSM 19719 [GCF 000426665.1]
Selenomonas noxia ATCC 43541 [GCF 000160555.1]
Centipeda periodontii DSM 2778 [GCF 000213975.1]
0.999
1
1
0.9
0.99
1

Our MAG5 compared to its closest relative Selenomonas Sputigena

MAG5:
2409 genes
1675 hits

Selenomonas Sputigena:
2268 genes
1641 hits



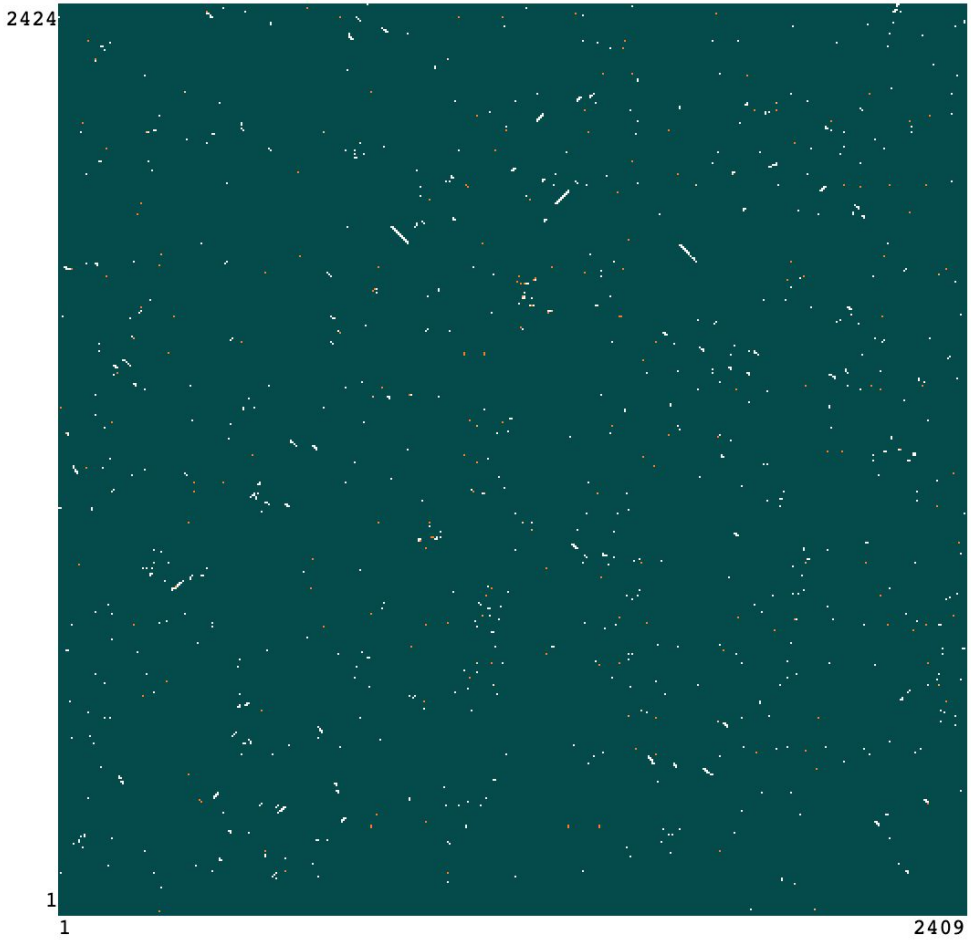

Anaerovibrio lipolyticus LB2005 [GCF 000621545.1]
MAG5 Assembly RASTAnnotation Reannotation PROKKA [User Genome 44101/7/1]
Selenomonas sputigena ATCC 35185 [GCF 000208405.1]
Selenomonas flueggei ATCC 43531 [GCF 000160695.1]
Selenomonas artemidis DSM 19719 [GCF 000426665.1]
Selenomonas noxia ATCC 43541 [GCF 000160555.1]
Centipeda periodontii DSM 2778 [GCF 000213975.1]
1
0.999
1
1
0.9
0.99
1

MAG5 compared to a second close relative slightly further on the tree.

MAG5:
2409 genes
1712 hits

Anaerovibrio Lipolyticus:
2424 genes
1739 hits

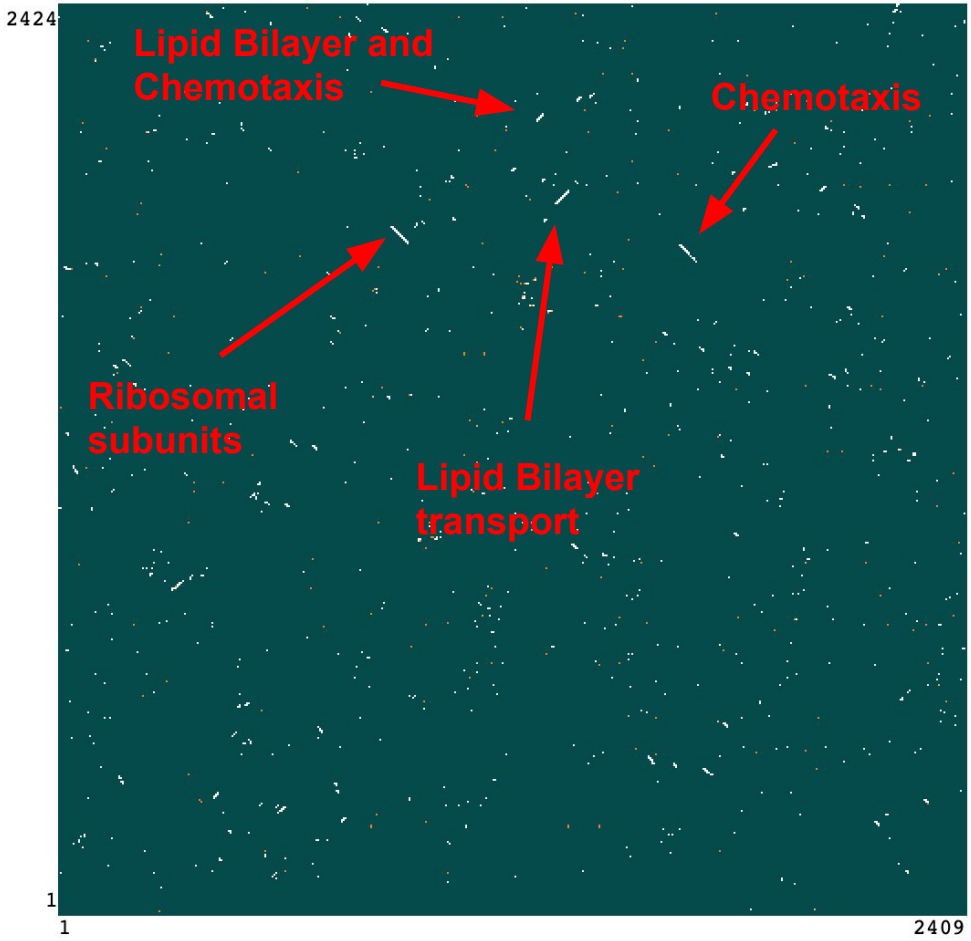
More Hits!



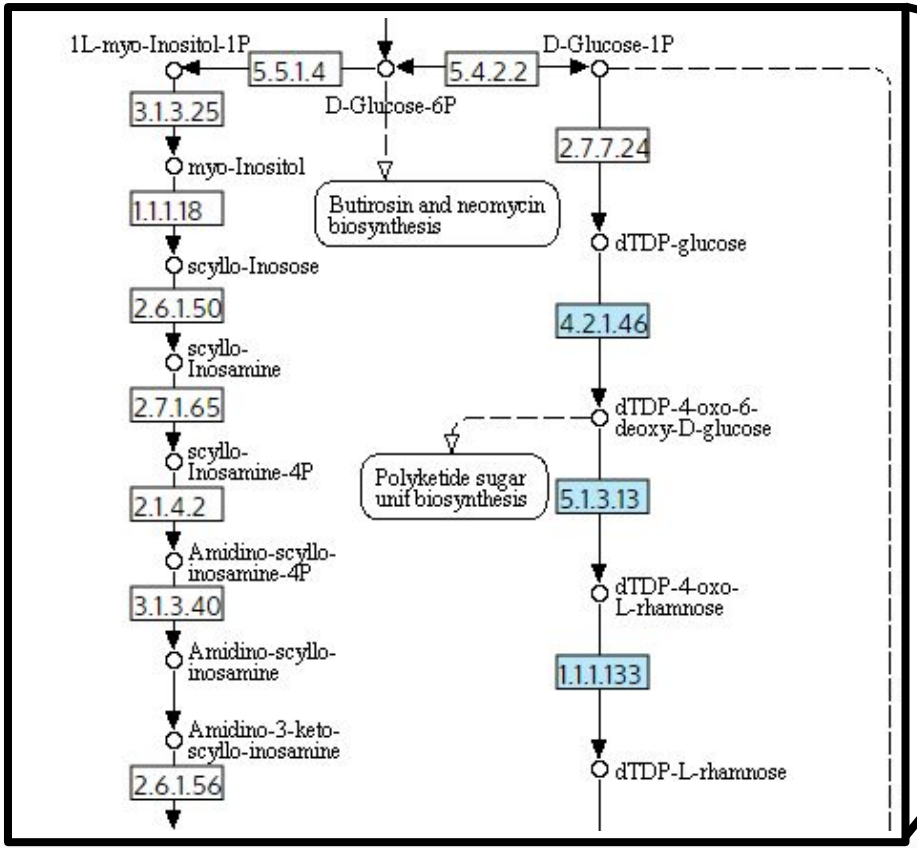
MAG5 compared to a second close relative slightly further on the tree.

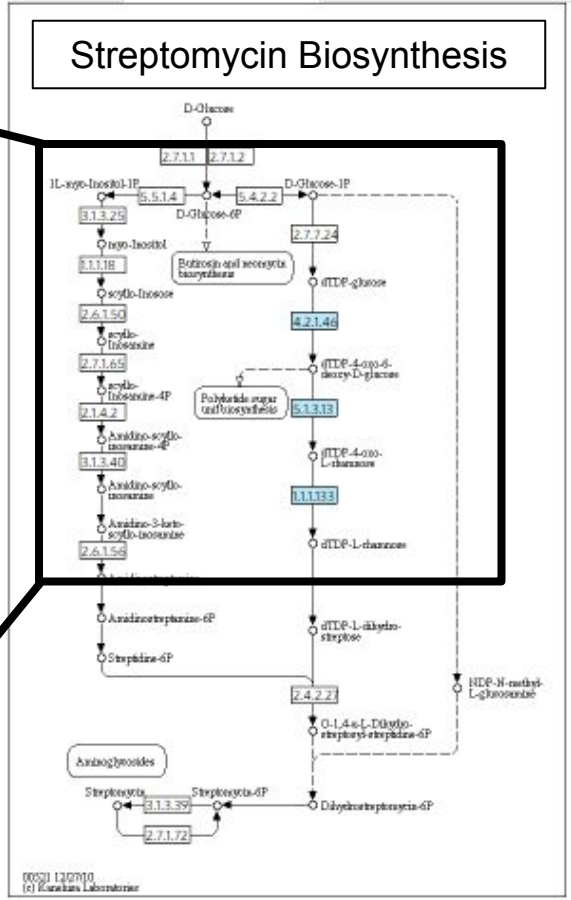
MAG5:
2409 genes
1712 hits

Anaerovibrio Lipolyticus:
2424 genes
1739 hits

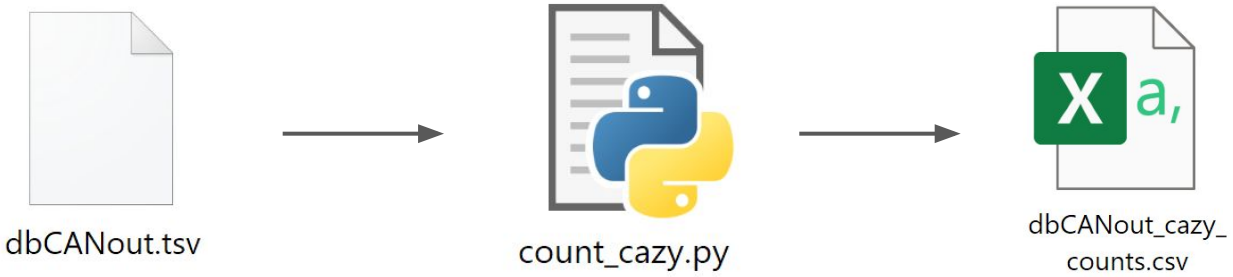


Weak Evidence Suggests No Potential for Antibiotic Production





CAZyme Pipeline



1. Determines CAZymes with 3 hidden Markov model (HMM) variations
2. Utilize python script to keep those present in all three databases - credit to Tom
3. Count and realize enzymes

Hidden Markov Model

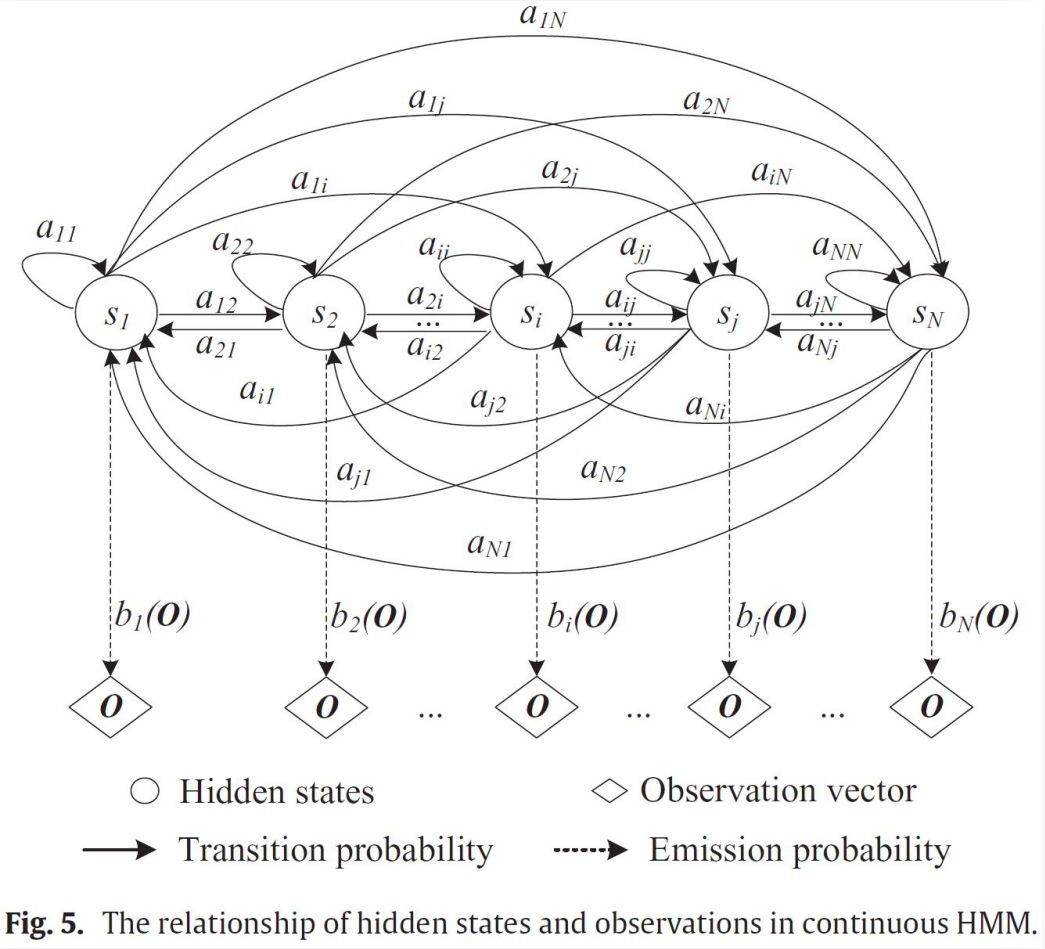


Fig. 5. The relationship of hidden states and observations in continuous HMM.

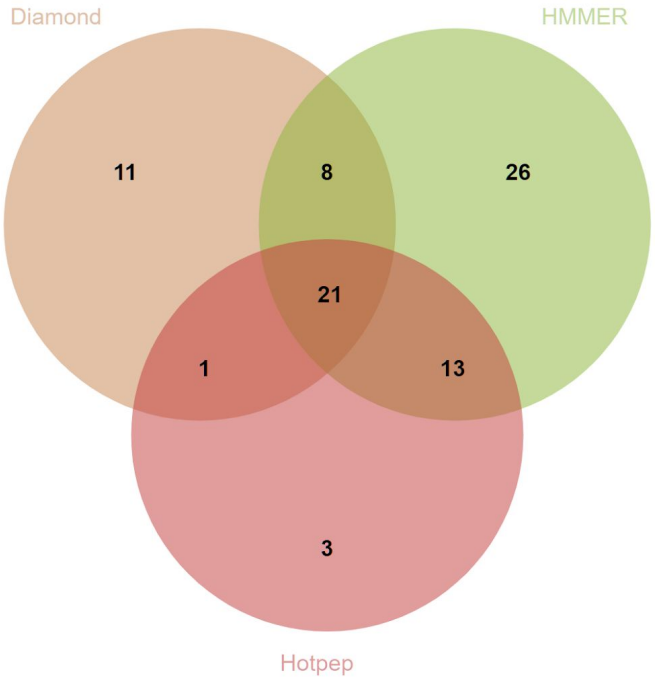
P.Jiang et al. *European Journal of Operational Research* (2016)

dbCAN meta server: Carbohydrate-active enzyme automated CAZyme annotation

Home | Annotate | Download | Help | About us

Cite us: NAR/gky418 and gkz479

DIAMOND:
Fast blast hits in CAZy database
(E-Value < 1e-102)



HMMER:
Domain boundaries according to the dbCAN CAZyme domain HMM database
(E-value < 1e-15, coverage > 0.35)

Hotpep:
Short conserved motifs in PPR library
(Frequency > 2.6, Hits > 6)

http://bcb.unl.edu/dbCAN2/

Overview | HMMER | DIAMOND | Hotpep | CGC-Finder

Download SignalP output **Download Prodigal predictions** **Download this table** (keep those with # of Tools >=2 will give you best result; and use dbCAN domain assignment is recommended) ?

Show 15 ▼ entries Search:

Gene ID	# of Tools	HMMER	DIAMOND	Hotpep	Signal Peptide
NODE_107_length_200687_cov_16.378152_148	1	N	GH6	N	N
NODE_107_length_200687_cov_16.378152_162	1	N	GT13	N	N
NODE_112_length_196477_cov_17.935031_129	2	CE4(77-234)+GH153(281-593)	CE4	N	Y (1-25)
NODE_112_length_196477_cov_17.935031_130	3	GT2_Glyco_tranf_2_3(55-275)	GT2	GT2	N
NODE_112_length_196477_cov_17.935031_24	1	CE10(72-238)	N	N	Y (1-35)
NODE_12_length_459105_cov_16.432523_146	1	N	GH13_30	N	N
NODE_12_length_459105_cov_16.432523_155	3	GT51(69-244)	GT51	GT51	N
NODE_12_length_459105_cov_16.432523_185	2	GH23(454-560)	GH23	N	Y (1-22)
NODE_12_length_459105_cov_16.432523_260	1	GT28(207-356)	N	N	Y (1-21)
NODE_12_length_459105_cov_16.432523_295	3	GT2_Glyco_tranf_2_3(45-268)	GT2	GT2	N
NODE_12_length_459105_cov_16.432523_298	2	GT26(58-226)	N	GT26	N
NODE_12_length_459105_cov_16.432523_391	1	GT2_Glycos_transf_2(2-165)	N	N	N
NODE_12_length_459105_cov_16.432523_468	2	GH73(123-249)	N	GH73	Y (1-29)
NODE_207_length_144259_cov_18.029109_10	2	GT9(74-322)	GT9	N	N
NODE_207_length_144259_cov_18.029109_2	1	GT2_Glycos_transf_2(13-113)	N	N	N

Count CAZymes by Class

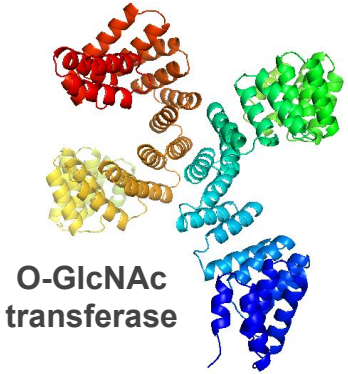
```
Gene ID HMMER   Hotpep  DIAMOND Signalp #ofTools
NODE_112_length_196477_cov_17.935031_129    CE4(77-234)+GH153(281-593)  N   CE4 Y(1-25) 2
NODE_112_length_196477_cov_17.935031_130    GT2_Glyco_tranf_2_3(55-275) GT2 GT2 N   3
NODE_112_length_196477_cov_17.935031_24 CE10(72-238)    N   N   Y(1-35) 1
NODE_12_length_459105_cov_16.432523_155 GT51(69-244)    GT51    GT51    N   3
NODE_12_length_459105_cov_16.432523_185 GH23(454-560)   N   GH23    Y(1-22) 2
NODE_12_length_459105_cov_16.432523_260 GT28(207-356)   N   N   Y(1-21) 1
NODE_12_length_459105_cov_16.432523_295 GT2_Glyco_tranf_2_3(45-268) GT2 GT2 N   3
NODE_12_length_459105_cov_16.432523_298 GT26(58-226)    GT26    N   N   2
NODE_12_length_459105_cov_16.432523_391 GT2_Glycos_transf_2(2-165)  N   N   N   1
NODE_12_length_459105_cov_16.432523_468 GH73(123-249)   GH73    N   Y(1-29) 2
NODE_207_length_144259_cov_18.029109_10 GT9(74-322) N   GT9 N   2
NODE_207_length_144259_cov_18.029109_2  GT2_Glycos_transf_2(13-113) N   N   N   1
NODE_207_length_144259_cov_18.029109_26 GT4(197-333)    GT4 GT4 N   3
NODE_207_length_144259_cov_18.029109_27 GT4(152-316)    N   N   N   1
NODE_207_length_144259_cov_18.029109_28 GT2_Glycos_transf_2(4-176)  GT2 N   N   2
NODE_207_length_144259_cov_18.029109_3  GT9(76-340) N   GT9 N   2
NODE_207_length_144259_cov_18.029109_30 GT2_Glycos_transf_2(9-127)  N   N   N   1
NODE_207_length_144259_cov_18.029109_49 GT4(205-353)    N   N   N   1
NODE_207_length_144259_cov_18.029109_8  GT9(7-262)   N   N   1
NODE_207_length_144259_cov_18.029109_91 GT2_Glycos_transf_2(5-158)  GT2 GT2 N   3
NODE_207_length_144259_cov_18.029109_98 GT35(106-513)   GT35    GT35    N   3
NODE_207_length_144259_cov_18.029109_99 CBM48(26-114)+GH13_9(179-479)   GH13+CBM48  CBM48   N   3
NODE_207_length_144259_cov_18.029109_90 GT5(2-475)  GT5 GT5 N   3
NODE_207_length_144259_cov_18.029109_91 GH13_20(185-517)+GH77(664-1129) GH13+GH77   CBM34   N   3
NODE_207_length_144259_cov_18.029109_99 GT83(3-366) GT83    GT83    N   3
NODE_239_length_191538_cov_22.771415_43 CE4(87-219) CE4 N   Y(1-26) 2
NODE_25_length_383736_cov_17.808888_13  CE9(22-384) N   CE9 N   2
NODE_25_length_383736_cov_17.808888_130 GT30(33-215)    GT30    GT30    N   3
NODE_25_length_383736_cov_17.808888_132 GT19(3-353) GT19    GT19    N   3
NODE_25_length_383736_cov_17.808888_136 CE11(12-278)    CE11    CE11    N   3
NODE_25_length_383736_cov_17.808888_18  GT9(208-361)    GT0 N   N   2
NODE_25_length_383736_cov_17.808888_19  GT8(17-255) N   N   N   1
NODE_25_length_383736_cov_17.808888_216 CE10(103-261)   N   N   N   1
NODE_25_length_383736_cov_17.808888_222 GT41(4-566) N   GT41    N   2
NODE_25_length_383736_cov_17.808888_230 GH84(28-323)    GH84    GH84    N   3
NODE_25_length_383736_cov_17.808888_27  GT2_Glycos_transf_2(5-119)   N   N   N   1
NODE_25_length_383736_cov_17.808888_299 GH3(117-356)    N   N   N   1
NODE_25_length_383736_cov_17.808888_30  GT2_Glycos_transf_2(19-146) GT2 N   N   2
NODE_25_length_383736_cov_17.808888_31  GT4(277-428)    N   N   N   1
NODE_25_length_383736_cov_17.808888_35  GT9(290-439)    N   N   N   1
NODE_25_length_383736_cov_17.808888_47  GT2_Glyco_tranf_2_4(11-108) N   N   N   1
NODE_25_length_383736_cov_17.808888_48  GT2_Glycos_transf_2(6-174)   N   N   N   1
NODE_25_length_383736_cov_17.808888_49  GT2_Glyco_tranf_2_5(7-218)   N   N   N   1
NODE_25_length_383736_cov_17.808888_56  GT2_Glyco_tranf_2_5(8-197)+GT2_Glycos_transf_2(230-363) N   N   N   1
NODE_25_length_383736_cov_17.808888_57  GT2_Glyco_tranf_2_5(12-220)  N   N   N   1
NODE_25_length_383736_cov_17.808888_60  GT41(10-526)    GT41    GT41    N   3
NODE_25_length_383736_cov_17.808888_62  GT41(56-488)    GT41    GT41    N   3
NODE_25_length_383736_cov_17.808888_63  GT41(4-319)+GT41(339-834)   N   GT41    N   2
NODE_25_length_383736_cov_17.808888_64  GT2_Glycos_transf_2(5-101)+GT41(318-593)+GT41(924-1454) GT41    GT41    N   3
NODE_25_length_383736_cov_17.808888_65  GT2_Glycos_transf_2(5-101)  GT2 N   N   2
NODE_25_length_383736_cov_17.808888_66  GT2_Glycos_transf_2(4-107)+GT2_Glycos_transf_2(345-485)   GT2 GT2 N   3
NODE_25_length_383736_cov_17.808888_67  GT2_Glycos_transf_2(6-115)   GT2 N   N   2
NODE_28_length_375312_cov_16.951897_124 GH23(86-204)    GH23    N   N   2
NODE_28_length_375312_cov_16.951897_170 GT28(188-355)   GT28    GT28    N   3
NODE_28_length_375312_cov_16.951897_178 CE4(101-229)    CE4 N   N   2
NODE_28_length_375312_cov_16.951897_259 GT9(147-354)    N   N   N   1
NODE_28_length_375312_cov_16.951897_274 GT11(1-271) N   N   N   1
NODE_28_length_375312_cov_16.951897_28  GT51(79-243)    GT51    N   N   2
NODE_28_length_375312_cov_16.951897_326 CE4(96-223) N   N   N   1
NODE_28_length_375312_cov_16.951897_34  CE1(202-482) N   N   Y(1-29) 1
NODE_28_length_375312_cov_16.951897_72  GH23(68-197)    GH23    N   N   2
NODE_28_length_375312_cov_16.951897_92  CE3(95-232) N   N   Y(1-32) 1
NODE_299_length_111202_cov_18.027123_22 GT41(106-547)   GT41    GT41    N   3
NODE_299_length_111202_cov_18.027123_80 GH18(95-356)    N   N   Y(1-23) 1
NODE_472_length_76310_cov_17.246075_43  GT2_Glycos_transf_2(5-163)   N   GT2 N   2
NODE_581_length_62691_cov_17.574552_51  GT9(76-318) GT9 GT9 N   3
NODE_726_length_50179_cov_16.117540_11  CE4(80-209) CE4 N   N   2
NODE_990_length_35575_cov_18.293481_18  GT9(128-380)    N   N   N   1
NODE_28_length_375312_cov_16.951897_107 N   GH23    N   N   1
NODE_207_length_144259_cov_18.029109_41 N   GT0 GT0 N   2
NODE_995_length_35575_cov_18.293481_11  N   GT0 N   N   1
NODE_207_length_144259_cov_18.029109_22 N   GT26    N   N   1
NODE_12_length_459105_cov_16.432523_146 N   N   GH13_30 N   1
NODE_25_length_383736_cov_17.808888_176 N   N   GH13_11 N   1
NODE_28_length_375312_cov_16.951897_70  N   N   GH28    N   1
NODE_28_length_375312_cov_16.951897_184 N   N   GT2 N   1
NODE_28_length_375312_cov_16.951897_273 N   N   GH6 N   1
NODE_107_length_200687_cov_16.378152_148    N   N   GH6 N   1
NODE_107_length_200687_cov_16.378152_162    N   N   GT13    N   1
NODE_299_length_111202_cov_18.027123_38 N   N   GT47    N   1
NODE_472_length_76310_cov_17.246075_64  N   N   CBM50   Y(1-25) 1
NODE_581_length_62691_cov_17.574552_19  N   N   GT0 N   1
NODE_581_length_62691_cov_17.574552_26  N   N   GT4 N   1
```

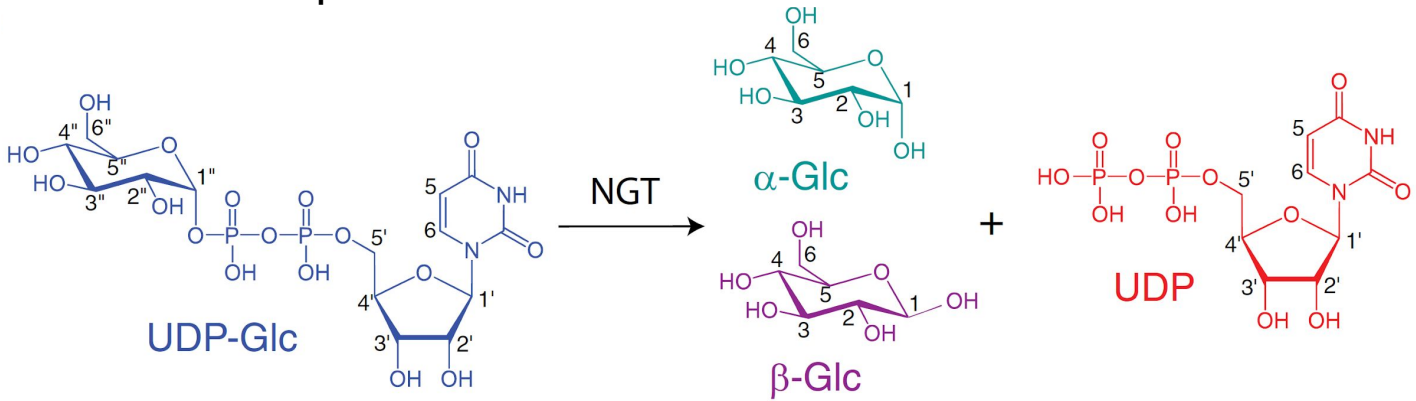
Glucosyltransferase
Glycoside Hydrolase
Carbohydrate Esterase
Carbohydrate Binding Module

GT2	6
GT51	1
GT4	1
GT35	1
CBM48	1
GT5	1
GT83	1
GT30	1
GT19	1
CE11	1
GH84	1
GT41	5
GT28	1
GT9	1
GH13	2
GH77	1

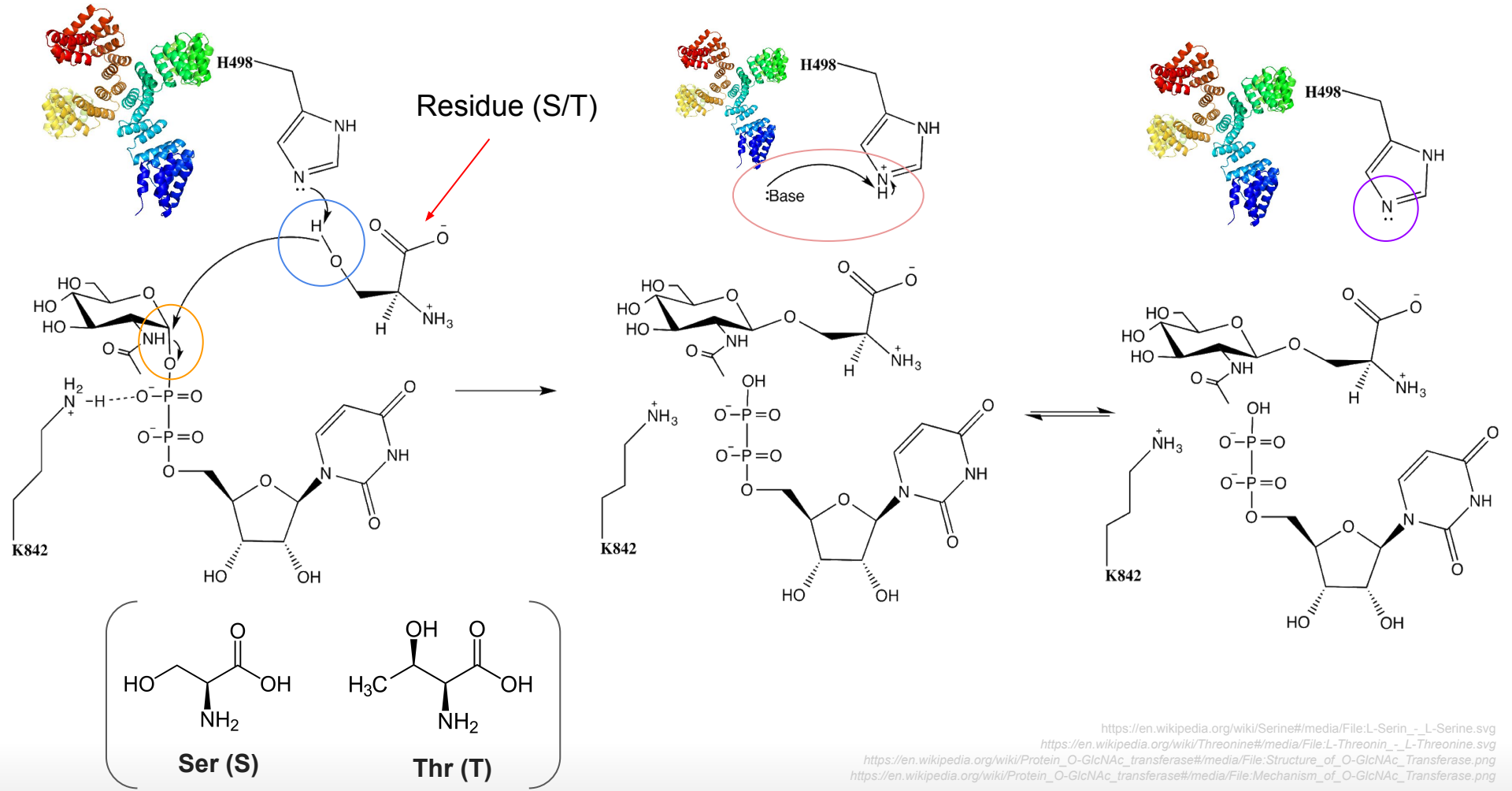
Glycosyl transferase family 41 (GT41) (N-glycosyltransferase)

- Attaches hexoses (saccharide moiety) to nitrogen atom in an amide group (asparagine) on target protein
- Found in bacteria *Actinobacillus pleuropneumoniae*
- Targets adhesion proteins for surface attachment stability
- Can hydrolyze sugar when no substrate present
- Resembles protein O-GlcNAc transferase





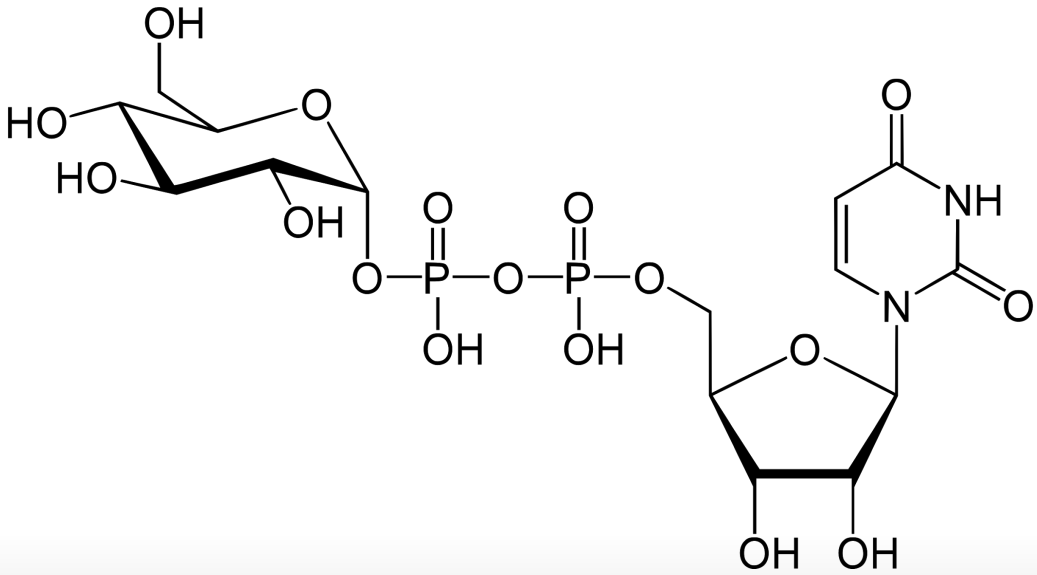
S. Grass *et al. Molecular Microbiology (2003)*
A. Naegeli *et al. Journal of Biological Chemistry (2014)*



https://en.wikipedia.org/wiki/Serine#/media/File:L-Serin_-_L-Serine.svg
https://en.wikipedia.org/wiki/Threonine#/media/File:L-Threonin_-_L-Threonine.svg
https://en.wikipedia.org/wiki/Protein_O-GlcNAc_transferase#/media/File:Structure_of_O-GlcNAc_Transferase.png
https://en.wikipedia.org/wiki/Protein_O-GlcNAc_transferase#/media/File:Mechanism_of_O-GlcNAc_Transferase.png

Glycosyl Transferase Family 2 (GT2)

- Diverse: > 3500 sequences from animal, plant, fungi, and bacteria, > 12 distinct GT functions have already been characterized
- Cellulose synthase, chitin synthase, mannosyltransferase, glucosyltransferase, galactosyltransferase, rhamnosyltransferase, etc.



UDP-glucose

Campbell et al., Biochem. J. (1997)

Other CAzyme genes found in MAG 5

- Glycoside Hydrolase (GH13, GH77, GH84): hydrolysis and rearrangement of glycosidic bonds
- Carbohydrate Esterase(CE11): UDP-3-0-acyl N-acetylglucosamine deacetylase
- Carbohydrate Binding Module (CBM48): glycogen-binding function, appended to GH13 modules

CAZyme Gene Clusters
(CGCs)

Show 15 ▼ entries Search:

Cluster	Contig ID	# of Genes	# of CAZymes	# of TC	# of TF	Cluster Start	Cluster End
CGC1		12	1	6	0	184995	200444
CGC10		2	1	1	0	151765	154876
CGC11		3	1	1	0	12762	15952
CGC12		6	1	2	0	125774	131698
CGC13		3	1	1	0	170482	174002
CGC2		4	1	1	0	295175	298678
CGC3		2	1	1	0	454119	456221
CGC4		3	1	2	0	9284	12401
CGC5		8	3	4	0	65624	82334
CGC6		3	2	1	0	142971	148385
CGC7		12	2	6	0	240595	257338
CGC8		2	1	1	0	1667	3655
CGC9		3	1	1	0	32391	35584

Showing 1 to 13 of 13 entries First Previous 1 Next Last

http://bcb.unl.edu/dbCAN2/blastation.php?jobid=20190527211827

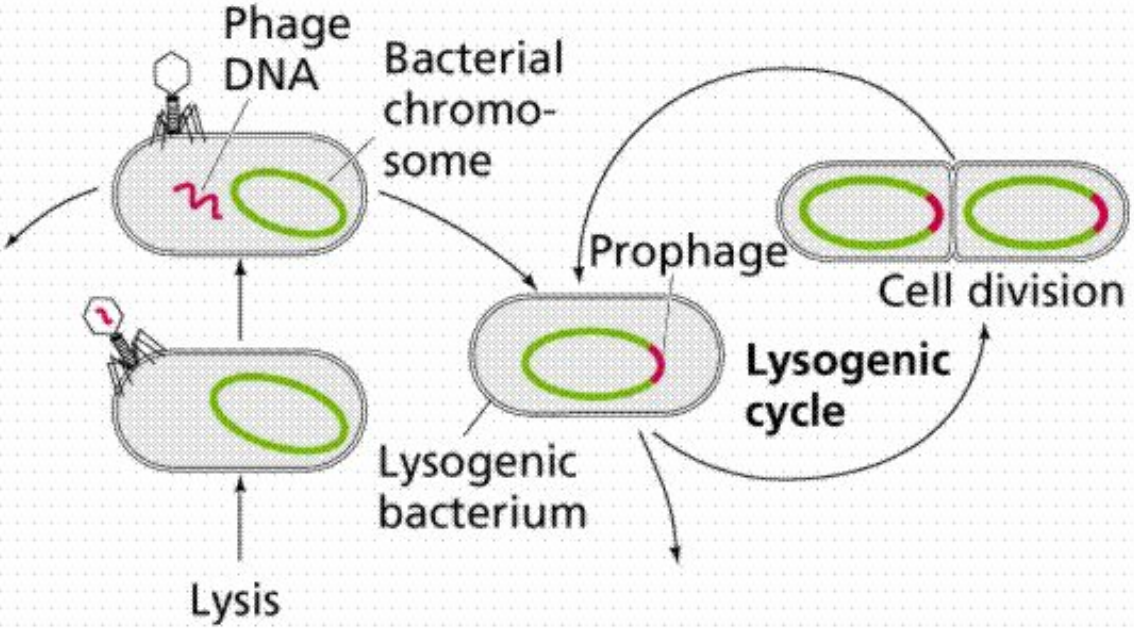
CyVerse Discovery Environment

Viral database used: **Viromes**

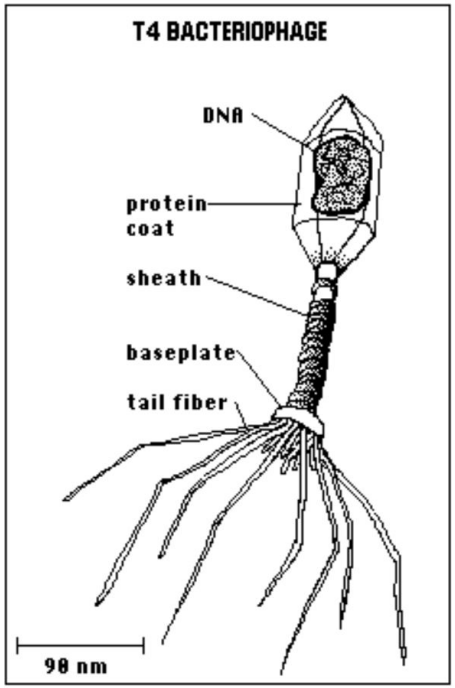
All bacterial and archaeal virus genomes in Refseq, plus non-redundant predicted genes from viral metagenomes

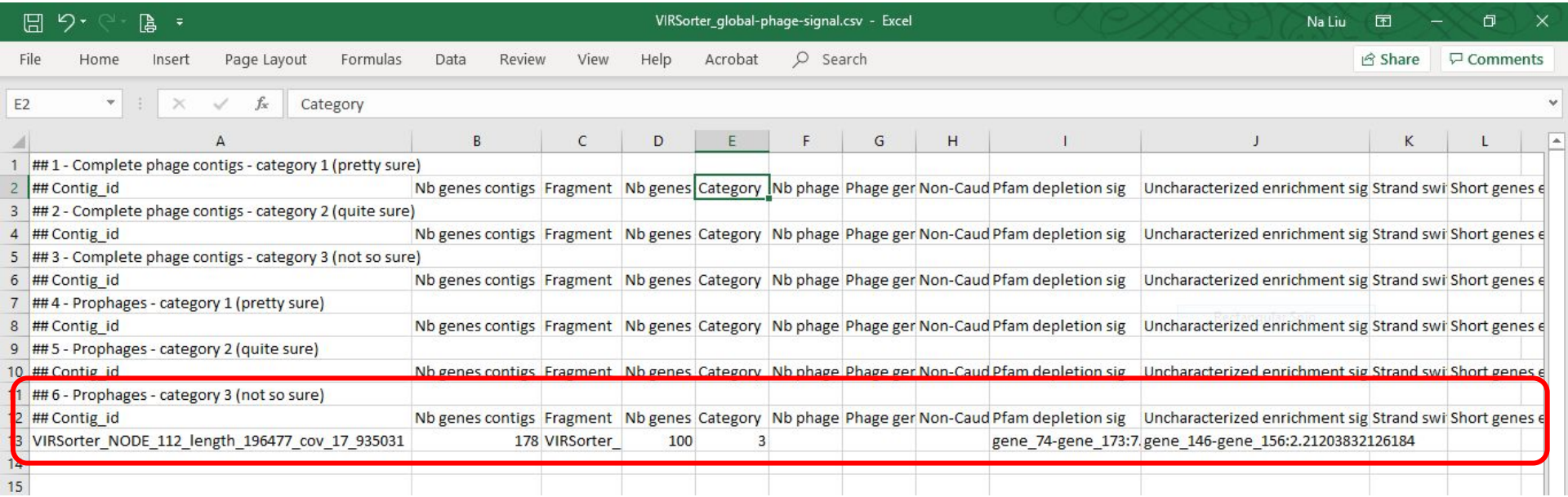
- Sequences **entirely viral**, from the more to less confident predictions (category 1, 2, and 3)
- **Prophages** (viral regions detected in a cellular contig), again from the more to less confident predictions (category 4, 5, and 6)

Roux *et al.*, *PeerJ*, 2015


Phage DNA
Bacterial chromo-some
Prophage
Cell division
Lysogenic cycle
Lysogenic bacterium
Lysis


Copyright 1996 ADAM (webmaster@adam.com)


T4 BACTERIOPHAGE
DNA
protein coat
sheath
baseplate
tail fiber
98 nm



	A	B	C	D	E	F	G	H	I	J	K	L
1	## 1 - Complete phage contigs - category 1 (pretty sure)											
2	## Contig_id	Nb genes contigs	Fragment	Nb genes	Category	Nb phage	Phage ger	Non-Caud	Pfam depletion sig	Uncharacterized enrichment sig	Strand swi	Short genes e
3	## 2 - Complete phage contigs - category 2 (quite sure)											
4	## Contig_id	Nb genes contigs	Fragment	Nb genes	Category	Nb phage	Phage ger	Non-Caud	Pfam depletion sig	Uncharacterized enrichment sig	Strand swi	Short genes e
5	## 3 - Complete phage contigs - category 3 (not so sure)											
6	## Contig_id	Nb genes contigs	Fragment	Nb genes	Category	Nb phage	Phage ger	Non-Caud	Pfam depletion sig	Uncharacterized enrichment sig	Strand swi	Short genes e
7	## 4 - Prophages - category 1 (pretty sure)											
8	## Contig_id	Nb genes contigs	Fragment	Nb genes	Category	Nb phage	Phage ger	Non-Caud	Pfam depletion sig	Uncharacterized enrichment sig	Strand swi	Short genes e
9	## 5 - Prophages - category 2 (quite sure)											
10	## Contig_id	Nb genes contigs	Fragment	Nb genes	Category	Nb phage	Phage ger	Non-Caud	Pfam depletion sig	Uncharacterized enrichment sig	Strand swi	Short genes e
11	## 6 - Prophages - category 3 (not so sure)											
12	## Contig_id	Nb genes contigs	Fragment	Nb genes	Category	Nb phage	Phage ger	Non-Caud	Pfam depletion sig	Uncharacterized enrichment sig	Strand swi	Short genes e
13	VIRSorter_NODE_112_length_196477_cov_17_935031	178	VIRSorter_	100	3				gene_74-gene_173:7.	gene_146-gene_156:2.21203832126184		
14												
15												

Prophage ranking of 3, **not so sure**

Conclusions

- *MAG 5* is a high-quality assembly: **99.4%** completeness and **0%** contamination
- *MAG 5* represents a bacterial genome in the phylum *Firmicutes*
- Closest relative on the phylogenetic tree: ***Selenomonas sputigena***
- No potential for antibiotics production
- **26 CAZyme genes** predicted by *dbCAN2* - sugar metabolism and glycosylation
- **1 not-so-sure prophage** predicted by *Virsorter*

References

1. Grass, S., Buscher, A. Z., Swords, W. E., Apicella, M. A., Barenkamp, S. J., Ozchlewski, N., and Geme, J. W. S. (2003). “The Haemophilus influenzae HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis.” *Molecular Microbiology*, 48(3), 737–751.
2. Naegeli, A., Neupert, C., Fan, Y.-Y., Lin, C.-W., Poljak, K., Papini, A. M., Schwarz, F., and Aebi, M. (2014). “Molecular Analysis of an Alternative N-Glycosylation Machinery by Functional Transfer from Actinobacillus pleuropneumoniae to Escherichia coli.” *Journal of Biological Chemistry*, 289(4), 2170–2179.
3. “Pfam: Family: Glyco_transf_41 (PF13844).” (n.d.). <http://pfam.xfam.org/family/Glyco_transf_41> (May 29, 2019).