Active sulfur and carbon cycling in deep Greenland groundwaters

Sophie L Nixon[1], Malin Bomberg[2], Lillemor Claesson Liljedahl[3], Tiina Lamminmäki[4], Anne Kontula[4]

Bioinformatic Working Group:
Using Big Data Approaches to Understand Microbial Communities
10.02.22

The University of Manchester

SCIENCE OF THE ENVIRONMENT

Greenland deep biosphere

Spin out from **Greenland**

- nr Kangarlussuaq, We
- 1800-1900 Ma Archae
- Candidate location for
- 2 deep boreholes stud
  - DH-GAP01 **191m**
  - DH-GAP04 **649m**

The Greenland Analogue Project: Final report

Claesson Liljedahl L[1], Kontula A[2], Harper J[3], Näslund J-O[1], Selroos J-O[1], Pitkänen P[2], Puigdomenech I[1], Hobbs M[4], Follin S[5], Hirschorn S[4], Jansson P[6], Kennell L[4], Marcos N[7], Ruskeeniemi T[8], Tullborg E-L[9], Vidstrand P[1]

1 Svensk Kärnbränslehantering AB
2 Posiva Oy
3 University of Montana
4 Nuclear Waste Management Organization (NWMO)
5 Golder associates
6 Department of Physical Geography, Stockholm University
7 Saanio & Riekkola Oy
8 Geological Survey of Finland
9 Terralogica AB

Svensk Kärnbränslehantering AB
Swedish Nuclear Fuel
and Waste Management Co
Box 250, SE-101 24 Stockholm
Phone +46 8 459 84 00

August 2016

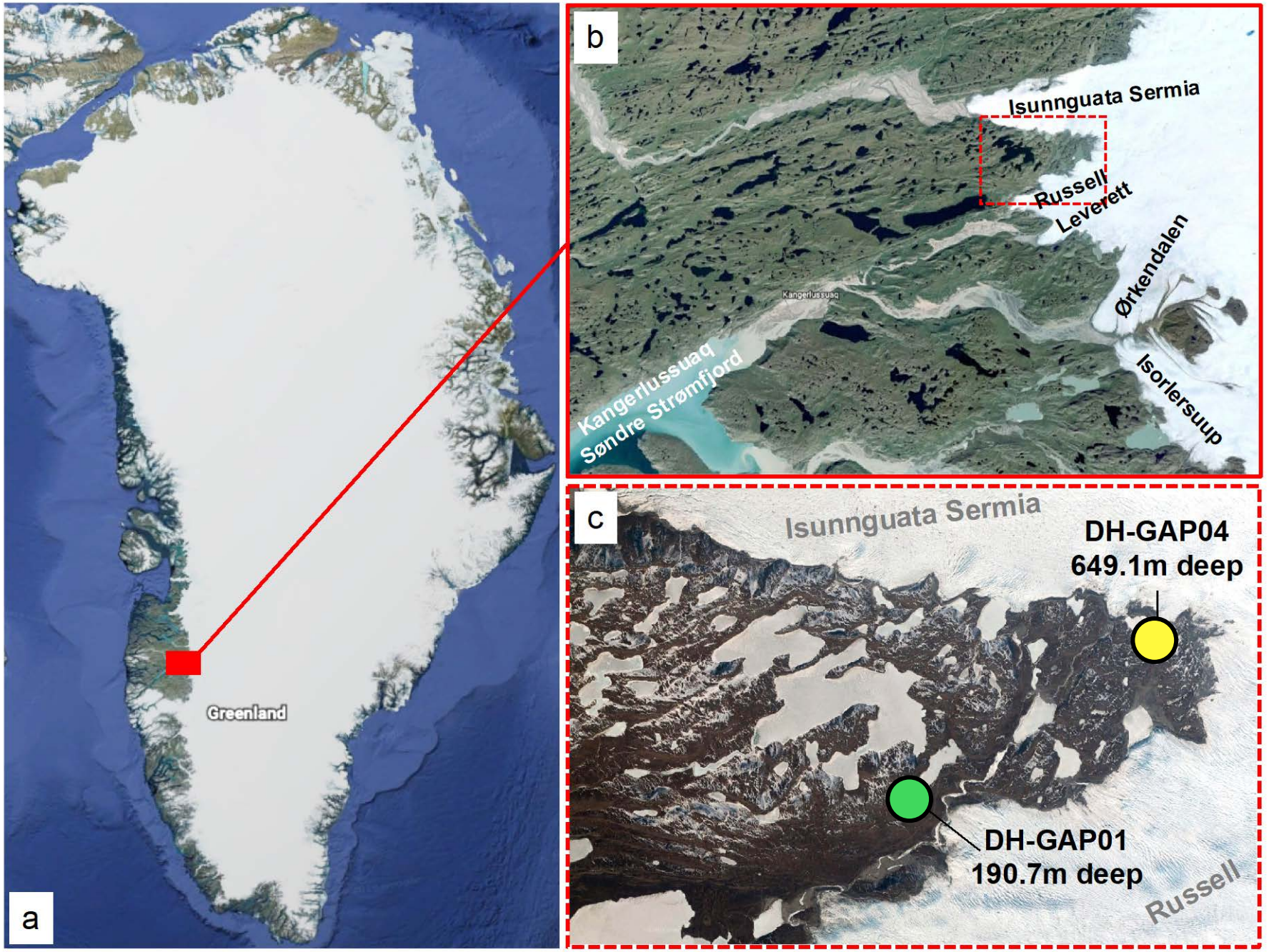
ORIGINAL RESEARCH ARTICLE
Front. Microbiol., 11 July 2019 | https://doi.org/1

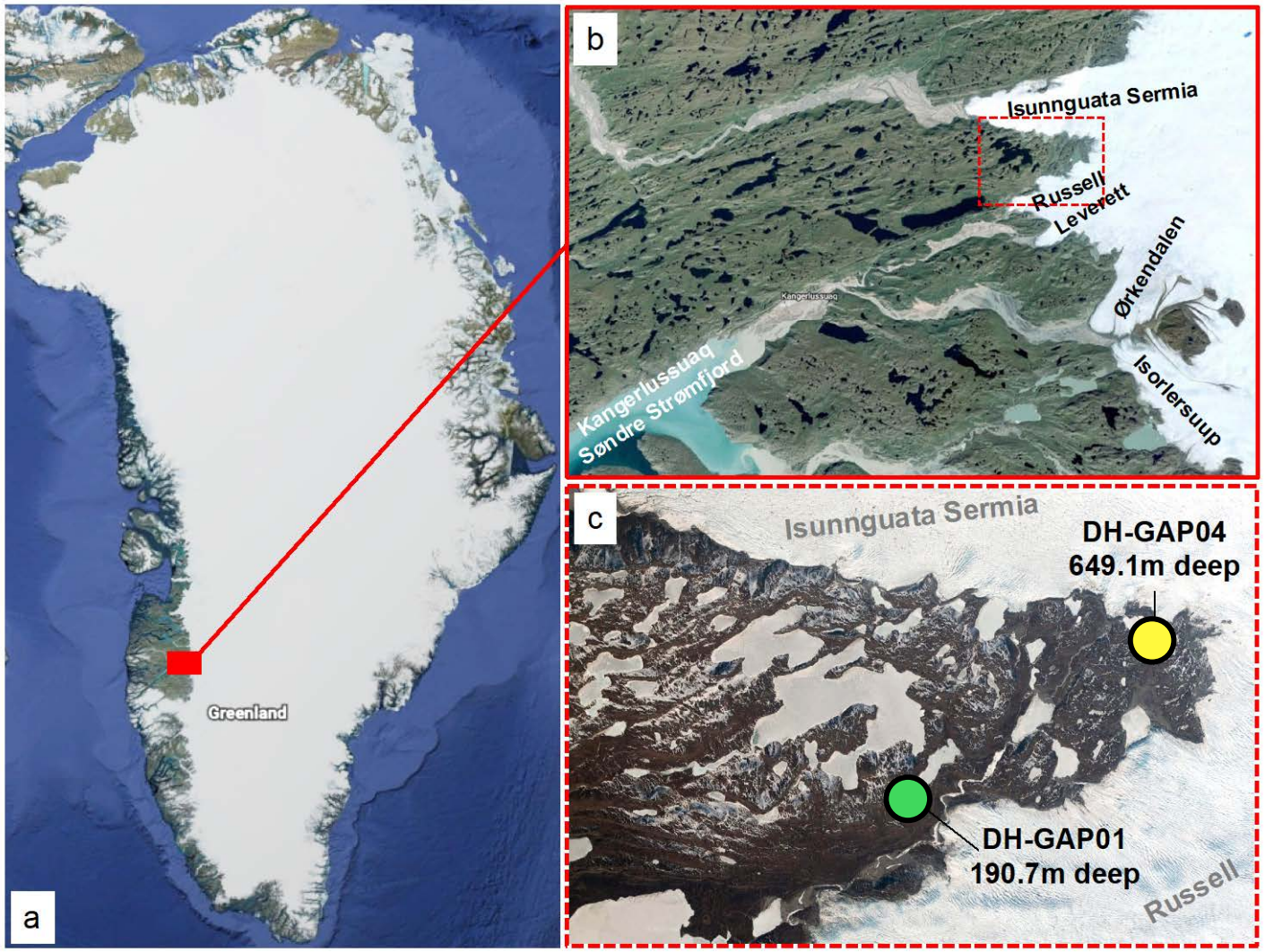
Highly Diverse A
Separated by Per
Distinct Features
Niches

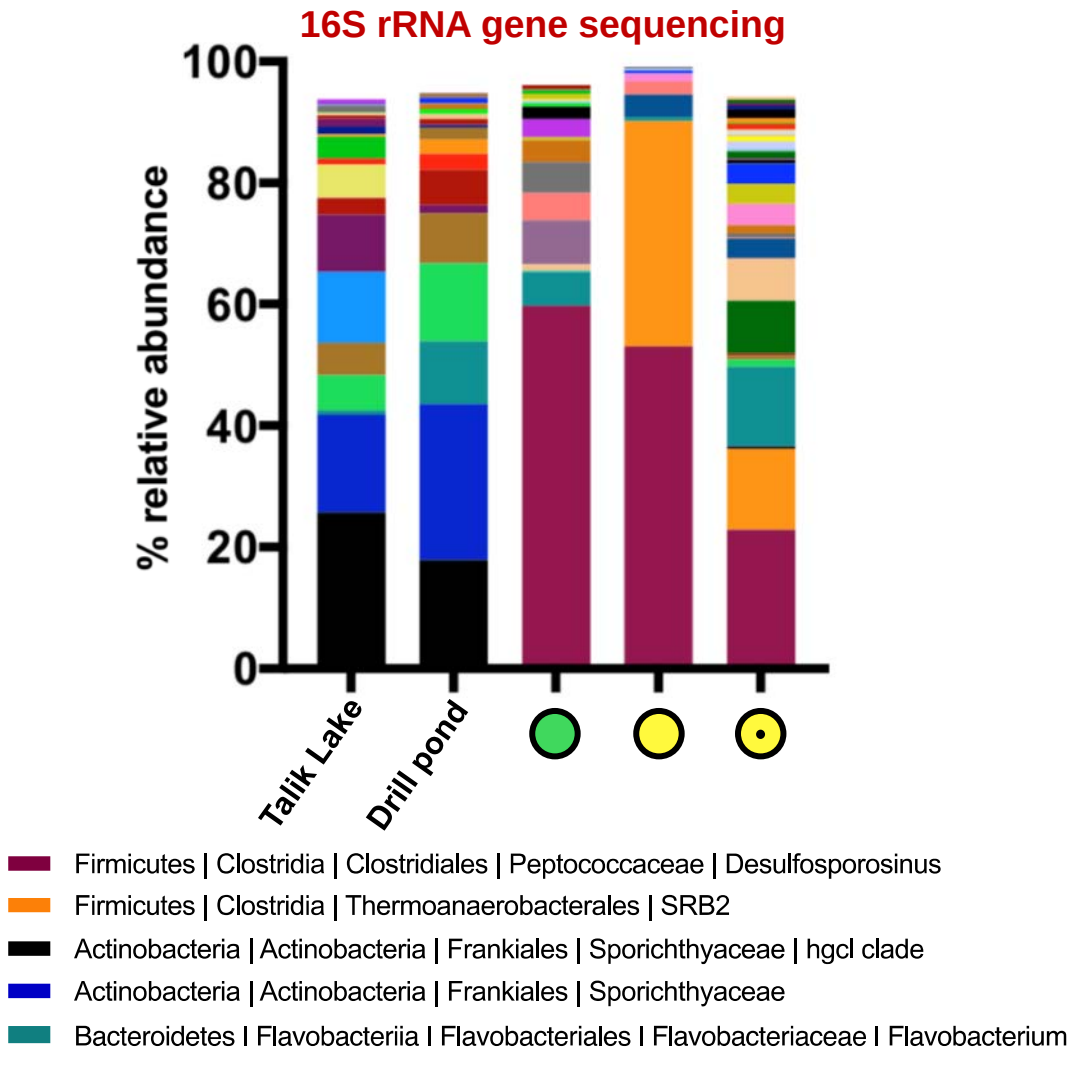
Malin Bomberg[1*], Lillemor Claesson Liljedahl[2], Tiina Lamminmäki[3] and Anne Kontula[3]

Why study it?

- Subsurface harbours '**unseen majority**' but diversity and function not well understood
- Deep biosphere is an **extreme environment** – understanding life in extremes helps define limits of life, and search for life elsewhere
- Extremely challenging to access – **makes use of existing boreholes**







	Depth (m)	Volume (L)							
									0.4
DH-GAP04-LOW									

(Bomberg *et al.*, 2019)

Methods

extraction

sequencing

bioinformatics

Talik Lake
Drill pond

DH-GAP01
DH-GAP04

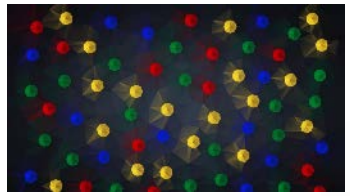
DH-GAP04

DNA

2x metaG libraries per sample

RNA

all RNA
&
rRNA removed



metaG

metaT

Who's there and what can they do?

- *metaG*
- *functional potential*

Who's active and what are they doing?

- *metaT*
- *genes expressed*

Key biogeochemical processes?

- *metaGand metaT*
- *pathways encoded / expressed*

Viral interactions?

- *metaG and metaT*
- *phage and prophage*
- *CRISPRs*

Bioinformatics

R reads-based
A assembly-based
G genome-based

R1.fastq
R2.fastq

quality check & read trimming — FastQC & Trimmomatic[R]

de novo assembly — IDBA-UD (metaG), metaSPAes (metaT)

binning — manual (Quicklooks[RA]), Maxbin 2.0[RA], checkM[G]

annotation — Prokka[G], VirSorter[G], Dram[A] CRISPR recognition tool[G]

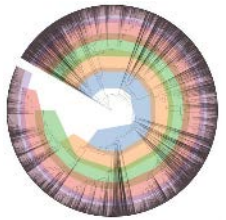
analysis of whole community

annotation

- Prokka[A] Dram[A]
- Kaiju[R]
- VirSorter[A]
- CRASS[R]

KEGG KAAS[AG]

metabolism *(what?)* — **phylogeny *(who?)*** — GTDB-tk[G]



analysis of individual genomes

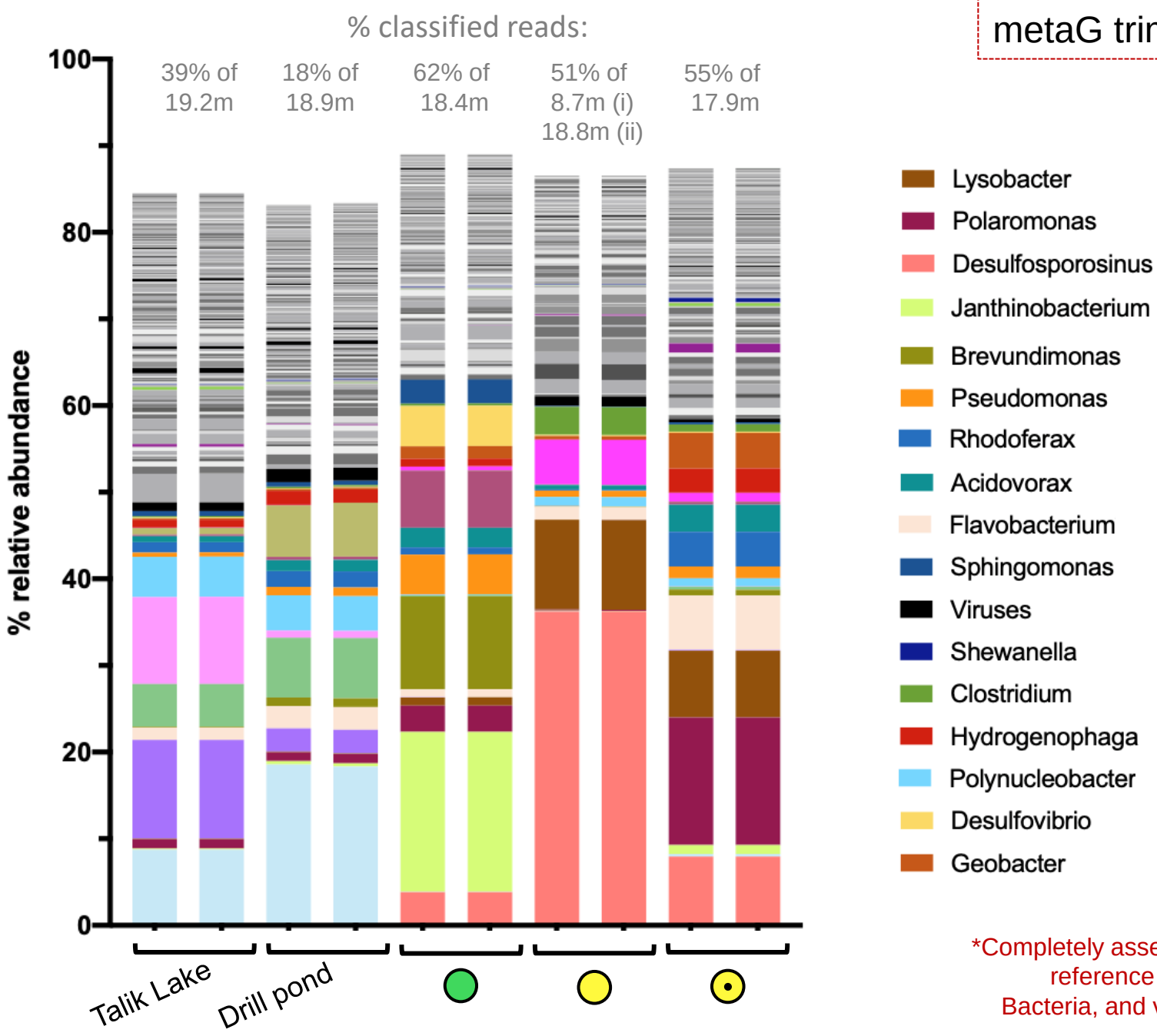
MRC Cloud Infrastructure for Microbial Bioinformatics

Who's there?

KAIJU

with refseq database*

- **Highly diverse** communities in all samples
- **Distinct composition** in above vs below permafrost samples



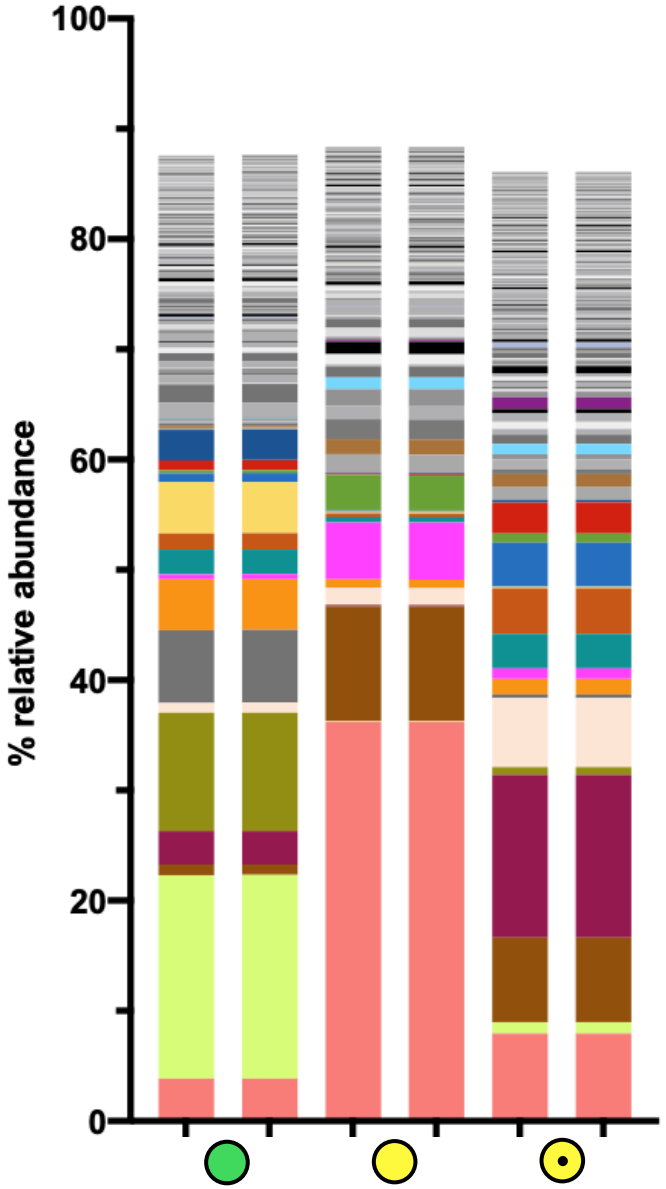
metaG trimmed reads

Genus-level taxonomic assignment of trimmed metagenomic reads (Kaiju). Only taxa >0.3% across all samples are shown

*Completely assembled and annotated reference genomes of Archaea, Bacteria, and viruses from the NCBI RefSeq database.

Who's there?

KAIJU

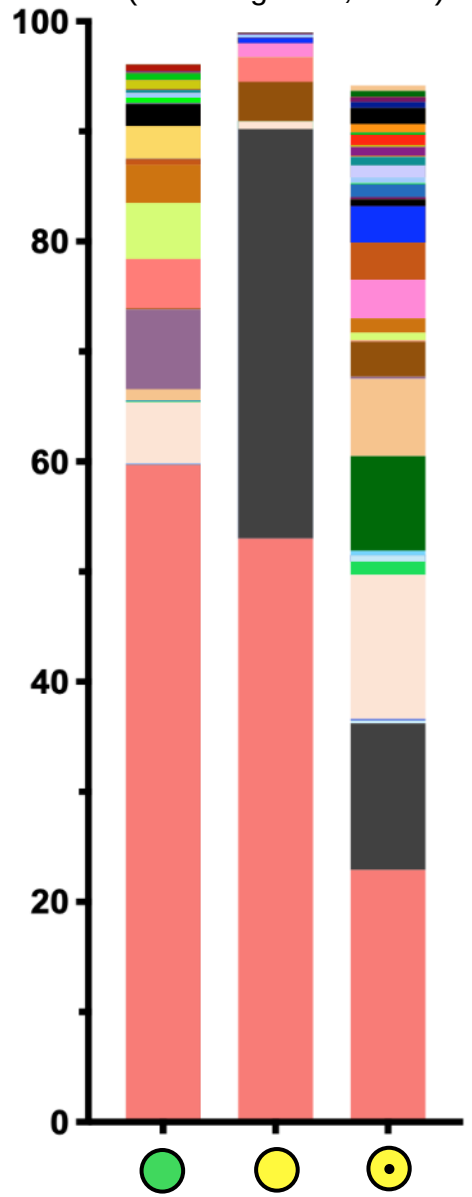


- Significantly **more diversity in metaG** vs 16S rRNA seq
- **Desulfosporosinus less dominant** than in published 16S rRNA gene seq data (but different DNA extracts)

Polaromonas
Desulfosporosinus
Janthinobacterium
Brevundimonas
Pseudomonas
Rhodoferax
Acidovorax
Flavobacterium
Sphingomonas
Viruses
Shewanella
Clostridium
Hydrogenophaga
Polynucleobacter
Desulfovibrio
Geobacter
Firmicutes | SRB2

metaG trimmed reads vs published 16S rRNA gene seq

(Bomberg *et al.*, 2019)

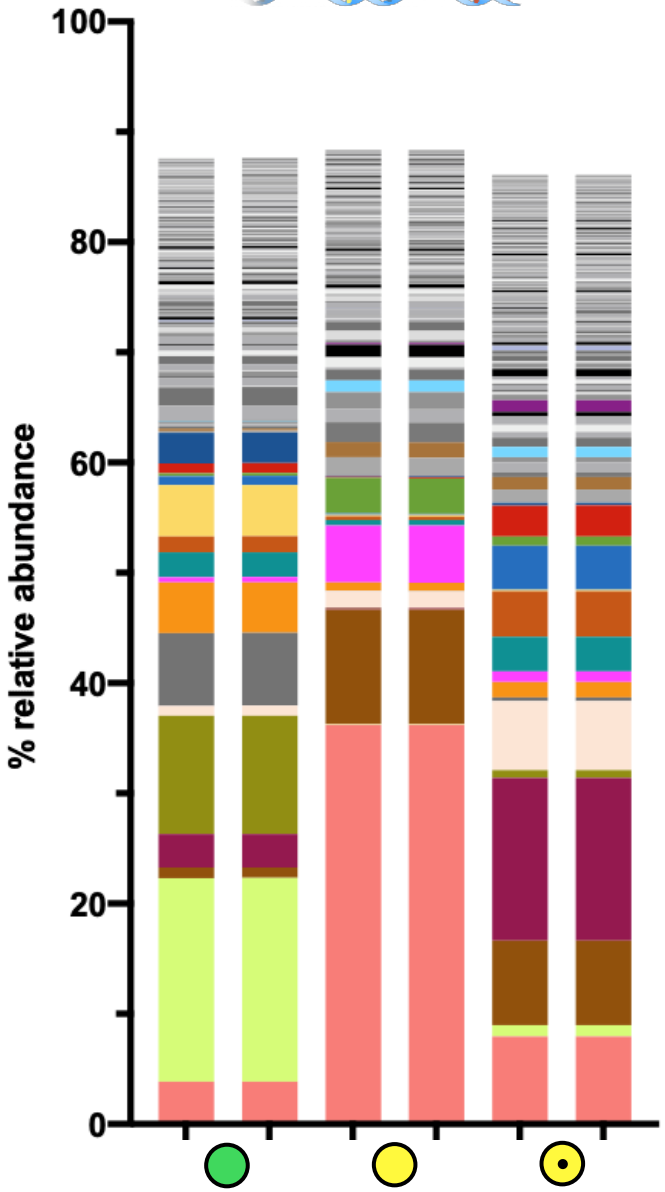


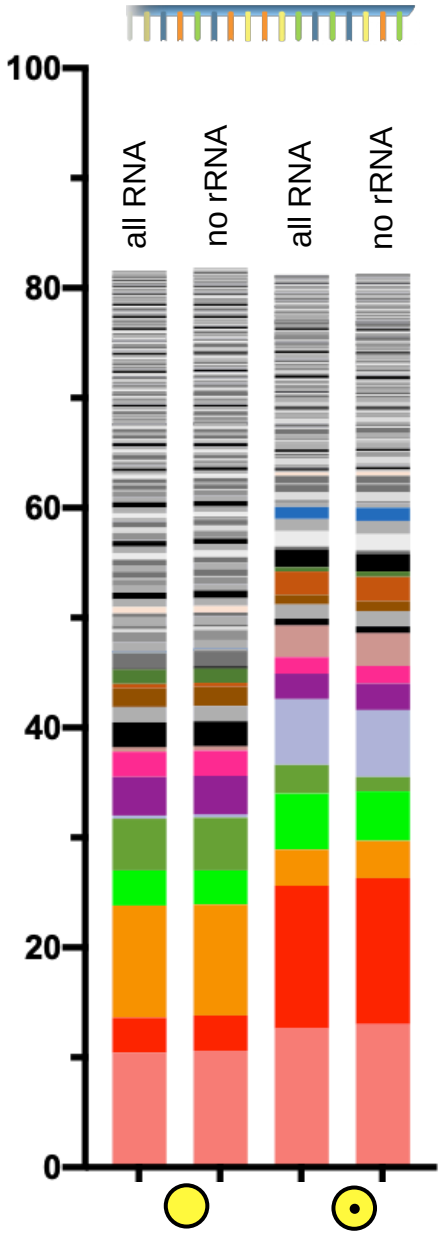
Who's active?

KAIJU

- Active communities are **highly diverse**
- **Low abundance taxa** appear to be **highly active** (Legionella (!), Bacillus, Lacunisphaera...)

metaG vs metaT trimmed reads





Community functional potential

KEGG-annotated assembled metaG

	C fixation (Wood Ljungdhal)	N fixation (*nifDKH*)	NO_3^- reduction	NO_2^- reduction	SO_4^{2-} reduction	Thiosulfate reduction (SOX)
191 m	✓	✓	✓	✓	✓	✓
561 – 571 m	✓	✓	✓	✓	✓	✓
571 – 600 m	✓	✓	✓	✓	✓	✓
Talik Lake	✗	✓	✓	✓	✗	✓
Drill pond	✗	✓	✗	✓	✓	✗

Community activity

KEGG-annotated assembled metaT

	C fixation (Wood Ljungdhal)	**N fixation** (*nifDKH*)	**NO_3^- reduction**	**NO_2^- reduction**	**SO_4^{2-} reduction**	**Thiosulfate reduction** (SOX)
561 – 571 m	✓	✗	✓	✓	✓	✓
571 – 600 m	✗	✓	✗	✓	✓	✓

- **Key genes / pathways expressed** in one or both deep borehole sample
- **Active C, N and S cycling**

Community functional potential

assembled metaG

DRAM

		C fixation (Wood Ljungdhal)	N fixation (*nifDKH*)	NO_3^- reduction	NO_2^- reduction	SO_4^{2-} reduction	Thiosulfate reduction (SOX)
191 m	100%	✓	✓	✓	✓	✓	✓
561 – 571 m	86%	✓	✓	✗	✓	✓	✓
571 – 600 m	100%	✓	✓	✓	✓	✓	✓
Talik Lake	43%	✗	✗	✗	✗	✗	✓
Drill pond	43%	✗	✗	✗	✗	✗	✓

Community activity

assembled metaT

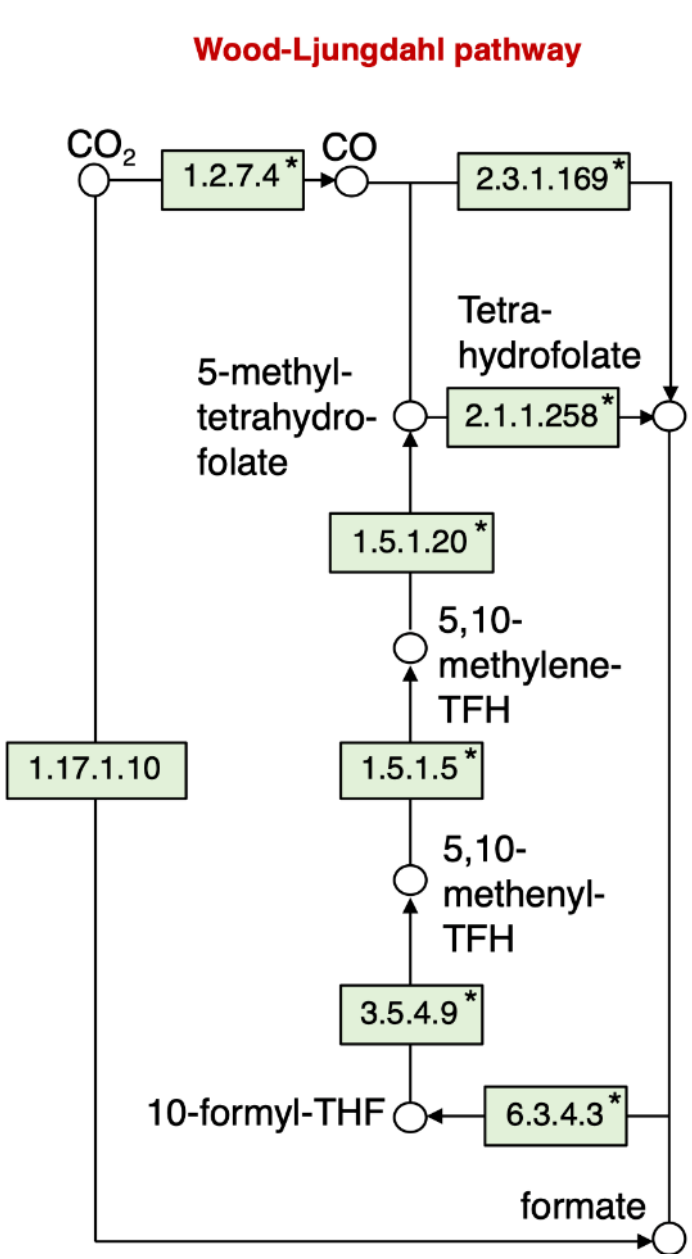
DRAM

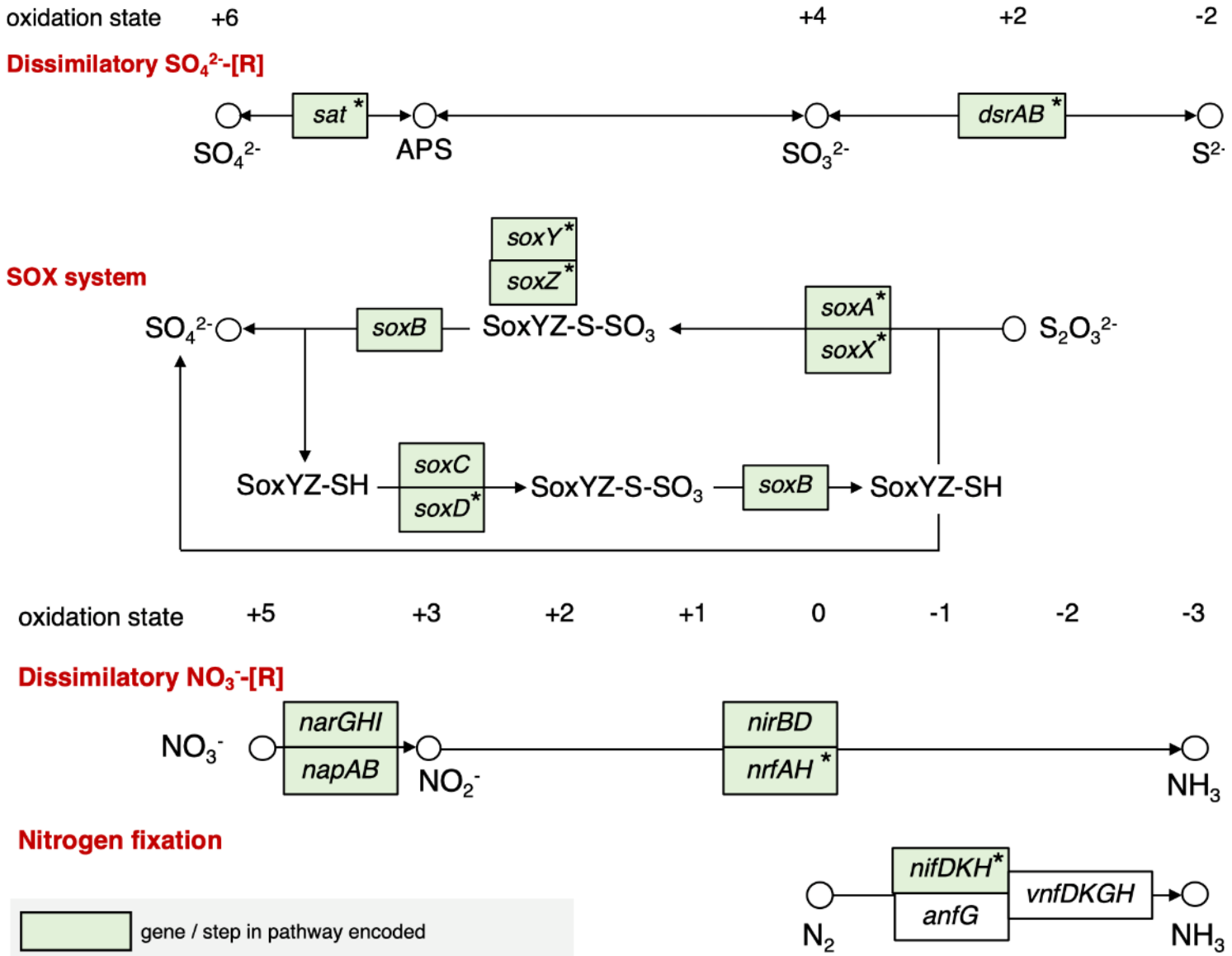
		C fixation (Wood Ljungdhal)	**N fixation** (*nifDKH*)	**NO_3^- reduction**	**NO_2^- reduction**	**SO_4^{2-} reduction**	**Thiosulfate reduction** (SOX)
561 – 571 m	43-71%	✗	✗	✗	✓	✗
571 – 600 m	29%	✗	✗	✗	✓	✗

- **Key genes / pathways expressed** in one or both deep borehole sample
- **Active C and S cycling**

Expressed biogeochemical pathways

KEGG-annotated assembled metaGT

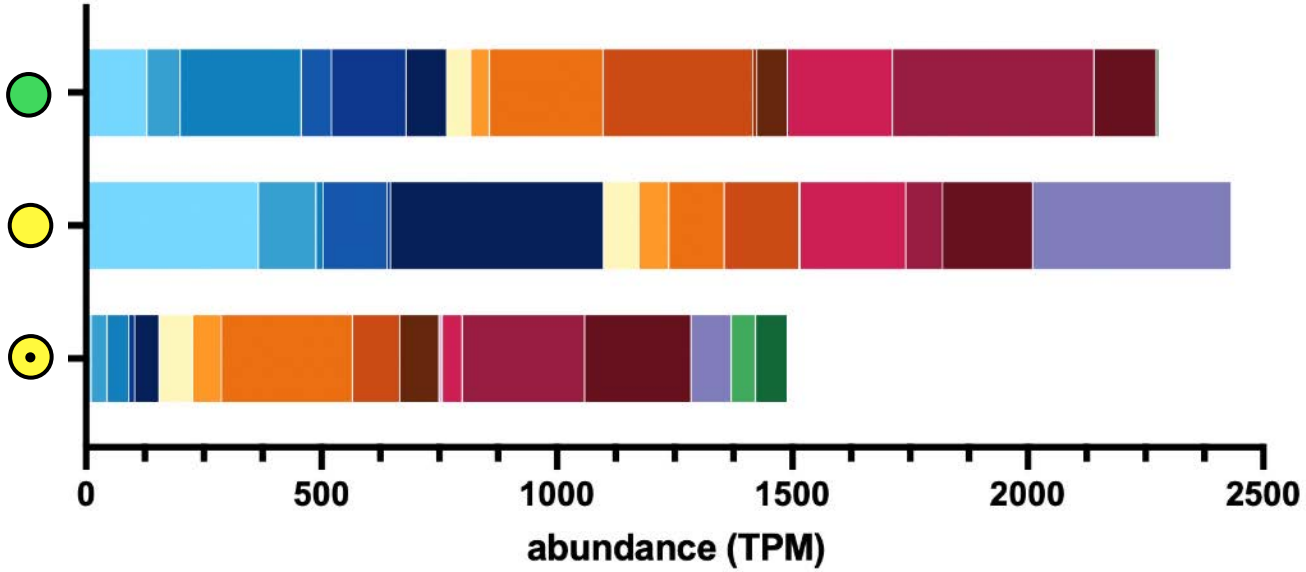




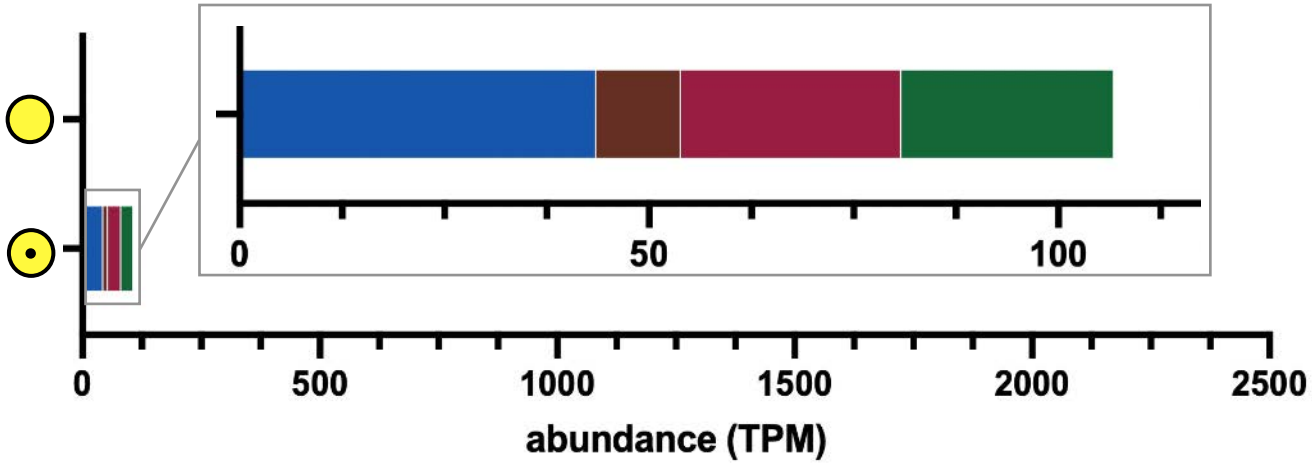
Carbon fixation

metaGT reads mapped to Prokka-annotated assembled metaGT

Encoded



Expressed



Reductive acetyl-CoA (Wood-Ljundahl) pathway

- anaerobic carbon-monoxide dehydrogenase*
- acetyl-CoA synthase*
- formate dehydrogenase (NADP+)*
- **formate dehydrogenase (hydrogenase)**
- methylenetetrahydrofolate reductase (NADH)*
- 5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase*

Hydroxypropionate and hydroxybutyrate cycles

- malonyl-CoA reductase / 3-hydroxypropionate dehydrogenase (NADP+)*
- 3-hydroxypropionyl-coenzyme A synthetase
- **3-hydroxypropionyl-coenzyme A dehydratase**
- acryloyl-coenzyme A reductase*
- propionyl-CoA carboxylase*
- succinyl-CoA:(S)-malate CoA-transferase*
- 2-methylfumaryl-CoA isomerase
- 3-methylfumaryl-CoA hydratase*
- succinyl-CoA reductase
- succinate semialdehyde reductase (NADPH)
- 4-hydroxybutyrate-CoA ligase
- 4-hydroxybutyryl-CoA dehydratase*
- **enoyl-CoA hydratase***
- 3-hydroxyacyl-CoA dehydrogenase*
- pyruvate carboxylase
- NADH-dependent fumarate reductase
- citryl-CoA synthetase
- citryl-CoA lyase

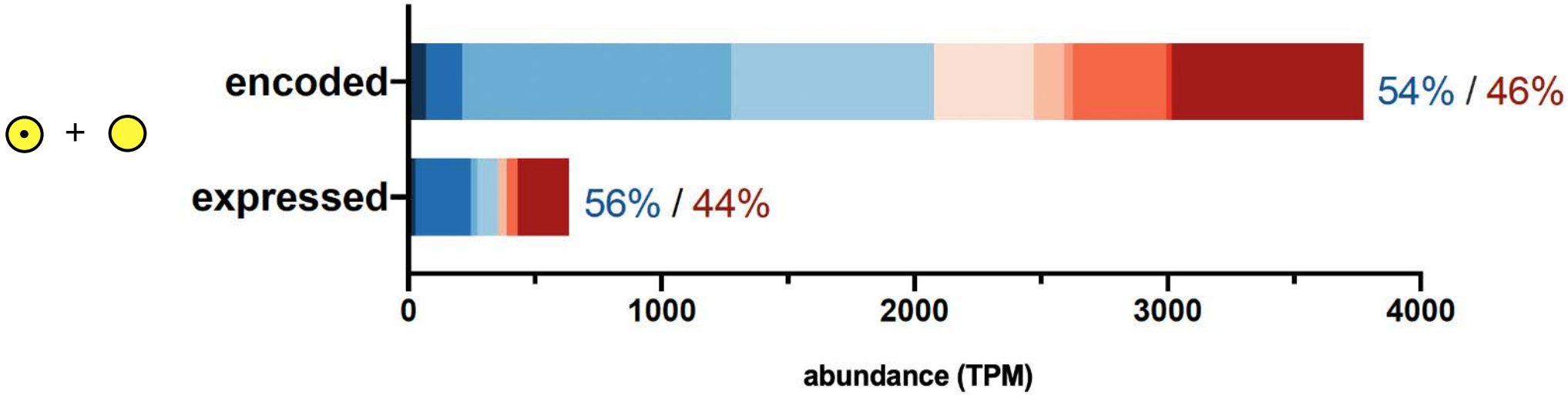
Calvin cycle

- glyceraldehyde-3-phosphate dehydrogenase (NADP+)
- phosphoribulokinase*
- **ribulose-bisphosphate carboxylase (RuBisCO)***

*encoded in all metaGs
absent in all metaGs
expressed in metaT

Net carbon fixation?

metaGT reads mapped to Prokka-annotated assembled metaGT



CO_2-fixing enzymes

- **ribulose-1,5-bisphosphate carboxylase**
- **phosphoenolpyruvate carboxylase**
- **2-oxoglutarate synthase**
- **acetyl-CoA carboxylase**

CO_2-producing enzymes

- acetolactate decarboxylase
- **acetolactate synthase**
- pyruvate decarboxylase
- **formate dehydrogenase**
- 6-phosphogluconate dehydrogenase
- **2-oxoglutarate dehydrogenase**
- pyruvate dehydrogenase

encoded
expressed

High quality MAGs
(Metagenome-assembled genomes)

manually binned metaG assemblies

MAG	Binning method	Compl. / contam %	C fixation (Wood Ljungdhal)	N fixation (*nifDKH*)	NO_3^- reduction	NO_2^- reduction	SO_4^{2-} reduction	Thiosulfate reduction	Fe(III) reduction	Taxonomy
Desulfo_42_786	manual	97.4 / 0.6	+	+	-	+	-	-	+	g_Desulfosporosinus
Desulfo_43_18	manual	97.9 / 0.6	+	+	-	+	+	-	-	g_ Desulfosporosinus
Flavo_36_14	manual	92.1 / 3.5	-	-	-	+	-	-	-	g_Flavobacteria
Xantho_67_416	manual	96.1 / 1.3	-	-	+	+	-	-	-	g_Lysobacter
Elusi_67_181	manual	95.0 / 1.5	-	+	+	+	-	-	-	c_Elusimicrobia
Actino_63_150	manual	95.0 / 1.7	-	+	+	+	-	-	+	c_Coriobacteriia
Erysipelo_33_100	manual	94.3 / 0.9	-	-	-	-	-	-	-	f_Erysipelotrichaceae
Geobacter_60_60	manual	**99.4 / 0.7**	-	+	+	+	-	-	+	g_Geobacter
Actino_49_34	manual	93.2 / 0.8	-	-	-	-	-	-	-	f_Nanopelagicaceae
Erysipelo_41_14	manual	70.1 / 0.9	-	-	+	-	-	-	-	f_Erysipelotrichaceae
Desulfovib_63_448	manual	**99.1 / 0.0**	-	+	+	+	+	-	-	Desulfovibrio
Lyso_66_53	manual	95.1 / 1.2	-	-	-	-	-	-	+	Lysobacter*
Pseudorho_61_33	manual	95.3 / 1.8	-	-	-	-	-	-	-	Pseudorhodobacter*
Pseudom_60_25	manual	95.0 / 2.5	-	-	-	+	-	-	-	o_Pseudomonadales*

*CheckM taxonomic assignment

Manual binning

QUICKLOOKS pipeline

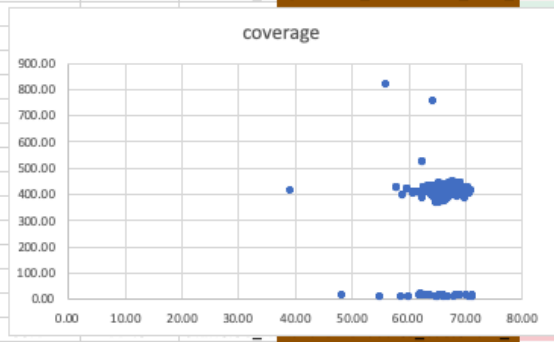
(adapted from WrightonGroup for UoM HPC)

	A	B	C	D	E	F	G	H	I	J	K
1	contig	gene		leng	GC	covera	sbjct_id	Taxon	Identity	evalue	bit_scor
102	_scaffold_596	1	#	22623	59.74	**457.58**	UniRef90_A1VW06_FAD-dependent_pyridine_nucleotide-disulphide_oxidore	Polaromonas_naphthalenivorans_(strain_CJ2)_TaxID=365044_	90.8	1.20E-85	324.3
103	_scaffold_596	2	#	22623	59.74	**457.58**	UniRef90_UPI000BBCD568_hypothetical_protein_n=1_Tax=Polaromonas_sp.	Polaromonas_sp._AER18D-145_TaxID=1977060_	56.8	1.50E-41	177.6
104	_scaffold_596	3	#	22623	59.74	**457.58**					
105	_scaffold_596	4	#	22623	59.74	**457.58**	UniRef90_A1VX29_Phage_integrase_family_protein_n=1_Tax=Polaromonas_	Polaromonas_naphthalenivorans_(strain_CJ2)_TaxID=365044_	92.4	2.70E-160	573.2
106	_scaffold_596	5	#	22623	59.74	**457.58**	UniRef90_N6Y9J2_Uncharacterized_protein_n=2_Tax=Thauera_TaxID=33057	Thauera_TaxID=33057_	76.2	8.00E-170	605.1
107	_scaffold_596	6	#	22623	59.74	**457.58**	UniRef90_UPI0005577216_hypothetical_protein_n=1_Tax=Herbaspirillum_Ta	Herbaspirillum_TaxID=963_	80.3	2.90E-88	333.2
108	_scaffold_596	7	#	22623	59.74	**457.58**	UniRef90_A1VX52_Relaxase/mobilization_nuclease_family_protein_n=2_Tax	Polaromonas_naphthalenivorans_TaxID=216465_	66.1	1.10E-224	788.5
109	_scaffold_596	8	#	22623	59.74	**457.58**	UniRef90_A1VX53_Uncharacterized_protein_n=2_Tax=Polaromonas_naphtha	Polaromonas_naphthalenivorans_TaxID=216465_	81	5.10E-39	168.7
110	_scaffold_596	9	#	22623	59.74	**457.58**	UniRef90_T0Z8C3_Uncharacterized_protein_n=1_Tax=Leptospirillum_sp._Gr	Leptospirillum_sp._Group_IV_'UBA_BS'_TaxID=1260983_	46.4	3.40E-19	103.2
111	_scaffold_596	10	#	22623	59.74	**457.58**	UniRef90_A0A1E7YTM2_Uncharacterized_protein_n=1_Tax=Acidithiobacillus	Acidithiobacillus_caldus_TaxID=33059_	69.9	6.60E-81	308.9
112	_scaffold_596	11	#	22623	59.74	**457.58**	UniRef90_A0A068YH56_Uncharacterized_protein_n=1_Tax=Polaromonas_sp	Polaromonas_sp._CG9_12_TaxID=1504672_	60.1	1.10E-76	295
113	_scaffold_596	12	#	22623	59.74	**457.58**	UniRef90_A1WDD9_Uncharacterized_protein_n=7_Tax=Burkholderiales_Tax	Burkholderiales_TaxID=80840_	62.7	8.20E-19	101.3
114	_scaffold_596	13	#	22623	59.74	**457.58**	UniRef90_UPI0004864034_KfrA_protein_n=1_Tax=Curvibacter_gracilis_TaxID	Curvibacter_gracilis_TaxID=230310_	65.9	2.70E-107	397.1
115	_scaffold_596	14	#	22623	59.74	**457.58**	UniRef90_UPI00082CA883_recombinase_family_protein_n=2_Tax=Marinoba	Marinobacterium_profundum_TaxID=1714300_	90	2.50E-95	356.7
116	_scaffold_596	15	#	22623	59.74	**457.58**	UniRef90_UPI000561A89A_hypothetical_protein_n=1_Tax=Methylobacter_w	Methylobacter_whittenburyi_TaxID=39770_	59.6	1.00E-129	471.9
117	_scaffold_596	16	#	22623	59.74	**457.58**	UniRef90_UPI00055F6179_transposition_protein_TniB_n=1_Tax=Methylobac	Methylobacter_whittenburyi_TaxID=39770_	81.4	3.80E-132	479.6
118	_scaffold_596	17	#	22623	59.74	**457.58**	UniRef90_UPI00055EC4F6_transposase_n=1_Tax=Methylobacter_whittenbur	Methylobacter_whittenburyi_TaxID=39770_	76.4	5.60E-246	858.6
119	_scaffold_596	18	#	22623	59.74	**457.58**	UniRef90_A0A2Z5VKH5_ISSod9,_transposase_n=46_Tax=Bacteria_TaxID=2_	Bacteria_TaxID=2_	83	0	1667.5
120	_scaffold_596	19	#	22623	59.74	**457.58**	UniRef90_A0A2S1FIJ2_Replication_initiation_protein_n=1_Tax=Polaromonas	Polaromonas_sp._H1N_TaxID=1840283_	95.8	1.20E-151	544.3
121	_scaffold_596	20	#	22623	59.74	**457.58**	UniRef90_A0A2S1FI71_Putative_partitioning_protein_ParB_n=1_Tax=Polaror	Polaromonas_sp._H1N_TaxID=1840283_	87.9	1.10E-44	187.6
122	_scaffold_596	21	#	22623	59.74	**457.58**	UniRef90_A0A2S1FI83_Partitioning_protein_ParA_n=1_Tax=Polaromonas_sp	Polaromonas_sp._H1N_TaxID=1840283_	97.4	2.10E-117	430.3
123	_scaffold_2500	1	#	8630	68.00	**446.03**	UniRef90_A0A0A0EQF8_Patatin_n=1_Tax=Lysobacter_concretionis_Ko07_=_	Lysobacter_concretionis_Ko07_=_DSM_16239_TaxID=1122185_	76.5	4.80E-116	426
124	_scaffold_2500	2	#	8630	68.00	**446.03**	UniRef90_A0A0Q9EP21_ATPase_n=7_Tax=Lysobacter_TaxID=68_RepID=A0A	Lysobacter_TaxID=68_	83.7	0	1875.9
125	_scaffold_2500	3	#	8630	68.00	**446.03**	UniRef90_E6WWY3_50S_ribosomal_protein_L33_n=35_Tax=Xanthomonadal	Xanthomonadales_TaxID=135614_	85.2	1.50E-18	99.8
126	_scaffold_2500	4	#	8630	68.00	**446.03**	UniRef90_A0A345ZMX2_50S_ribosomal_protein_L28_n=2_Tax=Lysobacter_T	Lysobacter_TaxID=68_	92.3	6.20E-34	151.4
127	_scaffold_2500	5	#	8630	68.00	**446.03**	UniRef90_A0A0Q8CTA7_Uncharacterized_protein_n=2_Tax=Xanthomonadac	Xanthomonadaceae_TaxID=32033_	91.9	1.20E-54	220.7
128	_scaffold_2500	6	#	8630	68.00	**446.03**	UniRef90_A0A0Q8CWE8_Dolichol-phosphate_mannosyltransferase_n=10_Ta	Lysobacter_TaxID=68_	87.9	1.20E-118	434.5
129	_scaffold_2500	7	#	8630	68.00	**446.03**	UniRef90_UPI000F429F70_ParB/RepB/Spo0J_family_partition_protein_n=1_	Lysobacter_sp._ZS60_TaxID=1327343_	85.2	4.50E-136	492.7
130	_scaffold_2500	8	#	8630	68.00	**446.03**	UniRef90_UPI000E6B464C_ParA_family_protein_n=1_Tax=Lysobacter_sp._C8	Lysobacter_sp._C8-1_TaxID=2315424_	83.3	5.30E-111	409.1
131	_scaffold_22027	1	#	2488	62.46	**443.68**	UniRef90_A0A2S1FI54_Transposase_n=2_Tax=Polaromonas_TaxID=52972_R	Polaromonas_TaxID=52972_	83.6	1.60E-74	287.3
132	_scaffold_22027	2	#	2488	62.46	**443.68**	UniRef90_UPI00049825CA_IS3_family_transposase_n=1_Tax=Polaromonas_	Polaromonas_glacialis_TaxID=866564_	91.6	5.60E-191	675.2
133	_scaffold_22027	3	#	2488	62.46	**443.68**	UniRef90_A1VVP6_Transposase_IS3/IS911_family_protein_n=1_Tax=Polaror	Polaromonas_naphthalenivorans_(strain_CJ2)_TaxID=365044_	87.7	5.10E-57	228.8
134	_scaffold_22027	4	#	2488	62.46	**443.68**	UniRef90_A1VVP7_IS66_Orf2_family_protein_n=5_Tax=Comamonadaceae_T	Comamonadaceae_TaxID=80864_	97.6	2.10E-40	172.9

Manual binning

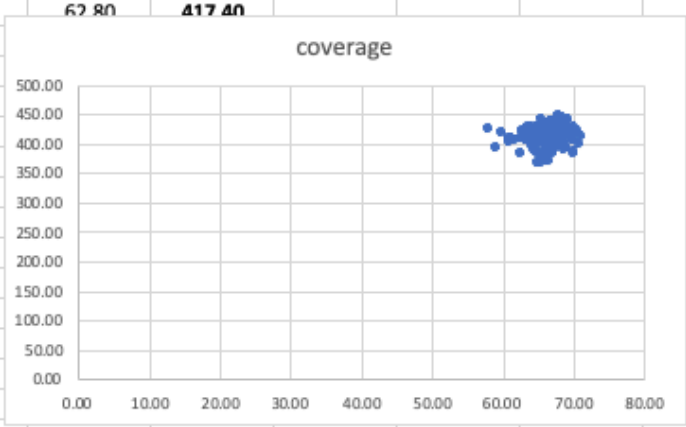
QUICKLOOKS pipeline
(adapted from WrightonGroup for CSF3)

	contig	gene	Reads	length	GC	coverage	sbjct_id	Taxon	Identity	evalue	bit_score
2	_scaffold_20	1	11747	1800	56.06	**815.76**	UniRef90_A0	Xanthomonas_fragariae_TaxID=	93	1.60E-39	170.2
3	_scaffold_20	2	11747	1800	56.06	**815.76**	UniRef90_A0	cellular_organisms_TaxID=1315	88	4.90E-16	91.3
4	_scaffold_31	1	7638	1269	64.30	**752.36**	UniRef90_A0	Xanthomonas_campestris_pv._t	73.5	6.70E-118	432.2
5	_scaffold_31	2	7638	1269	64.30	**752.36**	UniRef90_A0	Xanthomonas_TaxID=338_	70.8	1.30E-28	134
6	_scaffold_53	1	33334	7954	62.56	**523.86**	UniRef90_U	Massilia_sp._K1S02-61_TaxID=2	71.5	1.90E-174	620.9
7	_scaffold_53	2	33334	7954	62.56	**523.86**	UniRef90_U	Dyella-like_sp._DHo_TaxID=166	84.5	2.30E-171	610.1
8	_scaffold_53	3	33334	7954	62.56	**523.86**	UniRef90_A0	Gammaproteobacteria_bacteriu	84.9	4.30E-34	152.1
9	_scaffold_53	4	33334	7954	62.56	**523.86**	UniRef90_A0	Gammaproteobacteria_bacteriu	92	2.10E-31	142.9
10	_scaffold_53	5	33334	7954	62.56	**523.86**	UniRef90_U	Vulcaniibacterium_tengchongen	71.6	3.20E-17	95.9
11	_scaffold_53	6	33334	7954	62.56	**523.86**	UniRef90_U	Lysobacter_TaxID=68_	70.7	2.60E-147	530.4
12	_scaffold_53	7	33334	7954	62.56	**523.86**	UniRef90_A0	Lysobacter_arseniciresistens_ZS	78.1	3.20E-134	486.5
13	_scaffold_53	8	33334	7954	62.56	**523.86**					
14	_scaffold_25	1	30794	8630	68.00	**446.03**	UniRef90_A0	Lysobacter_concretionis_Ko07_=	76.5	4.80E-116	426
15	_scaffold_25	2	30794	8630					83.7	0	1875.9
16	_scaffold_25	3	30794	8630					85.2	1.50E-18	99.8
17	_scaffold_25	4	30794	8630					92.3	6.20E-34	151.4
18	_scaffold_25	5	30794	8630					91.9	1.20E-54	220.7
19	_scaffold_25	6	30794	8630					87.9	1.20E-118	434.5
20	_scaffold_25	7	30794	8630					85.2	4.50E-136	492.7
21	_scaffold_25	8	30794	8630					83.3	5.30E-111	409.1
22	_scaffold_33	1	33134	9357					77.2	0	3672.5
23	_scaffold_33	2	33134	9357					69.6	3.90E-190	672.9
24	_scaffold_44	1	87342	24667					72.2	3.80E-62	246.1
25	_scaffold_44	2	87342	24667					68.2	1.50E-106	394.4
26	_scaffold_44	3	87342	24667							
27	_scaffold_44	4	87342	24667					76.2	5.60E-154	552.4
28	_scaffold_44	5	87342	24667					64.1	4.10E-35	156
29	_scaffold_44	6	87342	24667	68.42	**442.61**	UniRef90_A0	Lysobacter_TaxID=68_	79.6	0	1305.8
30	_scaffold_44	7	87342	24667	68.42	**442.61**	UniRef90_A0	Lysobacter_TaxID=68_	87.8	1.30E-134	487.6
31	_scaffold_44	8	87342	24667	68.42	**442.61**	UniRef90_A0	Lysobacter_concretionis_Ko07_=	77	5.00E-158	565.8
32	_scaffold_44	9	87342	24667	68.42	**442.61**	UniRef90_A0	Xanthomonadaceae_TaxID=320	91.1	1.30E-75	290.8
33	_scaffold_44	10	87342	24667	68.42	**442.61**	UniRef90_A0	Xanthomonadaceae_TaxID=320	90.7	2.10E-35	156.8
34	_scaffold_44	11	87342	24667	68.42	**442.61**	UniRef90_U	Lysobacter_sp._ZS60_TaxID=132	82.6	2.50E-145	523.5



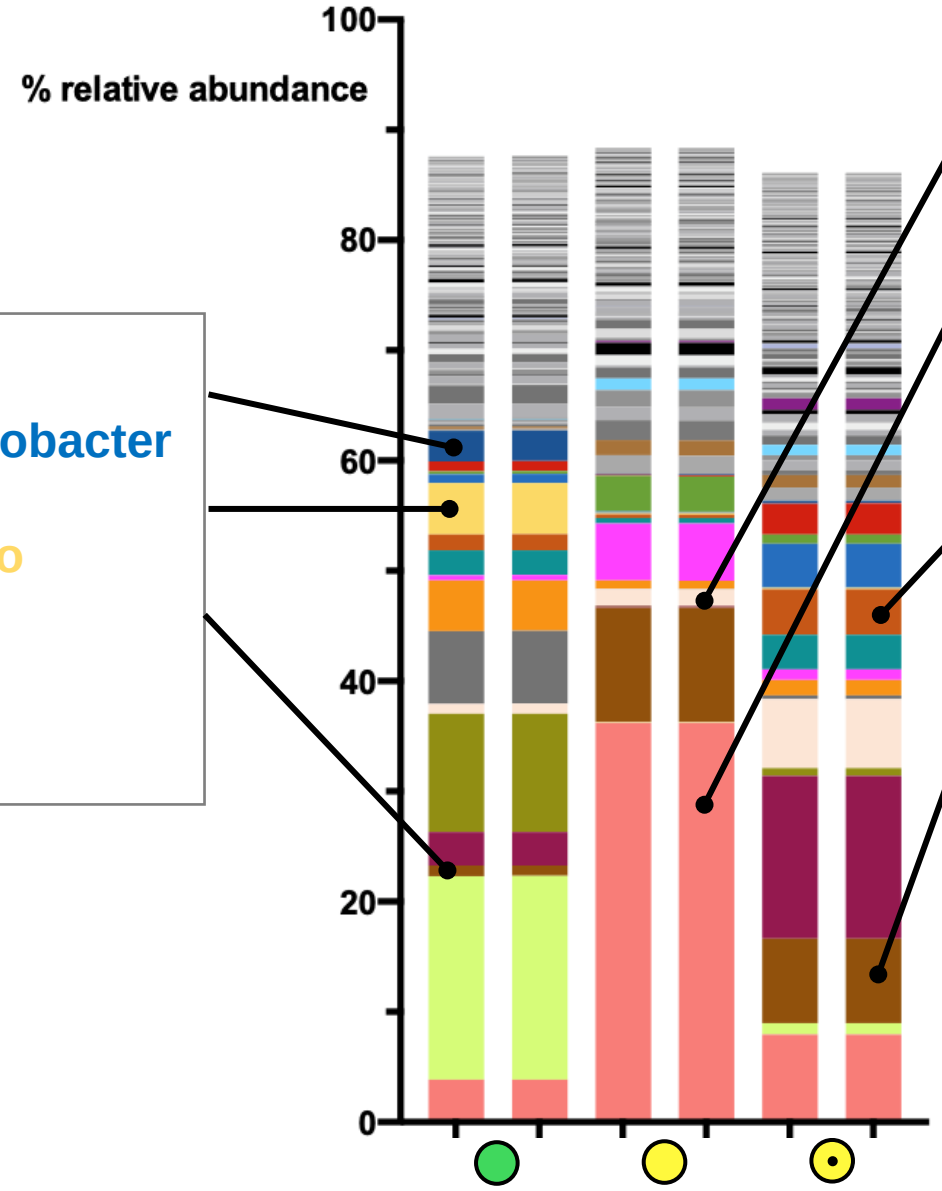
Xantho_67_416

	contig	length	GC	coverag
1	_scaffold_11	14563	58.11	**423.69**
2	_scaffold_20	2228	59.07	**392.67**
3	_scaffold_19	11025	59.86	**419.83**
4	_scaffold_28	11329	60.91	**408.48**
5	_scaffold_26	1445	61.04	**400.95**
6	_scaffold_16	3311	61.07	**408.30**
7	_scaffold_19	15637	61.77	**406.06**
8	_scaffold_39	1034	62.57	**383.70**
9	_scaffold_20	19984	62.72	**408.73**
10	_scaffold_59	13306	62.80	**417.40**
11	_scaffold_19	19769		
12	_scaffold_73	19638		
13	_scaffold_10	3006		
14	_scaffold_23	9360		
15	_scaffold_12	14532		
16	_scaffold_55	4889		
17	_scaffold_18	16980		
18	_scaffold_26	8297		
19	_scaffold_22	10768		
20	_scaffold_19	1849		
21	_scaffold_16	2137		
22	_scaffold_19	35945		
23	_scaffold_13	2553		
24	_scaffold_23	1600	64.44	**389.61**
25	_scaffold_73	45004	64.49	**414.25**
26	_scaffold_26	1467	64.55	**385.99**



18		2587259	66.51	415.68
19			1.98	14.73
20				
21		CheckM v1.0.18 25.5.20		
22		96.1% complete, 1.28% contaminated		
23				

MAGs in context



g_Pseudorhodobacter

g_Desulfovibrio

g_Lysobacter

g_Flavobacteria

g_Desulfosporosinus
- Desulfo_42_786
- Desulfo_43_18

g_Geobacter

g_Lysobacter

c_Coriobacteriia

f_Nanopelagicaceae

f_Erysipelotrichaceae

Expressed genes in *Desulfosporosinus*

annotated metaT
(vs binned metaG)

BLASTp to *Desulfo*_43_786

15,556 hits from annotated metaT to MAG!
(e-value e-10, bit score >60)

195 chemotaxis

163 putative multidrug resistance genes

99 osmoprotectant import

96 sporulation

33 acetyl-CoA dehydrogenase

24 pyruvate synthase

14 sulfite reductase

14 cyctochrome c

13 alkaline shock protein genes

...

QUICKLOOKS pipeline*

Genes from assembled metaT taxonomically assigned to *Desulfosporosinus*

Flagellin / flagellar synthesis

Dissimilatory sulfite reductase (MAG 42_18?)

cytochrome c family protein (MAG 43_786?)

oxidoreductases

CO dehydrogenase

Tellurium reductase & stress proteins

...

*WrightonLab github page

Viral community members

(water samples filtered through 0.22um)

metaG reads and assemblies

KAIJU

	Virus abundance (% classified reads)	**# Viral contigs** (# total)	**Viral contigs** (# per million read pairs)	**CRISPRs** (# per million read pairs)
191 m	0.1%	103	2.8	2.8
561 – 571 m	1.0%	103	3.7	1.0
571 – 600 m	0.4%	115	3.2	1.9
Talik Lake	0.6%	105	2.7	5.2
Drill pond	1.1%	...	...	6.0

- **Similar (low) abundance of viruses** across all communities, despite sample filtration
- **Lower viral immunity in subsurface communities** compared with surface lake / pond communities

Viral immunity in MAGs

MAGs

before I finished manual binning…

MAG	Binning method	Compl. / contam %	Prophage	CRISPR loci	# CRISPR spacers	Taxonomy
Desulfo_42_786	manual	97.4 / 0.6	1	2	9	g_Desulfosporosinus
Desulfo_43_18	manual	97.9 / 0.6	-	-	-	g_ Desulfosporosinus
Flavo_36_14	manual	92.1 / 3.5	-	1	4	g_Flavobacteria
108_001	Maxbin 2.0	97.3 / 2.0	1	2	115	g_Lysobacter
108_003	Maxbin 2.0	94.9 / 1.6	1	-	-	TBD
108_004	Maxbin 2.0	95.0 / 5.1	1	-	-	c_Coriobacteriia
108_006	Maxbin 2.0	95.1 / 3.8	1	-	-	TBD
108_007	Maxbin 2.0	98.7 / 1.6	-	-	-	g_Geobacter
108_012	Maxbin 2.0	100 / 4.7	-	-	-	f_Nanopelagicaceae
108_023	Maxbin 2.0	91.0 / 2.4	-	-	-	f_Erysipelotrichaceae
Desulfovib_63_448	manual	99.1 / 0.0	-	-	-	Desulfovibrio
110_007	Maxbin 2.0	98.9 / 9.8	-	-	-	Lysobacter
110_012	Maxbin 2.0	97.8 / 3.4	-	-	-	Pseudorhodobacter
110_013	Maxbin 2.0	94.3 / 5.2	-	1	55	f_Burkholderiaceae

Key findings

- **Highly diverse** communities in all samples
- **Distinct composition** in above vs below permafrost samples

- Active communities are **highly diverse**
- **Low abundance taxa** appear to be **highly active** (Legionella (!), Bacillus, Lacunisphaera...)

- ***Desulfosporosinus*** members of the community actively **expressing genes for flagellar / chemitaxis, sporulation, C fixation, sulfite [R] and stress response**

- Subsurface communities encode **C, N and S cycling pathways**
- metaT data demonstrates that these pathways **are active in deep borehole**

- Several **high quality MAGs** were recovered from deep borehole samples
- Many **have the capacity for N fixation and dissimilatory nitrate/nitrite reduction**, few MAGs encode C fixation and sulfate reduction / thiosulfate oxidation

- **Similar (low) abundance of viruses** across all communities, despite sample filtration

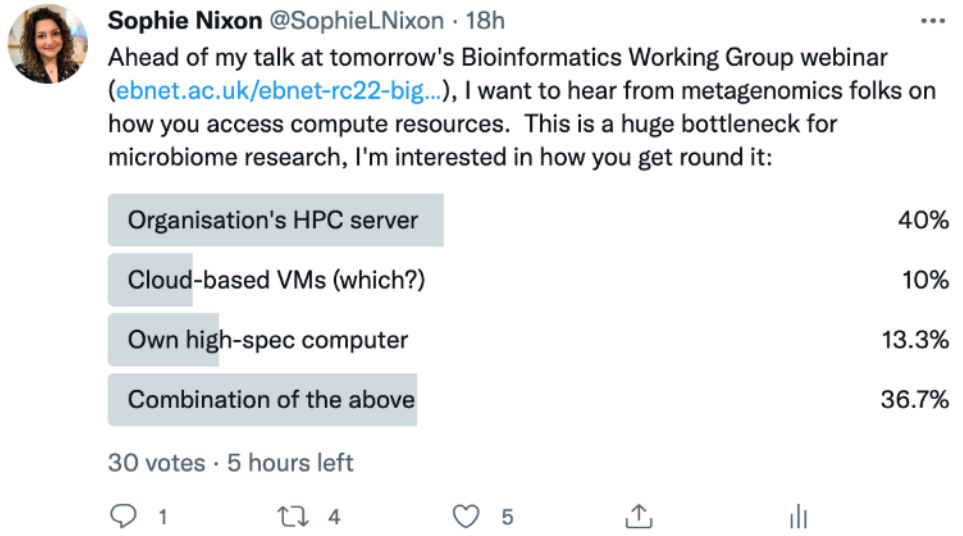
- **Lower viral immunity in subsurface communities** compared with surface lake / pond communities

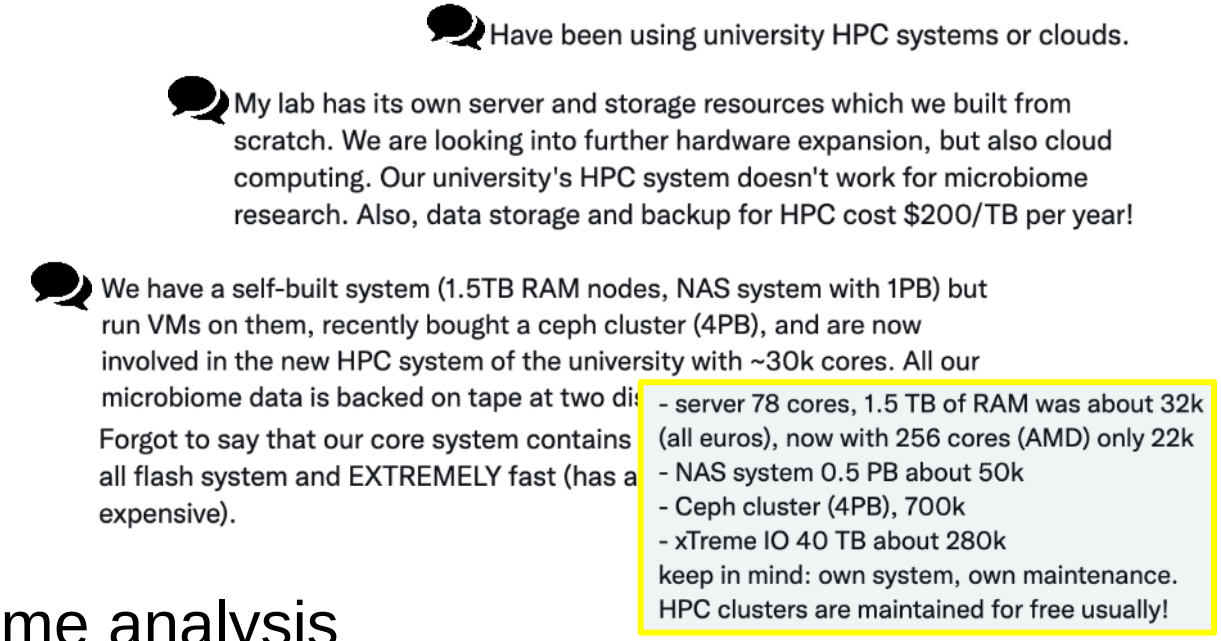
- **Some MAGs have prophage** and evidence of **viral immunity**

- **Net C fixation** – implications for geological CO_2 storage?

Challenges

- Access to (and cost of!) computational resources…





- and bioinformatics training for microbiome analysis

CLIMB BIG DATA

UKRI Biotechnology and Biological Sciences Research Council

(watch this space!)

PhD opportunity

Microbial carbon cycling in geological CO2 storage environments

The University of Manchester › Department of Earth and Environmental Sciences based in the Manchester Institute of Biotechnology

Dr Sophie Nixon, Prof R Breitling, Dr M Buckley, Prof D Johnson, Prof Perdita Barran

Thursday, March 31, 2022 Funded PhD Project (Students Worldwide)

- high pressure subsurface simulation
- anaerobic cultivation
- geochemical characterisation
- genome-resolved metagenomics
- metatranscriptomics
- metabolomics
- proteomics
- stable isotope probing
- biotechnological potential
- microbiome modeling

https://www.findaphd.com/phds/project/microbial-carbon-cycling-in-geological-co2-storage-environments/?p141491