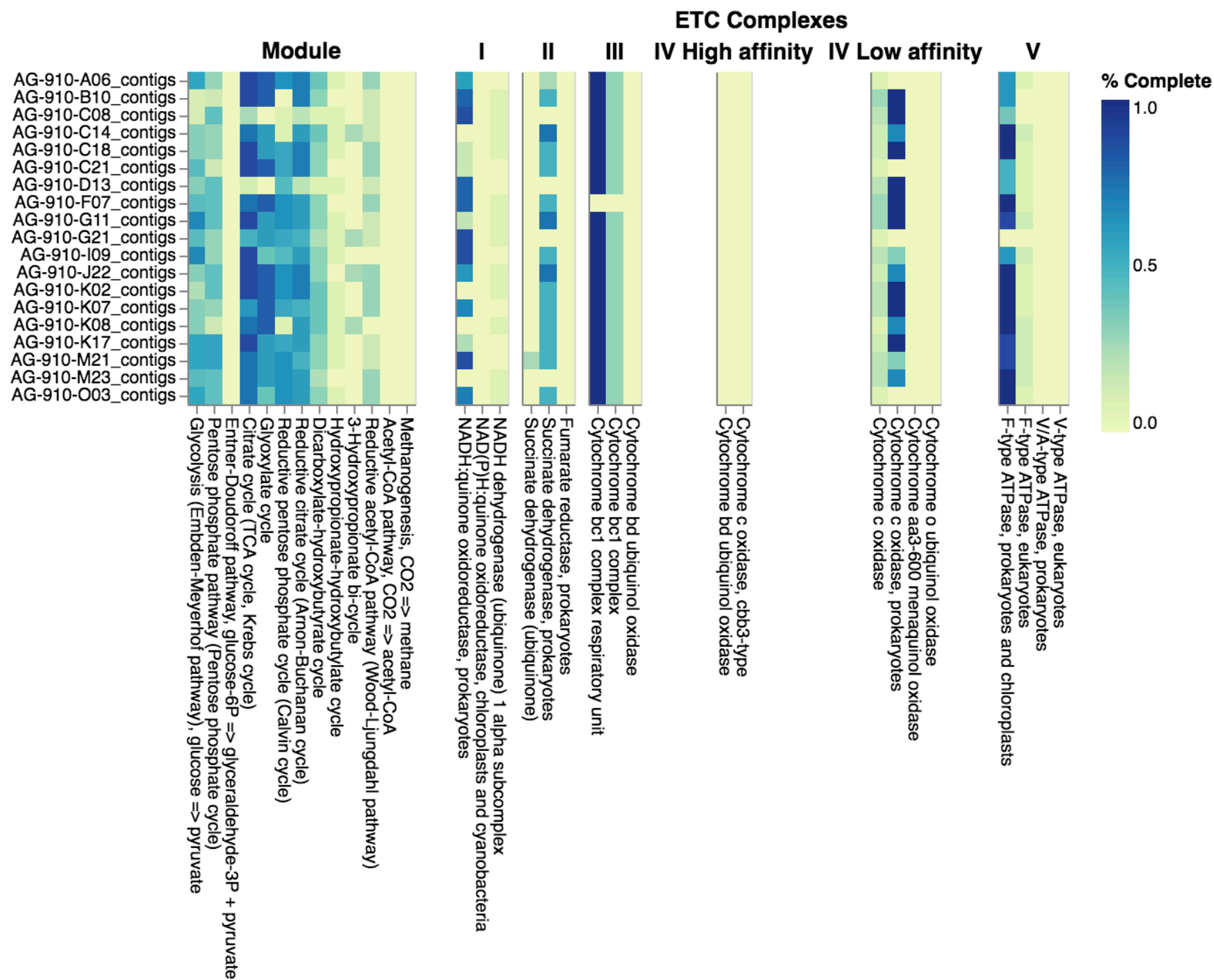# DRAM outputs

- First, let's go through the files you generated from the "annotate" and "distill" DRAM commands

- What are they, how can they be used? What will be most useful to you?

- Annotation summaries

- One example of visualization (made with DRAM in the "Distill" command) – we won't do a whole lot with manipulating this data, but will show you how to access what goes into a dataset like this!

A

Scaffer et al., 2020
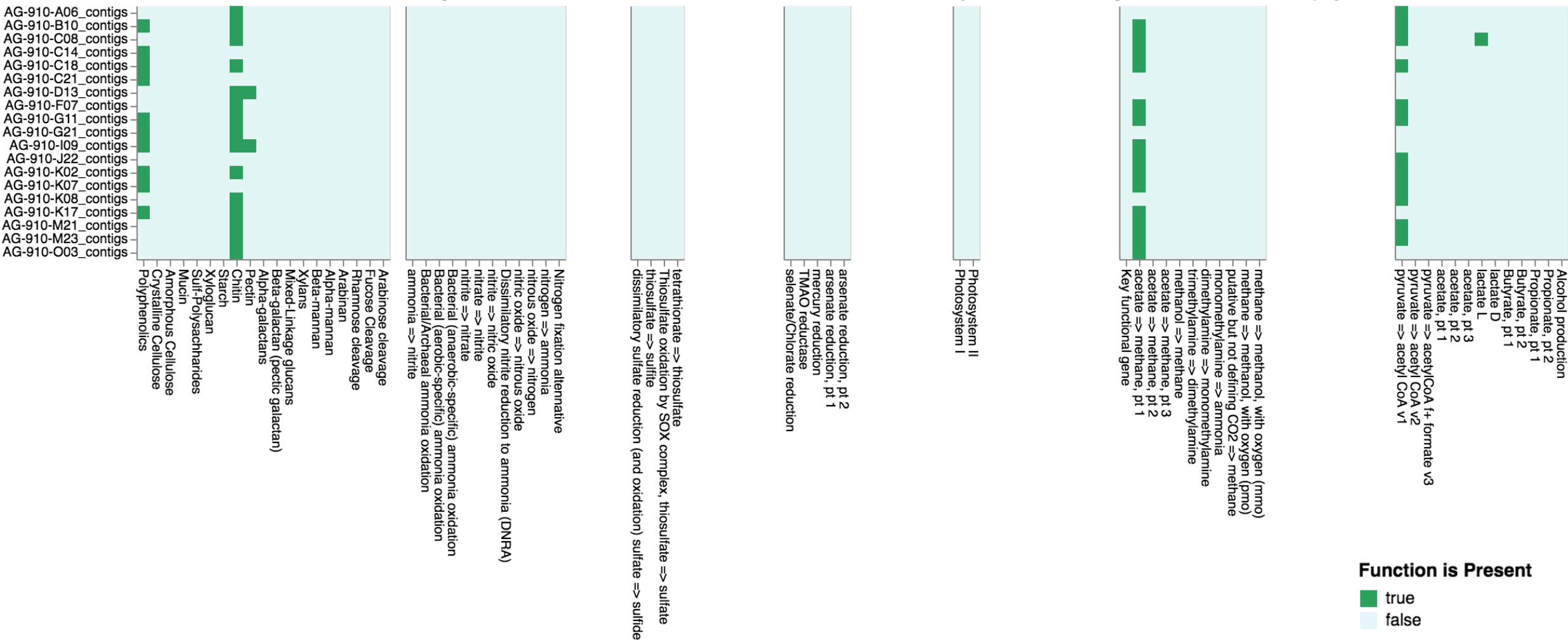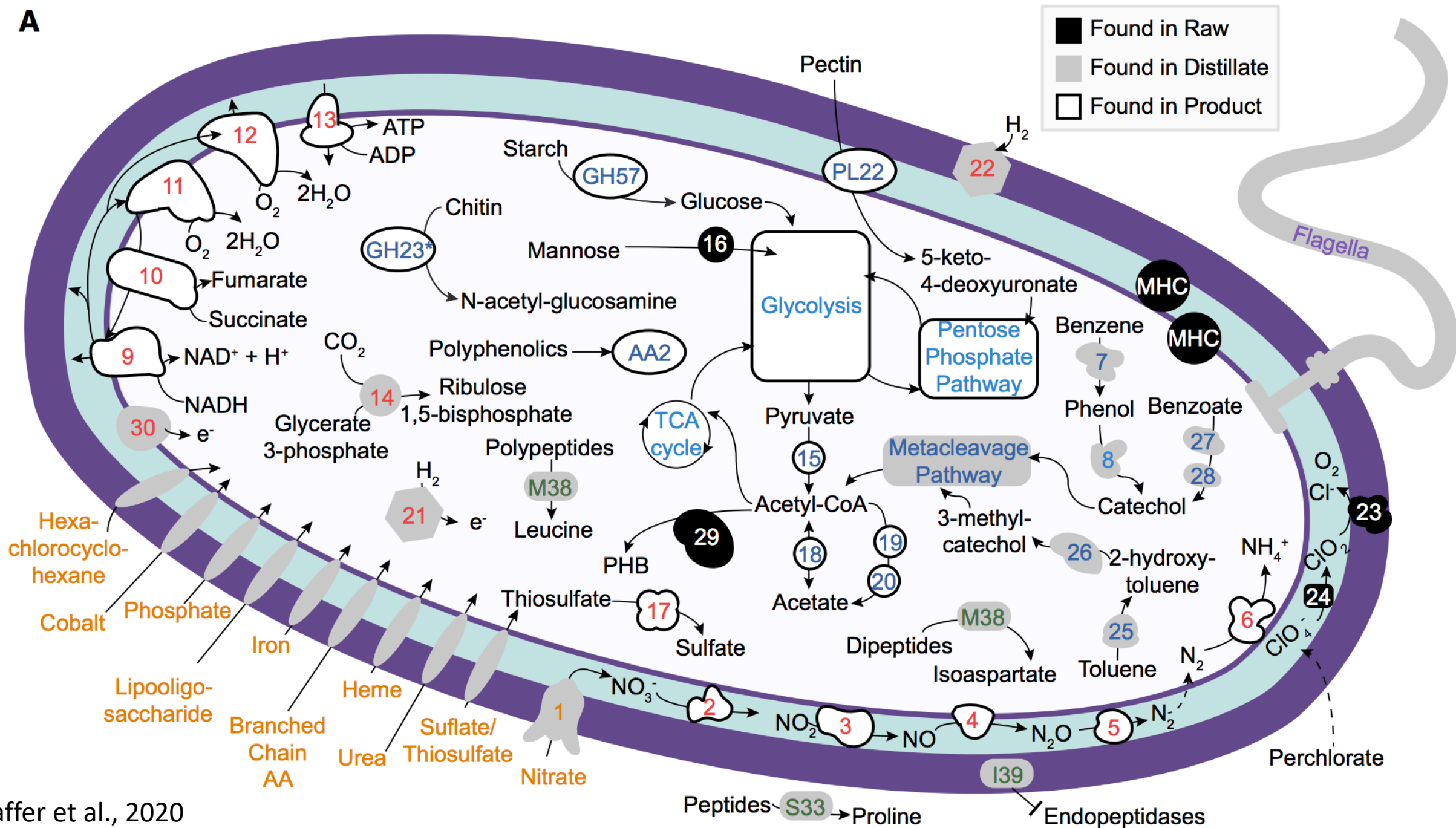
Misannotations, or genes involved in multiple pathways

- Rampant in databases, particularly as larger and larger sequence datasets are deposited using automated annotation programs
- How would misannotations become "the norm" for some genes?

- Do you know of any common misannotations in data from organisms you work with/might work with in the future?
- Common genes?

Notice anything suspicious?

How would you go about verifying?

Check databases (KEGG)

ORTHOLOGY: K01895

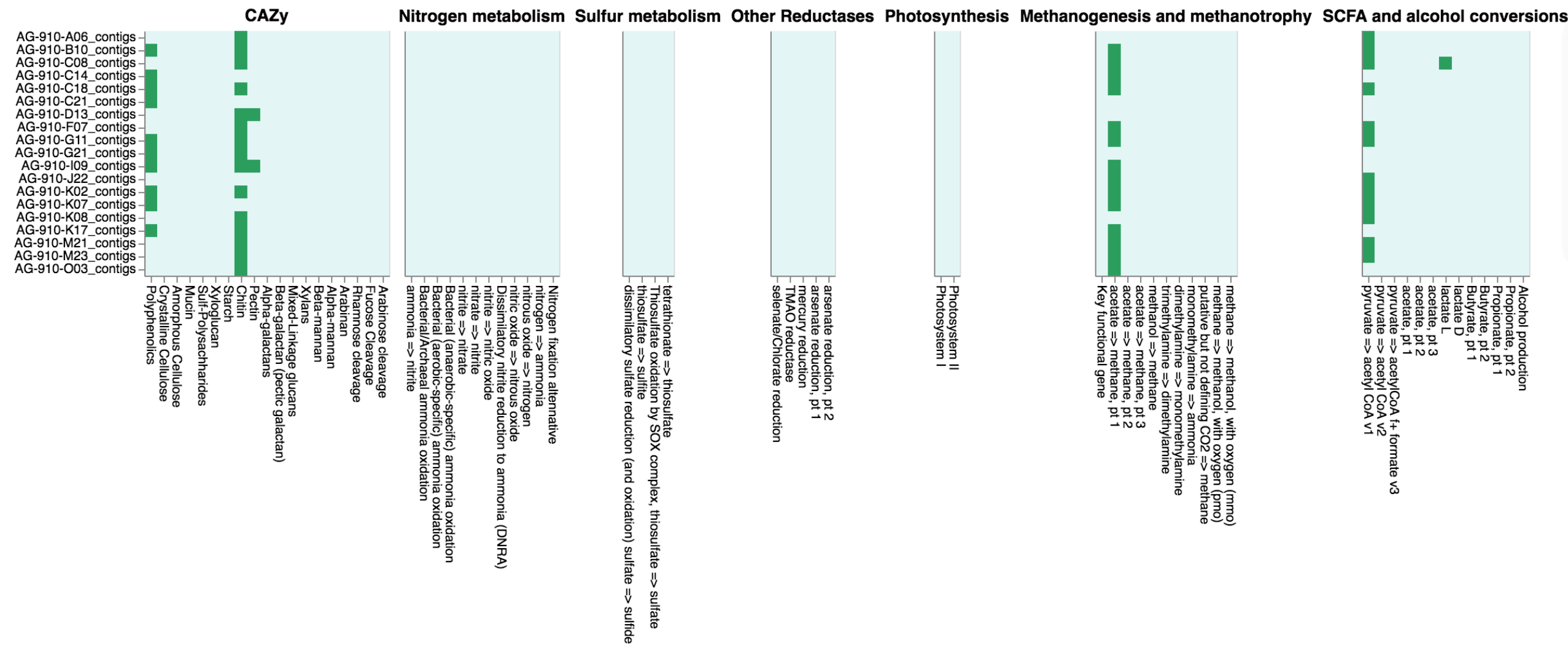| Entry	K01895 KO
Symbol	ACSS1_2, acs
Name	acetyl-CoA synthetase [EC:6.2.1.1]
Pathway	map00010 Glycolysis / Gluconeogenesis
	map00620 Pyruvate metabolism
	map00630 Glyoxylate and dicarboxylate metabolism
	map00640 Propanoate metabolism
	map00680 Methane metabolism
	map00720 Carbon fixation pathways in prokaryotes
	map01100 Metabolic pathways
	map01110 Biosynthesis of secondary metabolites
	map01120 Microbial metabolism in diverse environments
	map01200 Carbon metabolism
Module	M00357 Methanogenesis, acetate => methane

KEGG database

Other Misannotations?

- Gene involved in nitrogen assimilation – flagged as dissimilatory nitrate to ammonium

- Carbon fixation – TCA cycle, ATP citrate lyase involved in carbon fixation (appear to be TCA cycle, unless you blast with another ATP citrate lyase). **Challenge: how would you do this?**

- Arsenic reduction misannotation – one of the genes that is a regulation factor (actually in multiple pathways, need to verify that other genes are present in the arsenic reduction pathway!)

Other ways to check for potential misannotations? Or divergence?

- Do you have any commonly misannotated genes that you know of?

- Active sites for enzymes (do they possess key motifs)?
    - Example: cytochromes
    - Example: hydrogenases

- Databases of your choice
    - Examples: hydrogenases/HydDB, cytochrome P450 database

Hogendoorn et al., 2020

- Example: Cysteine residues key for binding site of hydrogenases (for [NiFe]-hydrogenases, need paired CxxC motifs)

Other ways to check for potential misannotations? Or divergence?

Lindsay et al., in prep
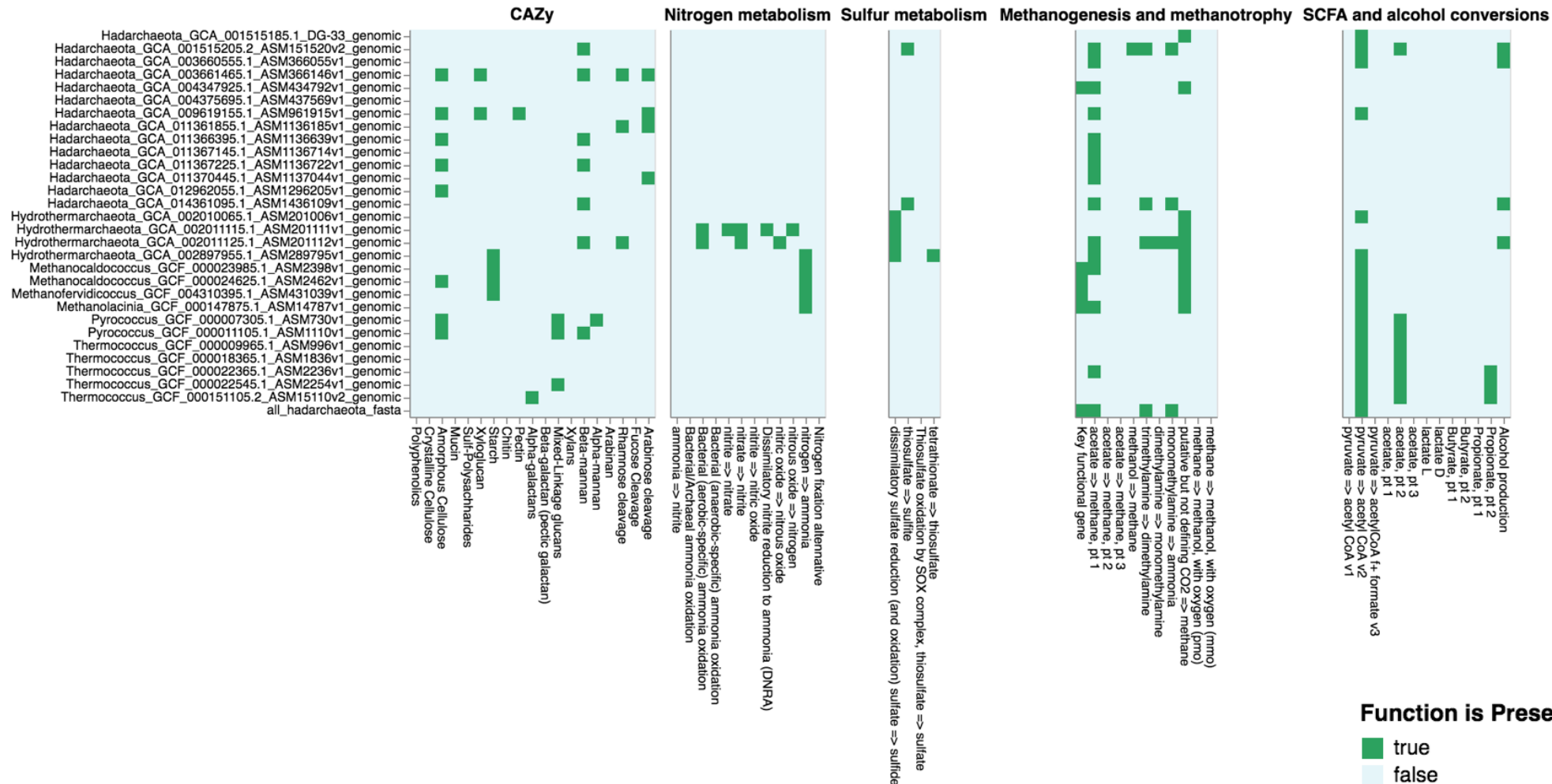
	A	B	C	D	G	H	I	J	K	L	M
1	gene_id	gene_description	module	header	294-B02_co	294-C15_co	294-C19_co	294-C23_co	294-D03_co	294-D11_co	294-D14_co
174	K00320	coenzyme F420-dependent N5,N10-methenyltetrahydrometham	Methanogenesis, CO2 => methane	C1-methane	0	0	0	0	0	0	0
175	K00399	methyl-coenzyme M reductase [EC:2.8.4.1] [RN:R04541]	Methanogenesis, CO2 => methane	C1-methane	0	0	0	0	0	0	0
176	K00401	methyl-coenzyme M reductase [EC:2.8.4.1] [RN:R04541]	Methanogenesis, CO2 => methane	C1-methane	1	0	0	0	1	1	0
177	K00402	methyl-coenzyme M reductase [EC:2.8.4.1] [RN:R04541]	Methanogenesis, CO2 => methane	C1-methane	0	0	0	0	0	0	0
178	K00577	tetrahydromethanopterin S-methyltransferase [EC:2.1.1.86] [RN	Methanogenesis, CO2 => methane	C1-methane	0	0	0	0	0	0	0

Tree scale: 1000

Hadarchaeota GCA 004347925.1 ASM434792v1 genomic QNVD01000218.1 6
BLM1_Hadarchaeota_AM-294-P03 NODE 15 2
BLM1_Hadarchaeota_AM-294-O15 NODE 21 3
BLM1_Hadarchaeota_AM-294-L20 NODE 19 3
BLM1_Hadarchaeota_AM-294-L18 NODE 26 4
BLM1_Hadarchaeota_AM-294-L08 NODE 14 4
BLM1_Hadarchaeota_AM-294-L04 NODE 14 3
BLM1_Hadarchaeota_AM-294-K10 NODE 2 13
BLM1_Hadarchaeota_AM-294-I22 NODE 37 3
BLM1_Hadarchaeota_AM-294-G07 NODE 22 2
BLM1_Hadarchaeota_AM-294-G05 NODE 14 3
BLM1_Hadarchaeota_AM-294-F21 NODE 14 4
BLM1_Hadarchaeota_AM-294-D11 NODE 31 2
BLM1_Hadarchaeota_AM-294-B02 NODE 33 3
BLM1_Hadarchaeota_AM-294-D03 NODE 23 4
Methanocaldococcus_jannaschii_Q58252
Ca.Methanodesulfokores_washburnensis_A0A3R9PGH0
Ca.Korarchaeota_archaeon_A0A523BJ61
Methanolacinia_petrolearia_GCF000147875
Methanosarcina_barkeri_P07955
Methanosarcinales_archaeon_A0A849QU72
Methanocella_sp.PtaU1_A0A1V4YWG9
Methanocaldococcus_vulcanius_GCF000024625
Methanocaldococcus_fervens_GCF000023985
Methanocaldococcus_vulcanius_GCF000024625
Methanolacinia_petrolearia_GCF000147875
Methanothermus_fervidus_P12972
Methanofervidicoccus_abyssi_GCF004310395
Methanococcus_vannielii_P07956
Methanococcus_voltae_P11561
Methanobacterium_sp.A39_A0A1D2W9L5
Methanobacterium_bryantii_A0A2A2H970
Methanothermus_fervidus_Q49171
Methanothermobacter_marburgensis_D9PXZ6
Methanothermobacter_sp.THM-2_A0A6B9TIS7
Methanothermobacter_wolfeii_A0A1C7D1E4
Methanothermobacter_sp.THM-1_A0A6B9TFZ0
Methanothermobacter_thermautotrophicus_A0A842YRG5
Methanothermobacter_thermautotrophicus_P21111
Methanothermobacter_sp_A0A842LRM8
Methanothermobacter_thermautotrophicus_A0A832RU37

mcrB tree

Lindsay et al., in prep

KAAS - KEGG Automatic Annotation Server
for ortholog assignment and pathway mapping

METHANE METABOLISM

Complete pathways?

Genome.jp