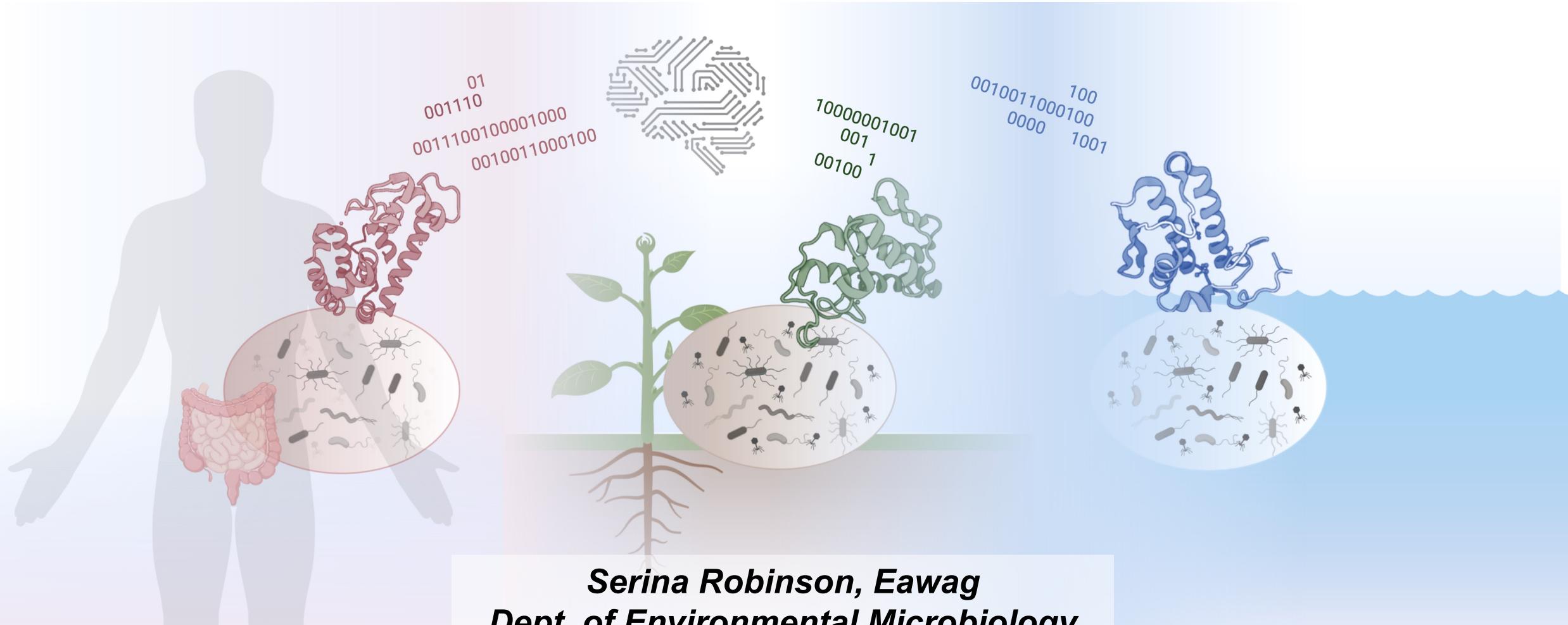


Metagenome mining to tap microbial functional potential



Serina Robinson, Eawag
Dept. of Environmental Microbiology
14.06.2023
Applied meta-omics workshop, ETHZ

Outline

Seminar (~30 minutes)

1. Introduction
2. Biosynthetic potential of the global ocean microbiome
 - a. Computational
 - b. Experimental
3. Future directions

Workshop module-3 introduction (~10 minutes)



Choose-your-own adventure

Acknowledgements



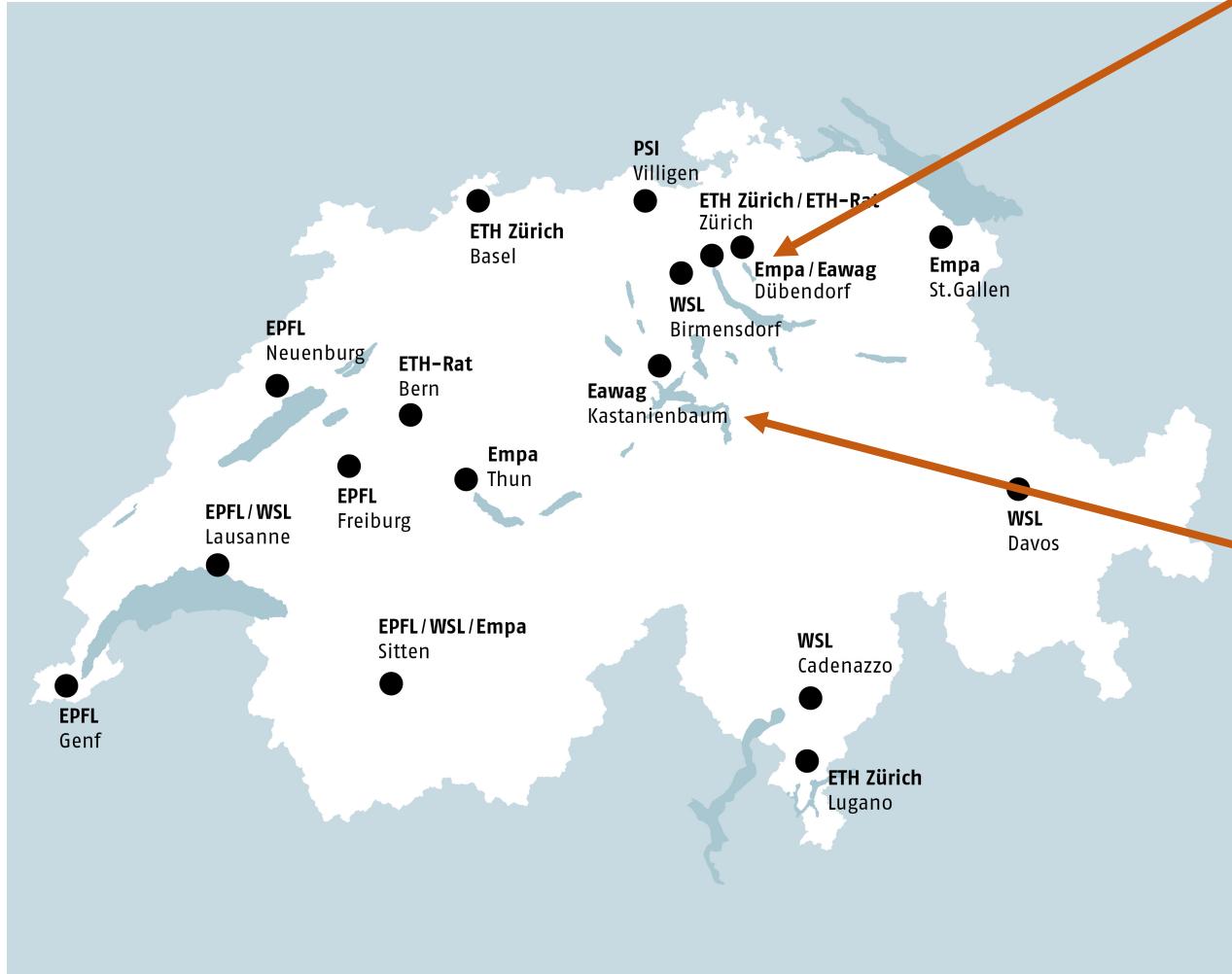
Piel group

Sunagawa group (& many more!)

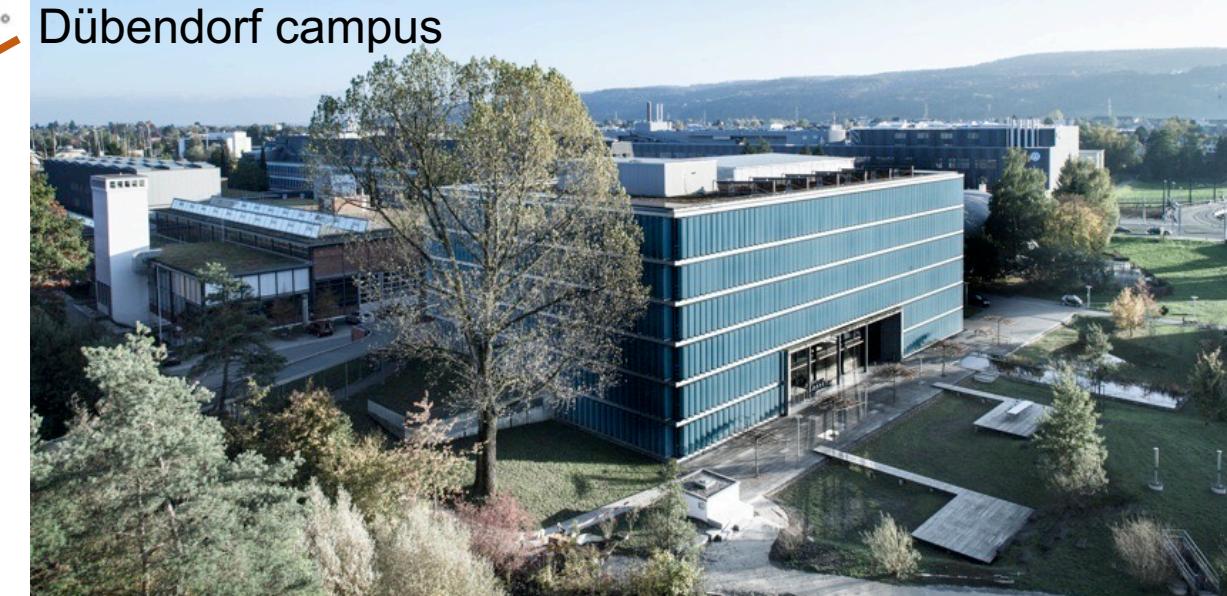


Swiss Federal Institute of Aquatic Science and Technology (Eawag)

ETH zürich EPFL



Dübendorf campus



Kastanienbaum campus



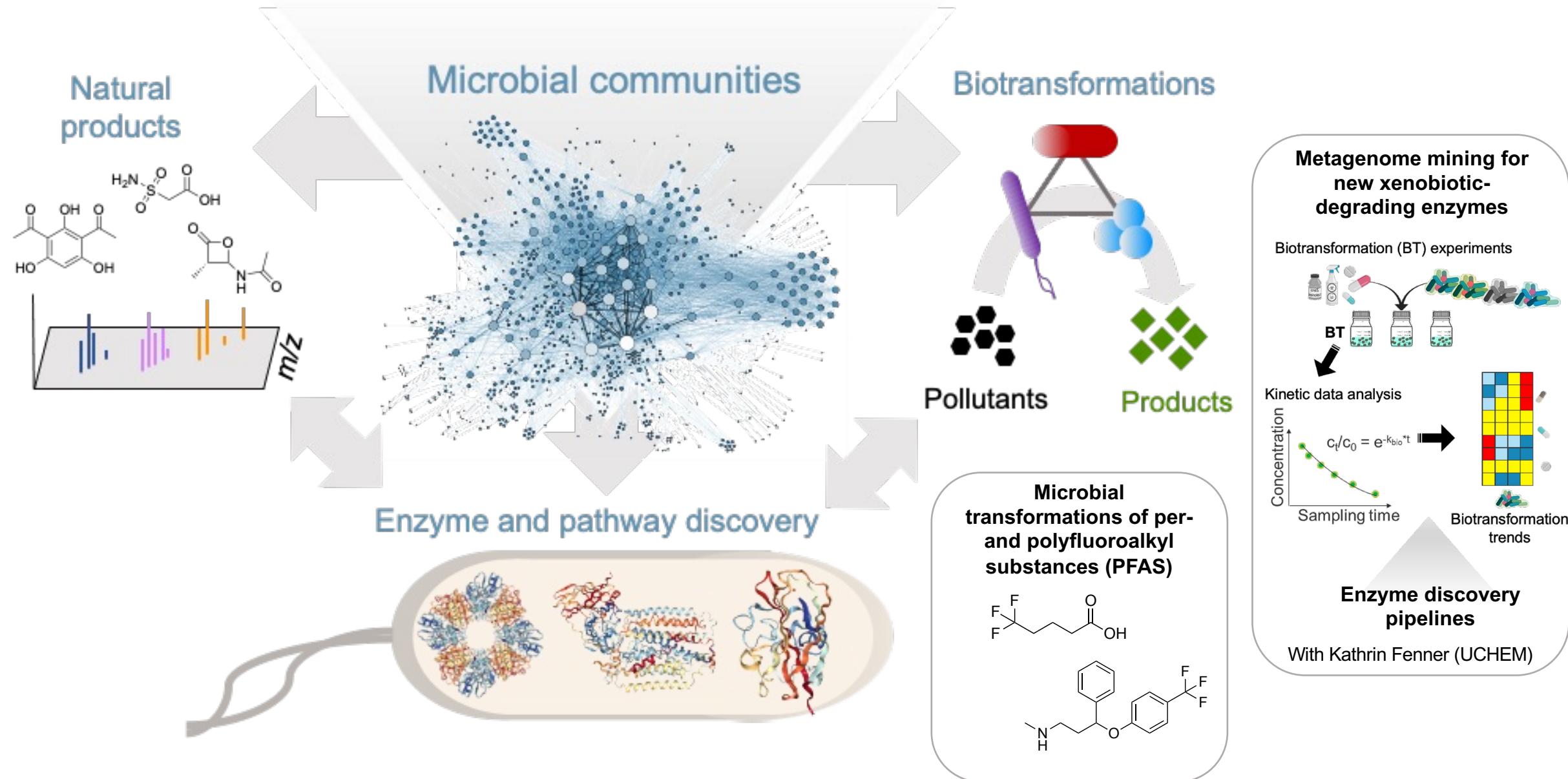


eawag

aquatic research ooo

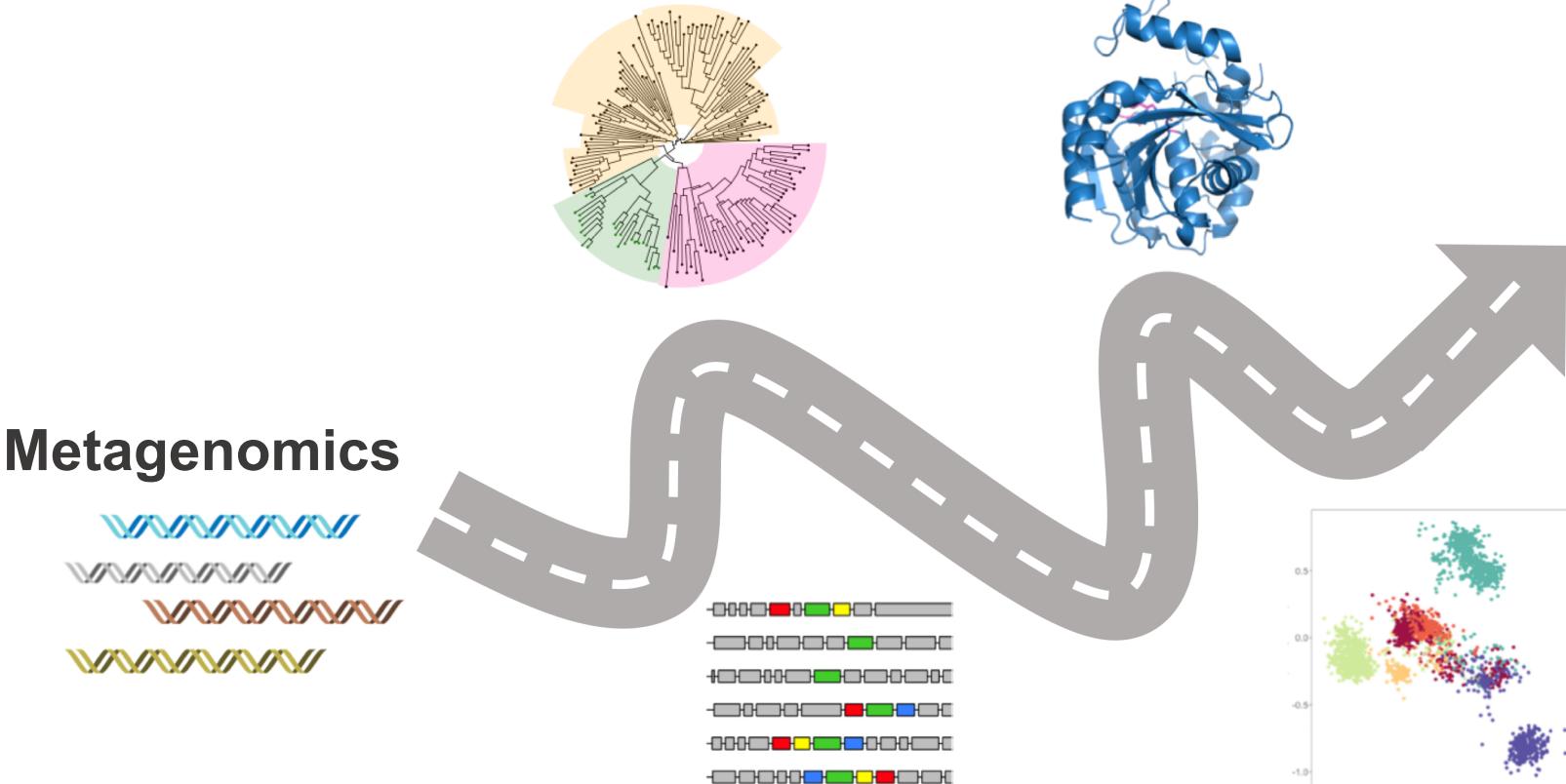
- Chemistry, physics, biology and microbiology of water
- Ecology of aquatic systems
- Drinking water and wastewater technologies

Microbial specialized metabolism group at Eawag

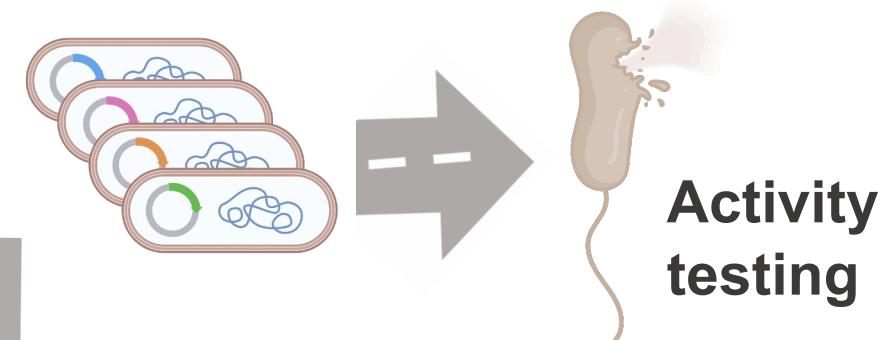


Roadmap for enzyme and pathway discovery

Computational

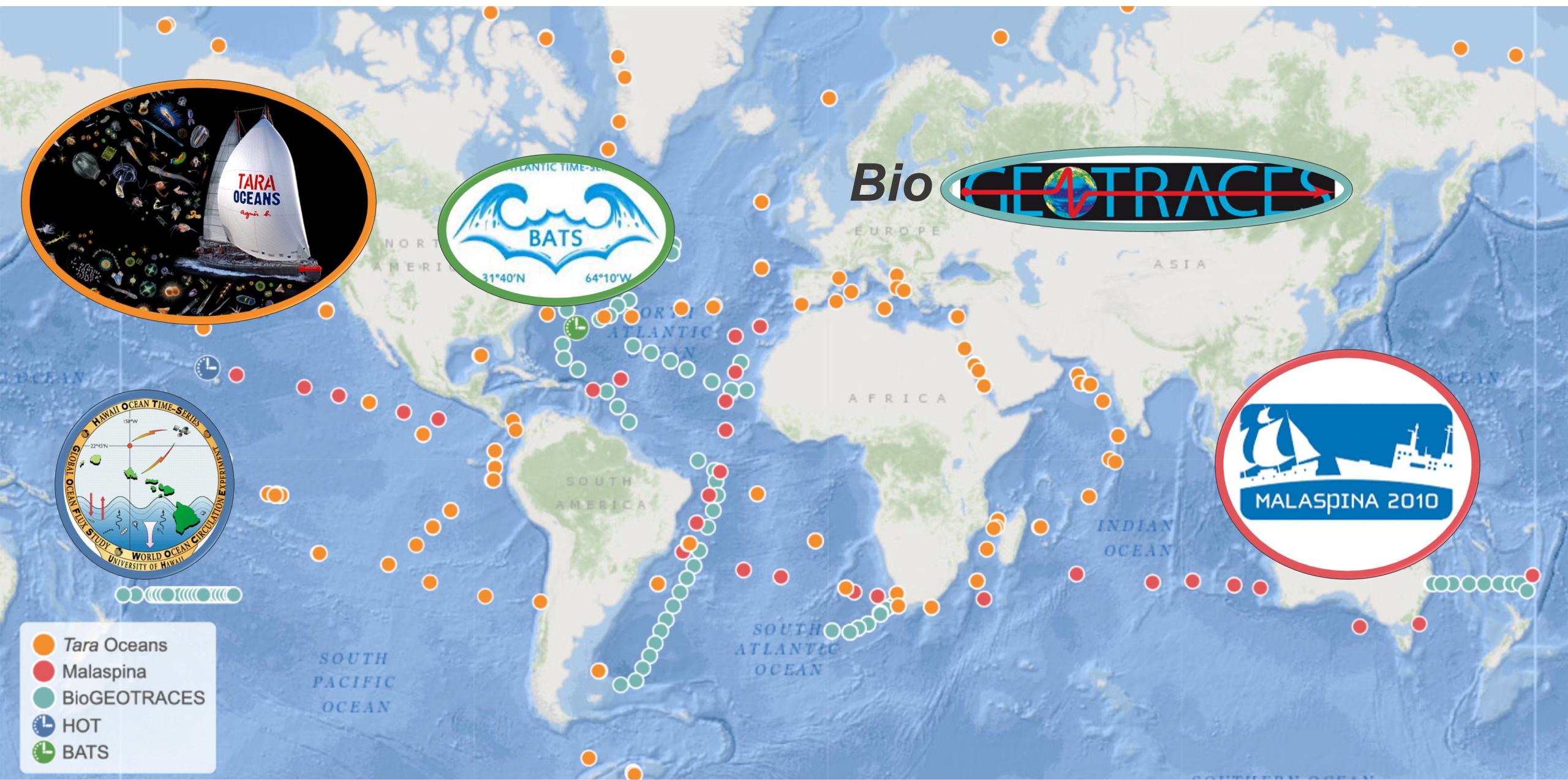


Experimental

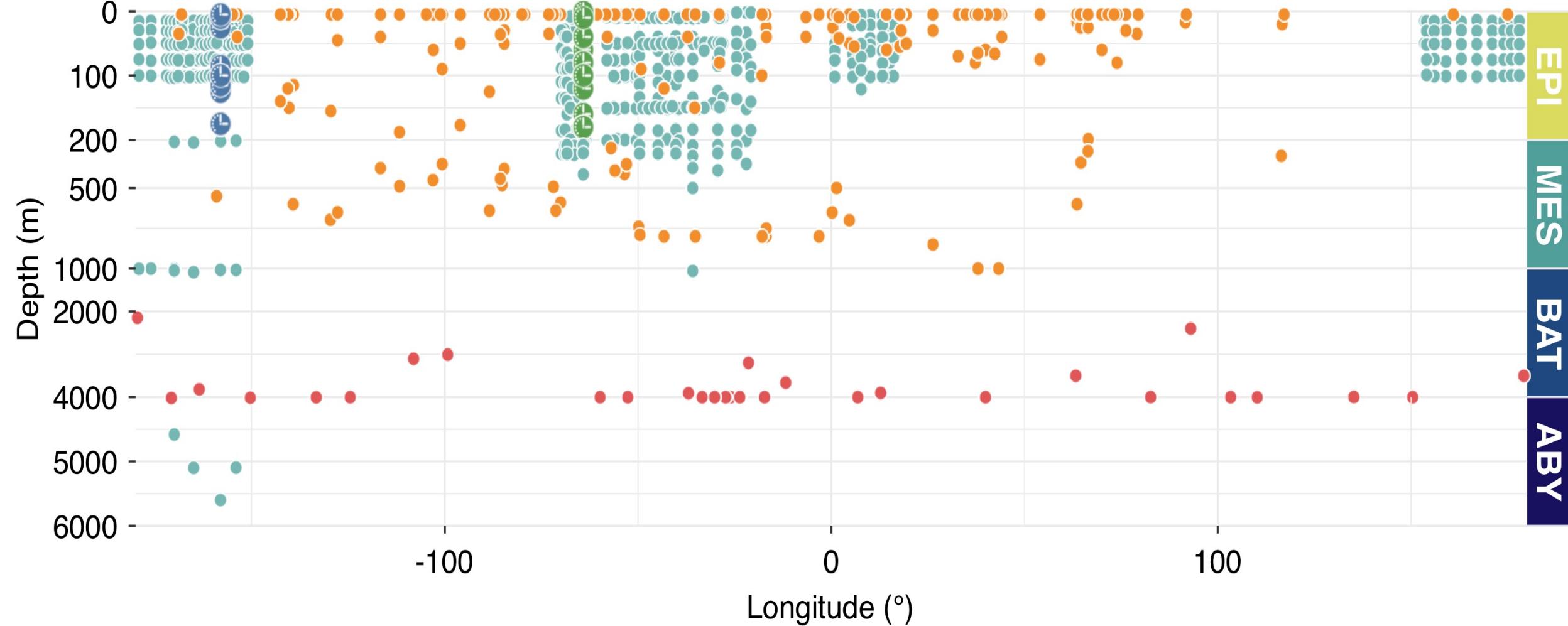


Activity
testing

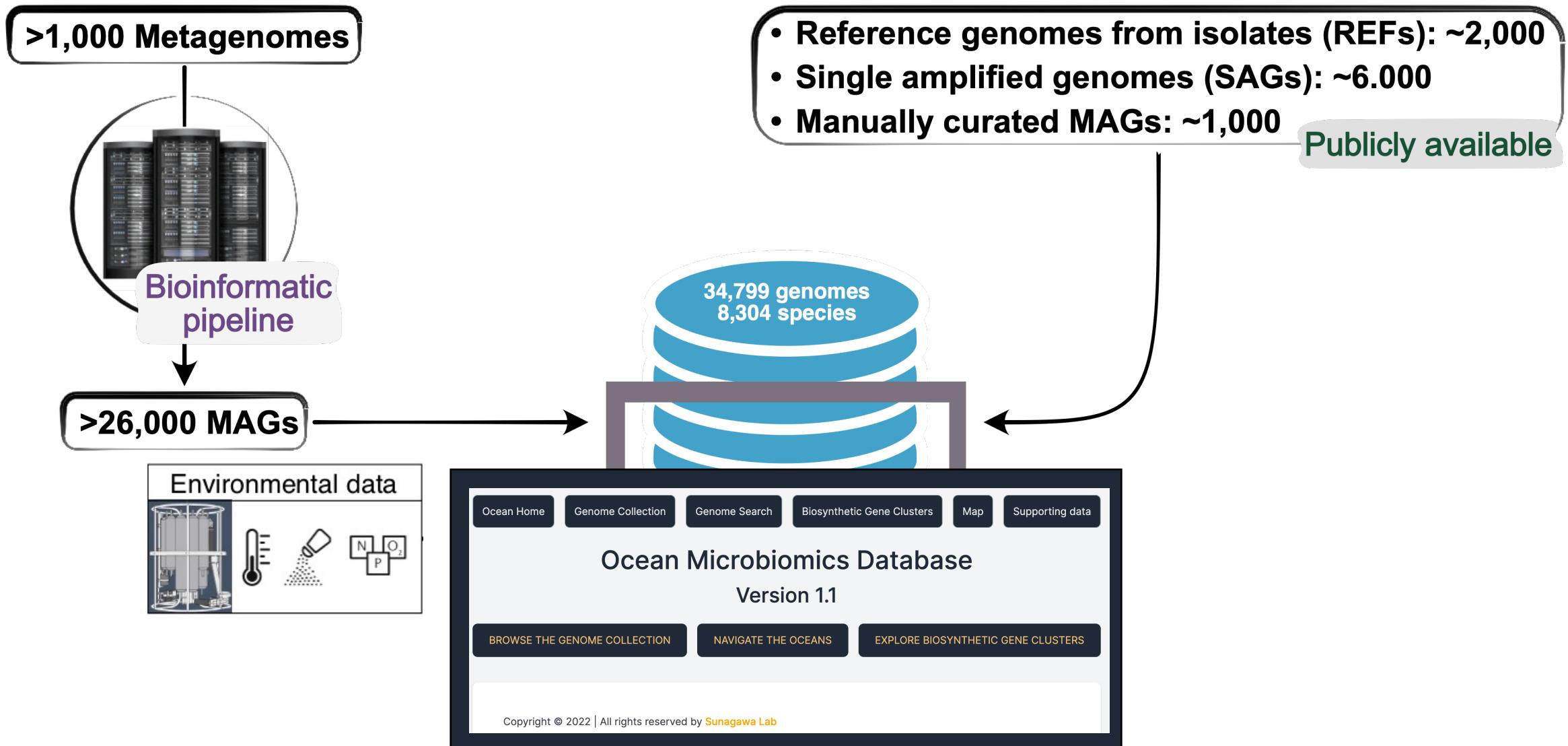
Biosynthetic potential of the global ocean microbiome



Aggregated >1k metagenomes across depth and time scales



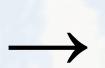
>26k MAGs in the Ocean Microbiomics Database (OMD)



Which microbes live in the open ocean?



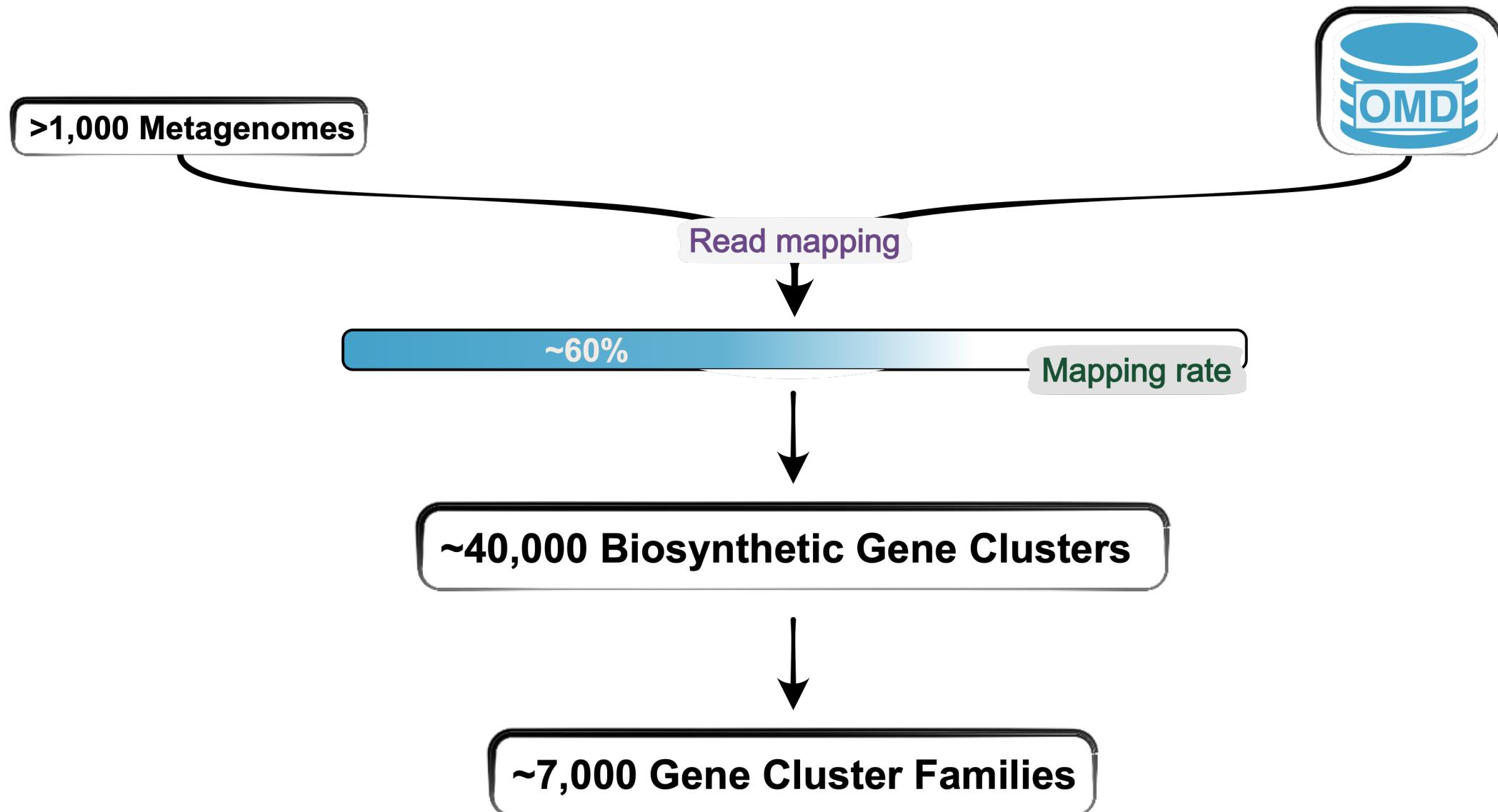
**Which microbes
live in the
open ocean?**



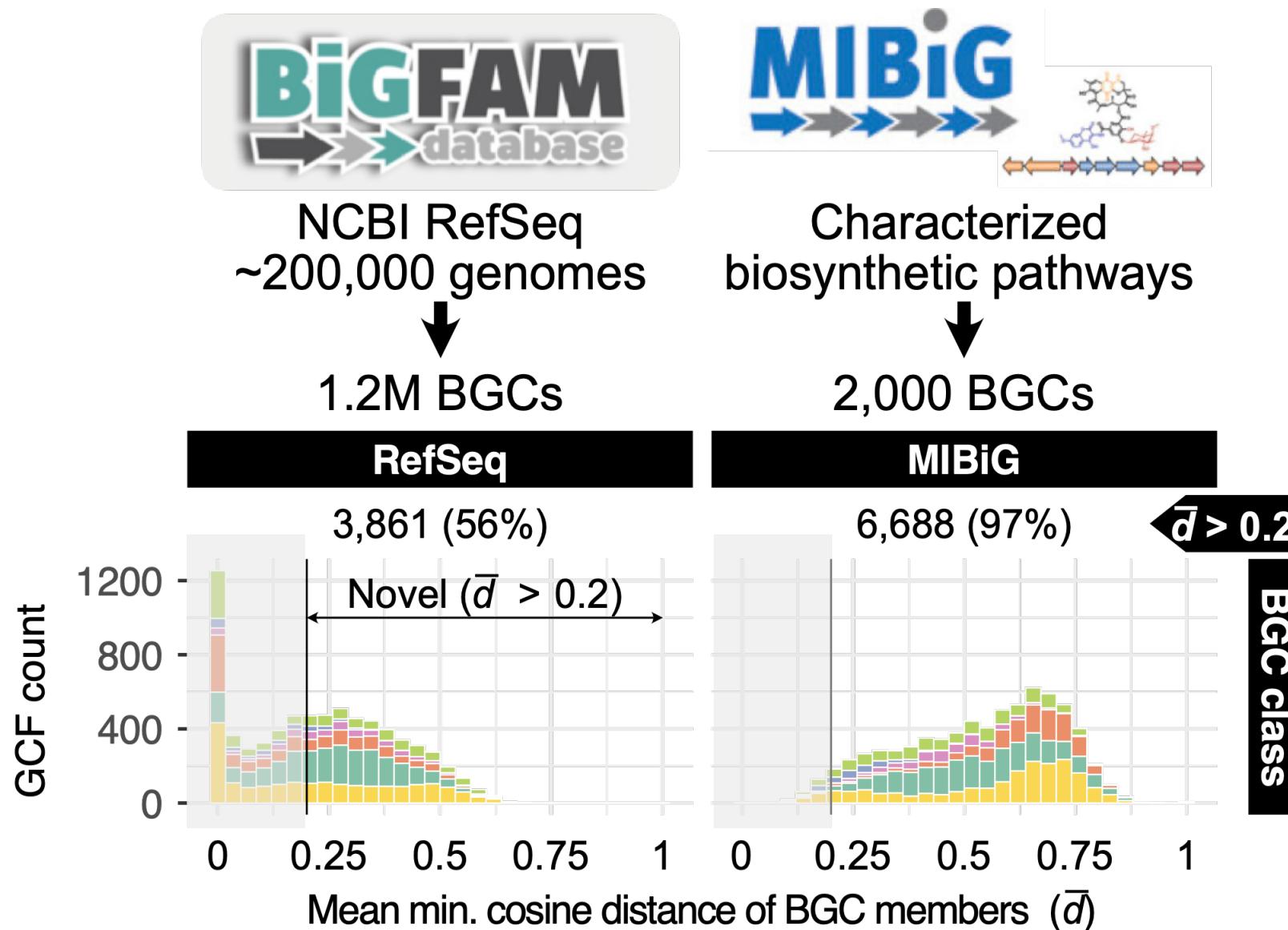
**Can we discover new enzymes,
secondary metabolites and
their ecological functions?**



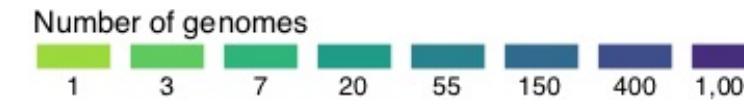
Ocean microbiome biosynthetic potential



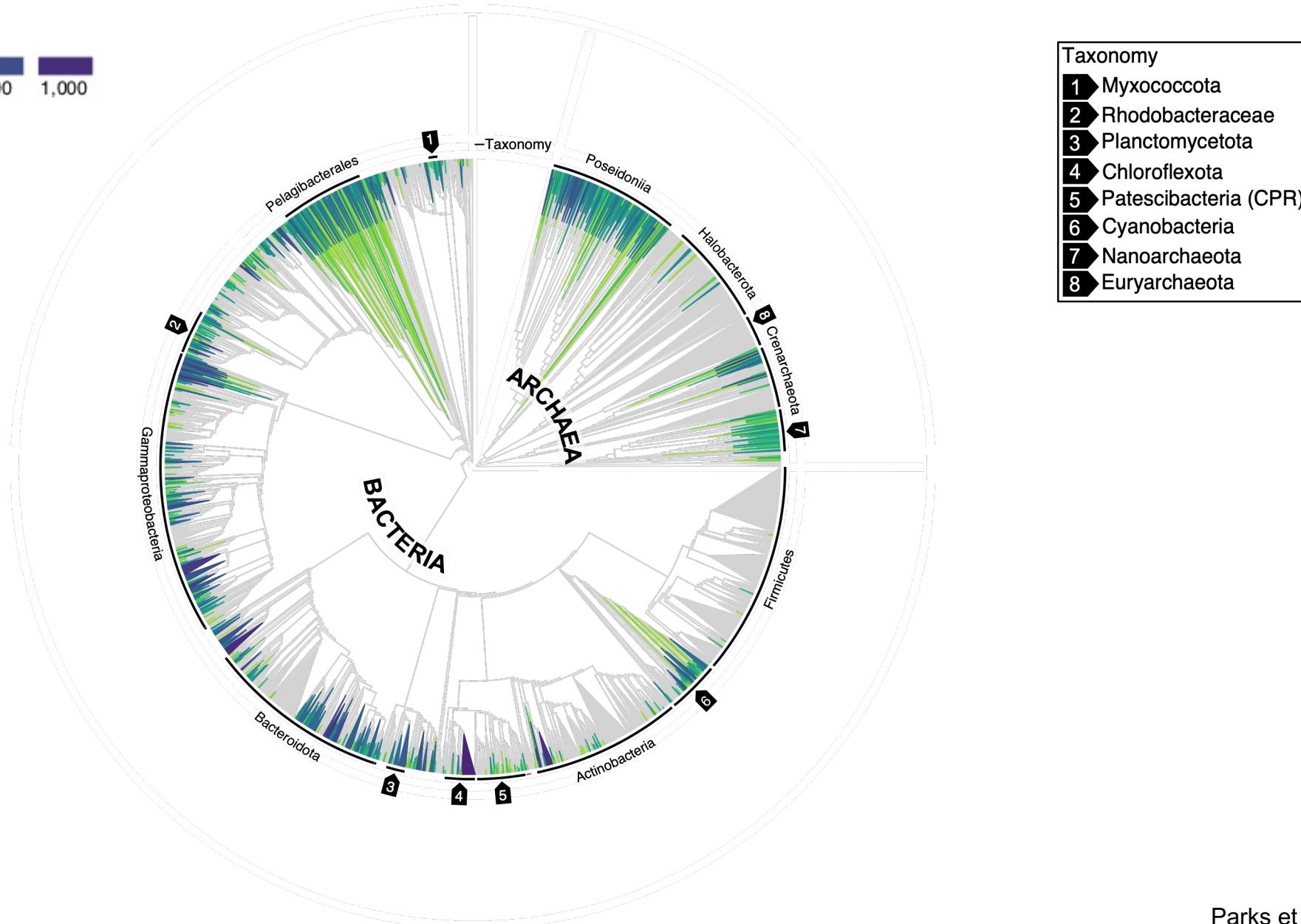
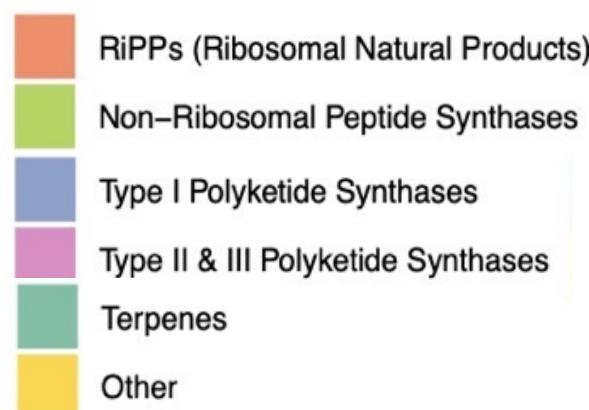
Large fraction of novel Gene Cluster Families (GCFs)



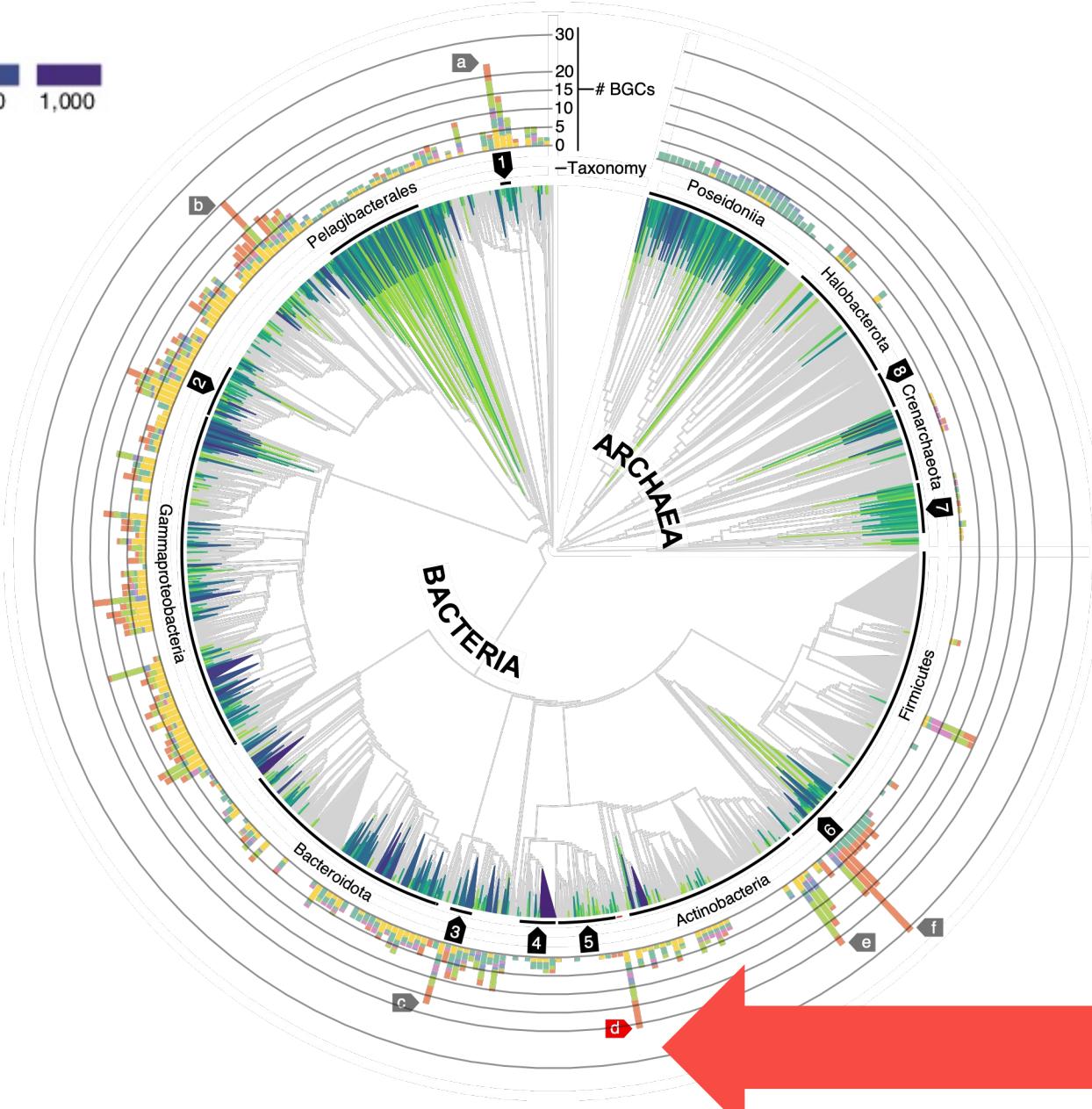
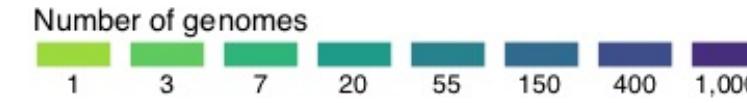
Phylogenomic distribution of the ocean biosynthetic potential



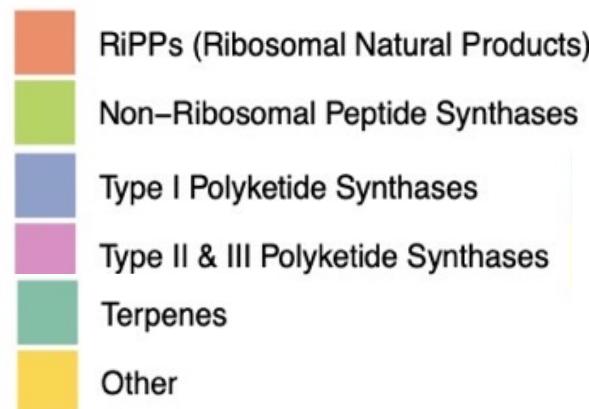
Biosynthetic gene cluster (BGC) type:



Phylogenomic distribution of the ocean biosynthetic potential



Biosynthetic gene cluster (BGC) type:



Taxonomy

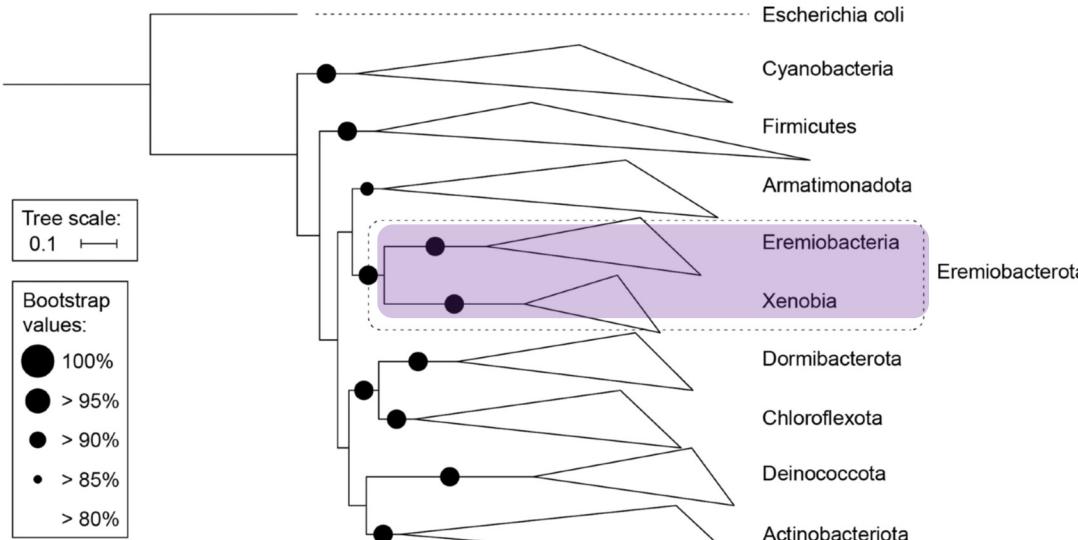
1 Myxococcota
2 Rhodobacteraceae
3 Planctomycetota
4 Chloroflexota
5 Patescibacteria (CPR)
6 Cyanobacteria
7 Nanoarchaeota
8 Euryarchaeota

BGC-rich lineages (>15 BGCs)

a Sandaracinaceae
b Tistrella
c Planctomycetota
d Eremiobacterota
e Rhodococcus
f Synechococcus

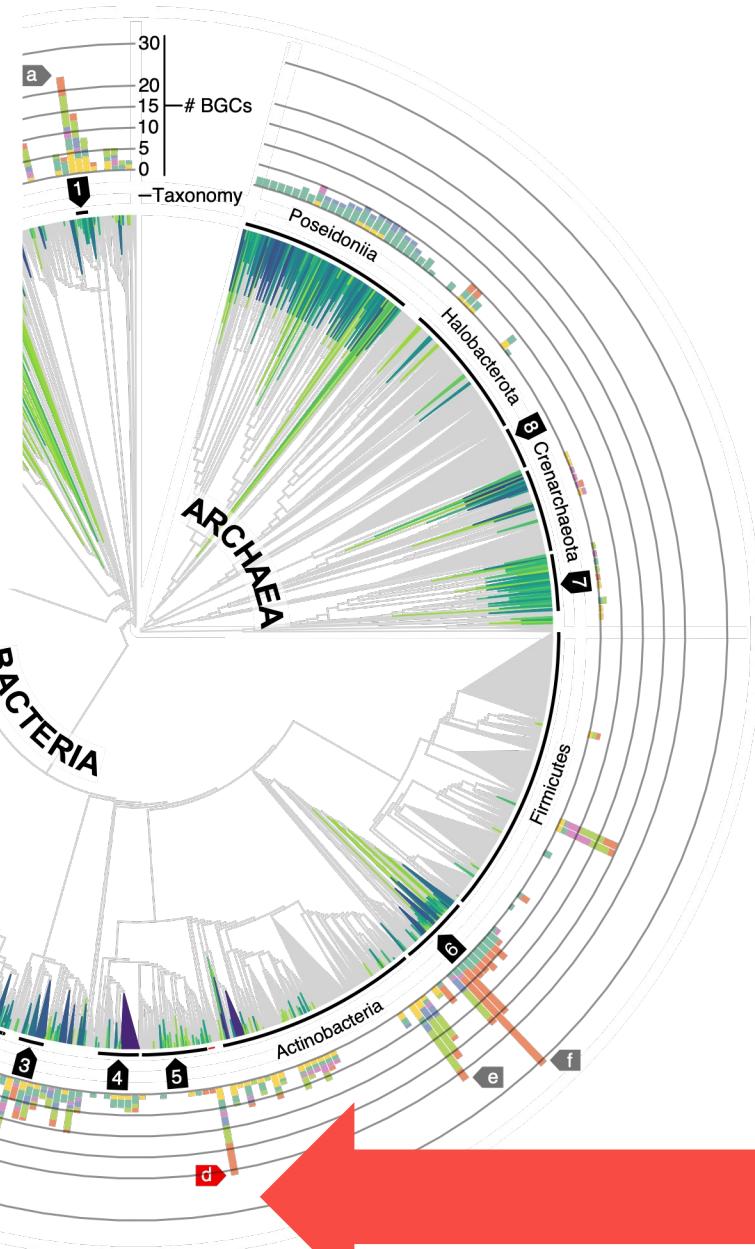
'Ca.
Eremiobacterota'

Eremio-who?



Biosynthetic gene cluster (BGC) type:

- █ RiPPs (Ribosomal Natural Products)
- █ Non-Ribosomal Peptide Synthases
- █ Type I Polyketide Synthases
- █ Type II & III Polyketide Synthases
- █ Terpenes
- █ Other

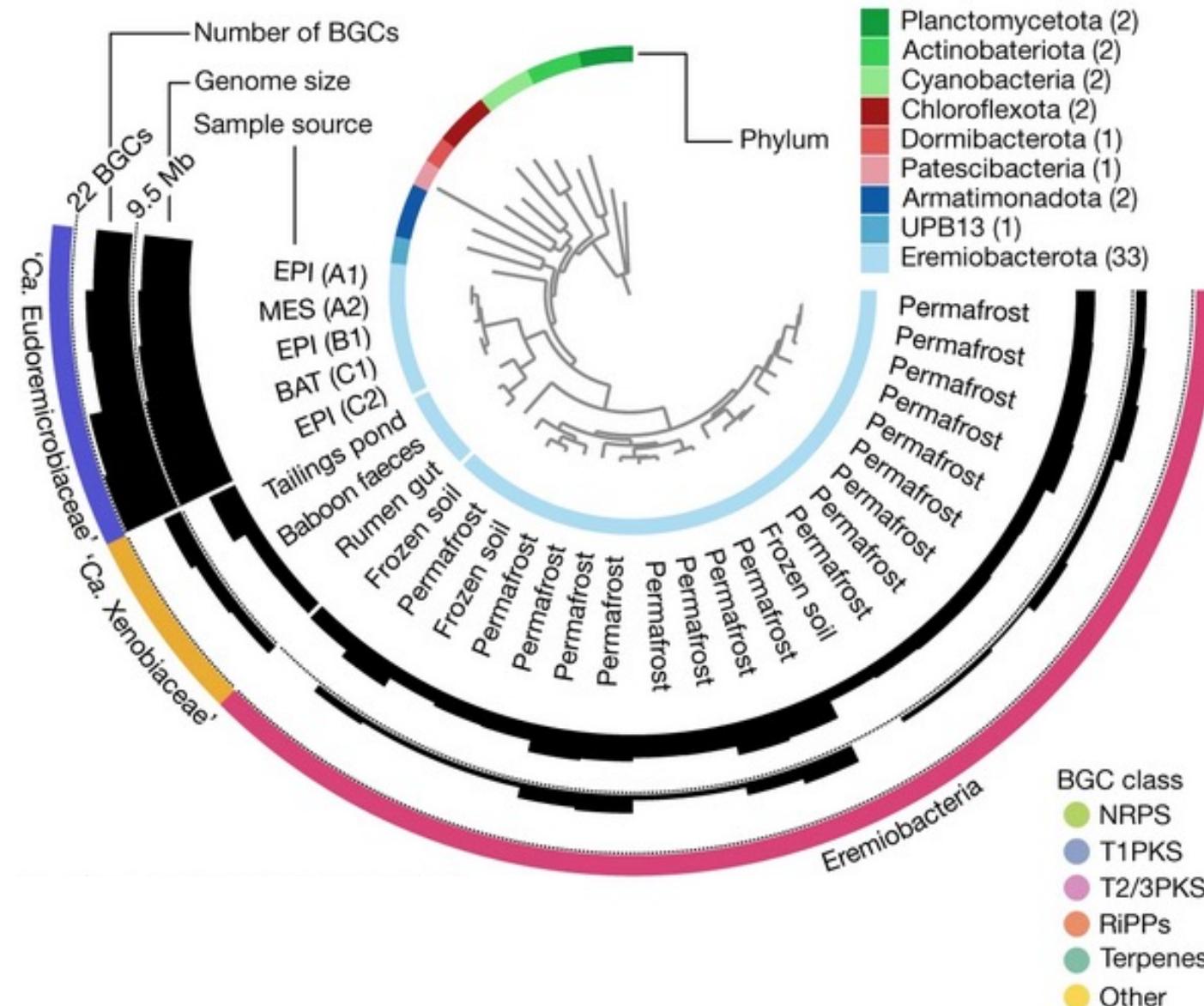


'Ca.
Eremiobacterota'

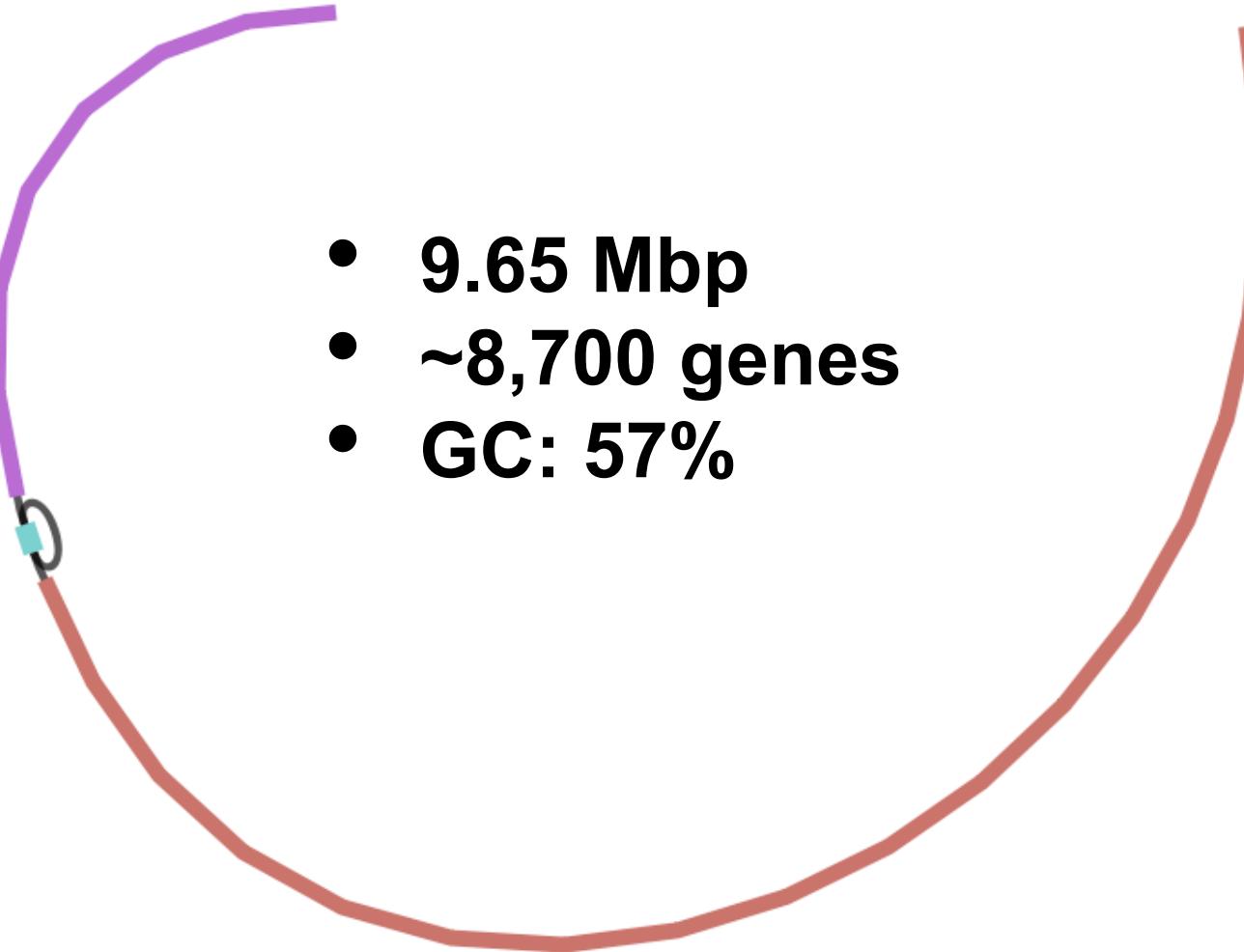
- Taxonomy**
- 1 Myxococcota
 - 2 Rhodobacteraceae
 - 3 Planctomycetota
 - 4 Chloroflexota
 - 5 Patescibacteria (CPR)
 - 6 Cyanobacteria
 - 7 Nanoarchaeota
 - 8 Euryarchaeota

- BGC-rich lineages (>15 BGCs)**
- a Sandaracinaceae
 - b Tistrella
 - c Planctomycetota
 - d Eremiobacterota
 - e Rhodococcus
 - f Synechococcus

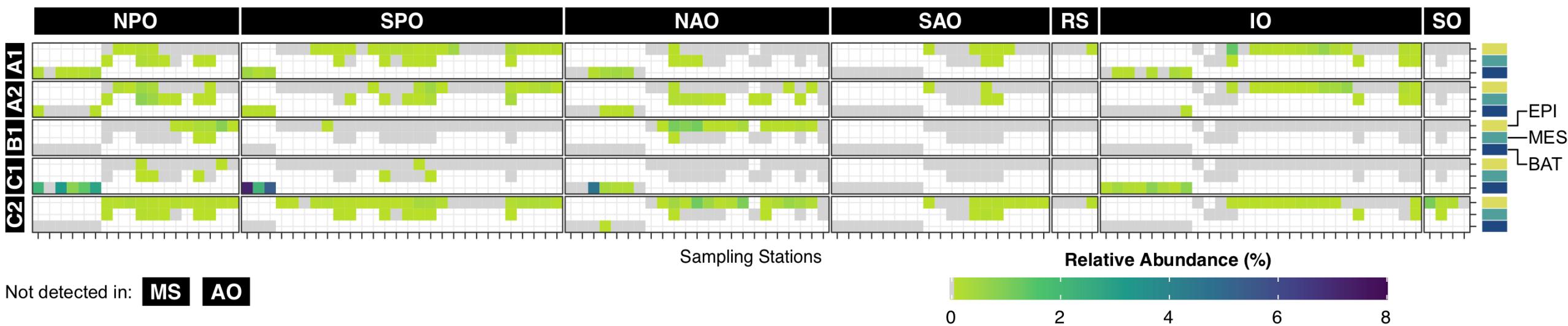
New ‘Ca. Eremiobacterota’ clade with enriched BGC content



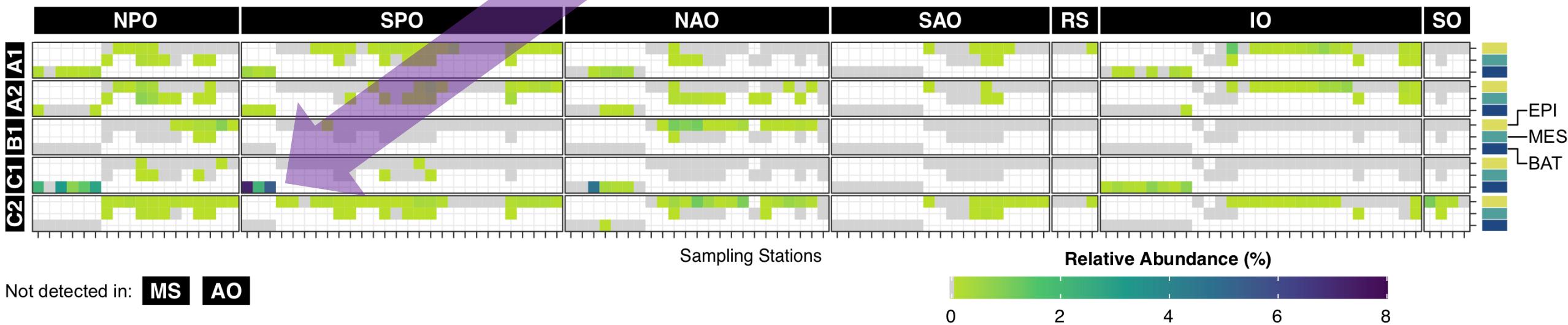
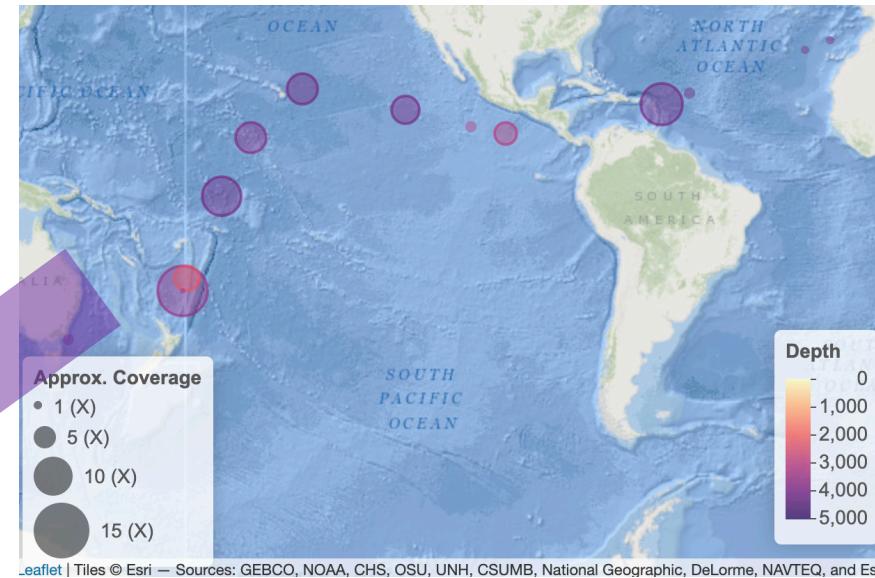
Closing the gaps: ultra-low-input (6 ng DNA!) for PacBio



Distribution of 'Ca. Eremiobacterota' across ocean layers



Distribution of 'Ca. Eremiobacterota' across ocean layers



C1: 'Ca. Eudoremycetum malaspini'
Locally accounting for up to 6% of ocean microbial communities

'Ca. Eudoremicrobium malaspini:' New biosynthetically-talented bacterial family

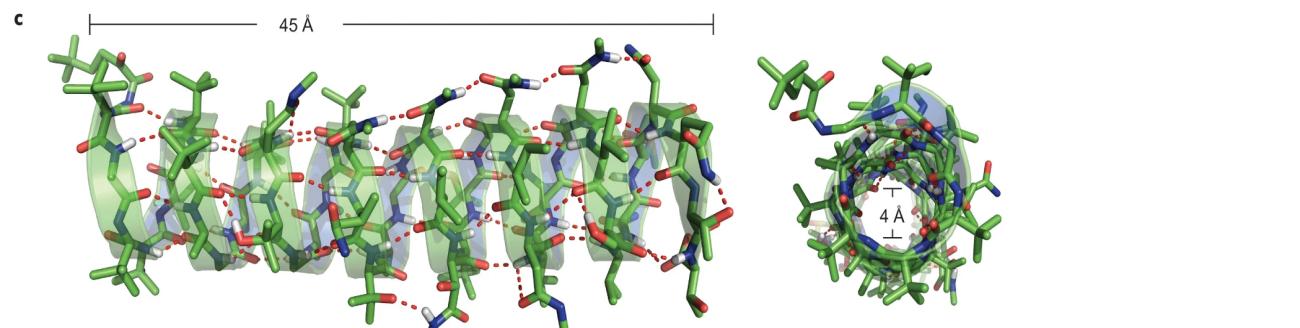
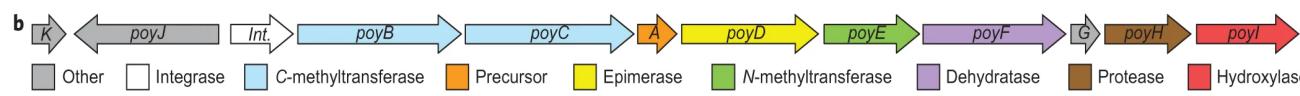
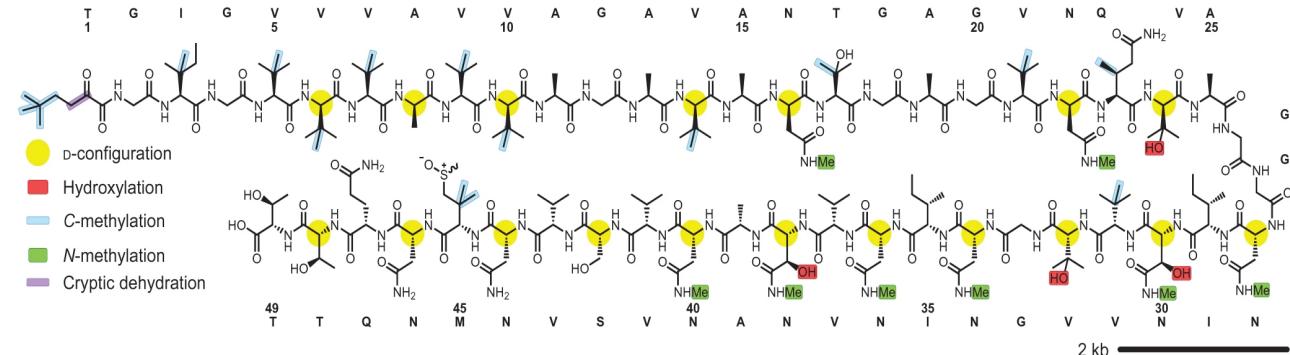
Region	Type
Region 1.1	NRPS-like <input checked="" type="checkbox"/>
Region 2.1	ectoine <input checked="" type="checkbox"/>
Region 2.2	T1PKS <input checked="" type="checkbox"/> , NRPS-like <input checked="" type="checkbox"/>
Region 3.1	proteusin <input checked="" type="checkbox"/> , bacteriocin <input checked="" type="checkbox"/>
Region 3.2	siderophore <input checked="" type="checkbox"/>
Region 7.1	arylpolyene <input checked="" type="checkbox"/>
Region 8.1	terpene <input checked="" type="checkbox"/>
Region 12.1	terpene <input checked="" type="checkbox"/>
Region 18.1	microviridin <input checked="" type="checkbox"/>
Region 18.2	hglE-KS <input checked="" type="checkbox"/> , T1PKS <input checked="" type="checkbox"/> , terpene <input checked="" type="checkbox"/>
Region 25.1	NRPS-like <input checked="" type="checkbox"/>
Region 26.1	bacteriocin <input checked="" type="checkbox"/>
Region 29.1	NRPS-like <input checked="" type="checkbox"/>
Region 32.1	phosphonate <input checked="" type="checkbox"/>
Region 34.1	proteusin <input checked="" type="checkbox"/> , bacteriocin <input checked="" type="checkbox"/>
Region 36.1	microviridin <input checked="" type="checkbox"/>
Region 37.1	terpene <input checked="" type="checkbox"/>
Region 52.1	T1PKS <input checked="" type="checkbox"/>
Region 54.1	lanthipeptide <input checked="" type="checkbox"/> , proteusin <input checked="" type="checkbox"/> , bacteriocin <input checked="" type="checkbox"/>
Region 62.1	T3PKS <input checked="" type="checkbox"/>
Region 75.1	bacteriocin <input checked="" type="checkbox"/>

'Candidatus Eudoremicrobium'
(Eu.do.re.mi.cro'bi.um; N.L.
fem. n. Eudore, the Nereid, sea
deities in Greek mythology, of
fine gifts from the sea)



'Ca. Eudoremicrobium' encodes highest # of proteusin BGCs

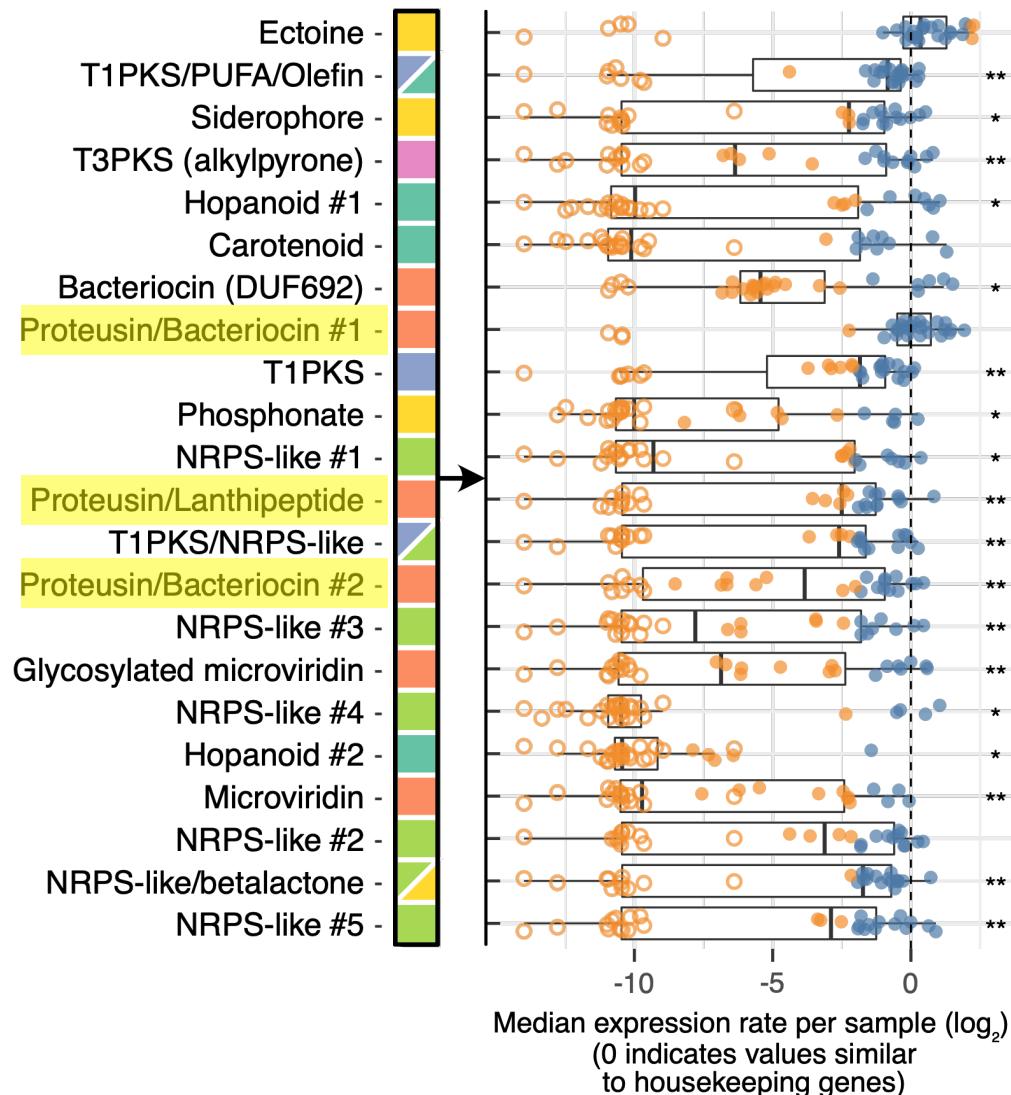
Region	Type
Region 1.1	NRPS-like ↗
Region 2.1	ectoine ↗
Region 2.2	T1PKS ↗, NRPS-like ↗
Region 3.1	proteusin ↗, bacteriocin ↗
Region 3.2	siderophore ↗
Region 7.1	arylpolyene ↗
Region 8.1	terpene ↗
Region 12.1	terpene ↗
Region 18.1	microviridin ↗
Region 18.2	hglE-KS ↗, T1PKS ↗, terpene ↗
Region 25.1	NRPS-like ↗
Region 26.1	bacteriocin ↗
Region 29.1	NRPS-like ↗
Region 32.1	phosphonate ↗
Region 34.1	proteusin ↗, bacteriocin ↗
Region 36.1	microviridin ↗
Region 37.1	terpene ↗
Region 52.1	T1PKS ↗
Region 54.1	lantipeptide ↗, proteusin ↗, bacteriocin ↗
Region 62.1	T3PKS ↗
Region 75.1	bacteriocin ↗



A diverse and actively expressed BGC repertoire

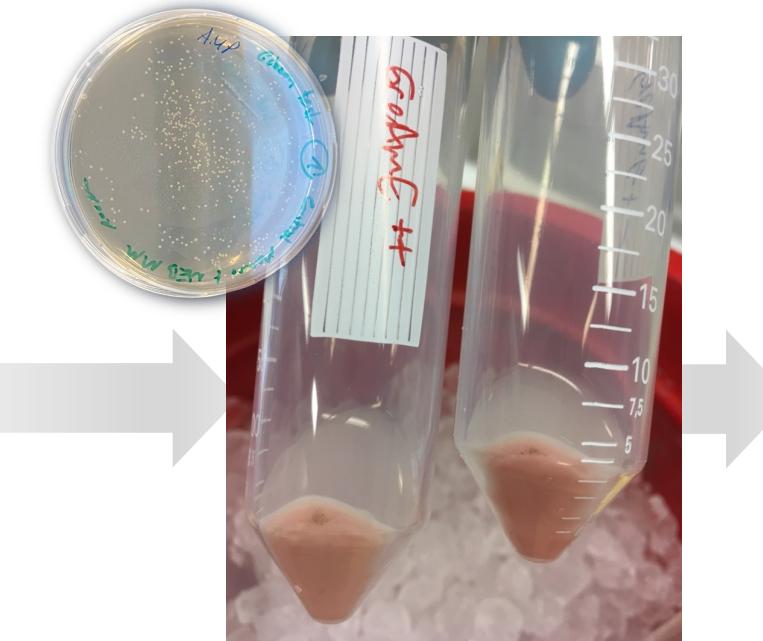
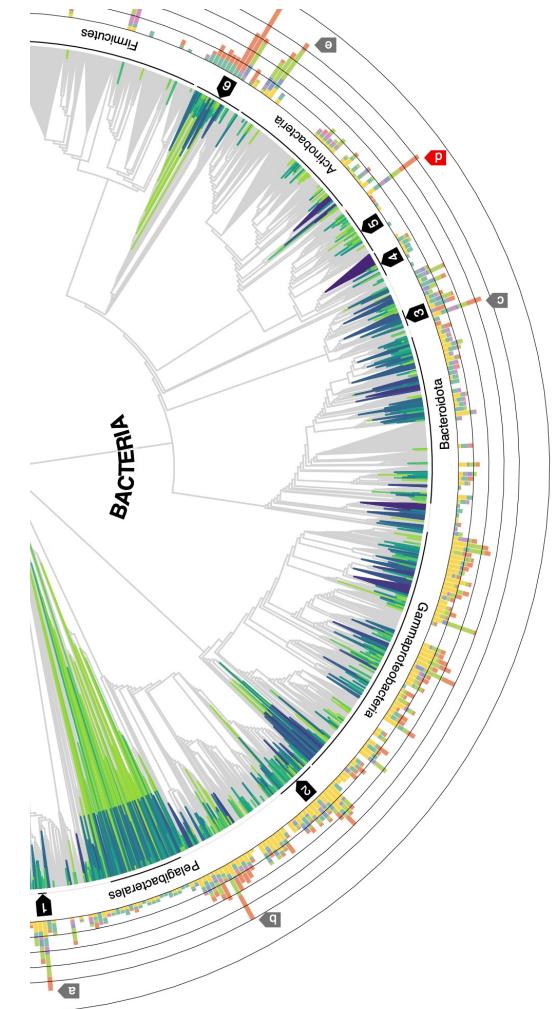
BGC class

- RiPPs (Ribosomal Natural Products)
- Non-Ribosomal Peptide Synthases
- Type I Polyketide Synthases
- Type II & III Polyketide Synthases
- Terpenes
- Other

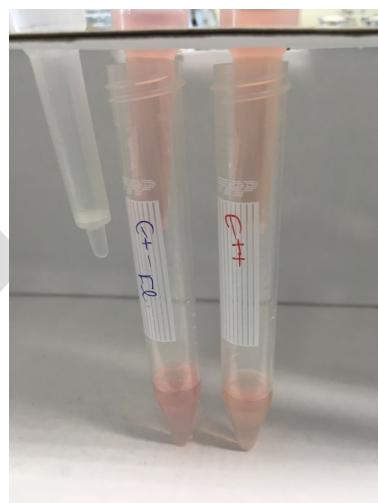


'Ca. Eudoremicrobium' proteusin BGC: characterization

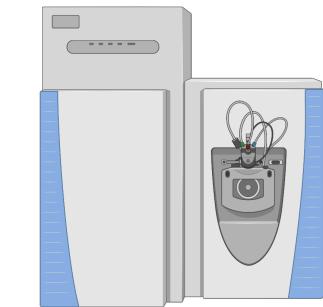
BGCs from metagenome-assembled genomes



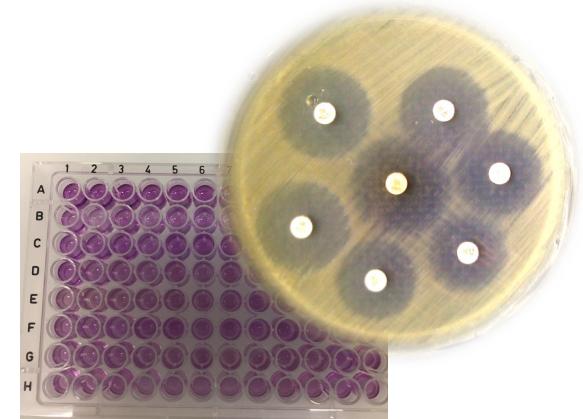
Clone and express in heterologous hosts



Purify peptides



Chemistry: mass spectrometry & NMR

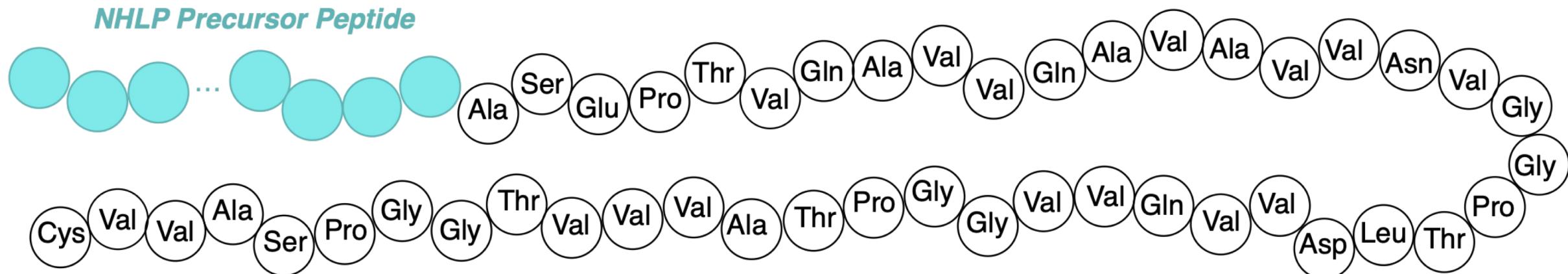


Function e.g., bioactivity

46 aa peptide encoded in novel ‘Ca. Eudoremicobium’ BGC

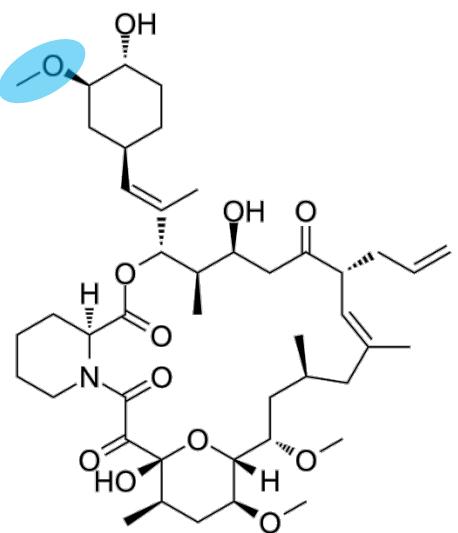


NHLP leader - ASEPTVQAVVQAVAVVNGGPTLDVVQVVGGPTAVVVTGGPSAVVC

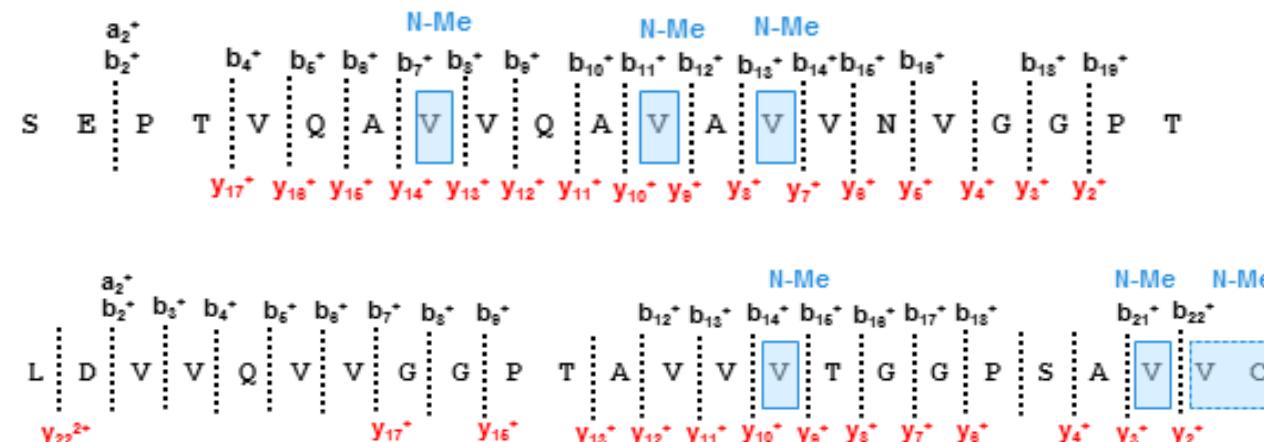


EreM: new peptide backbone *N*-methyltransferase family

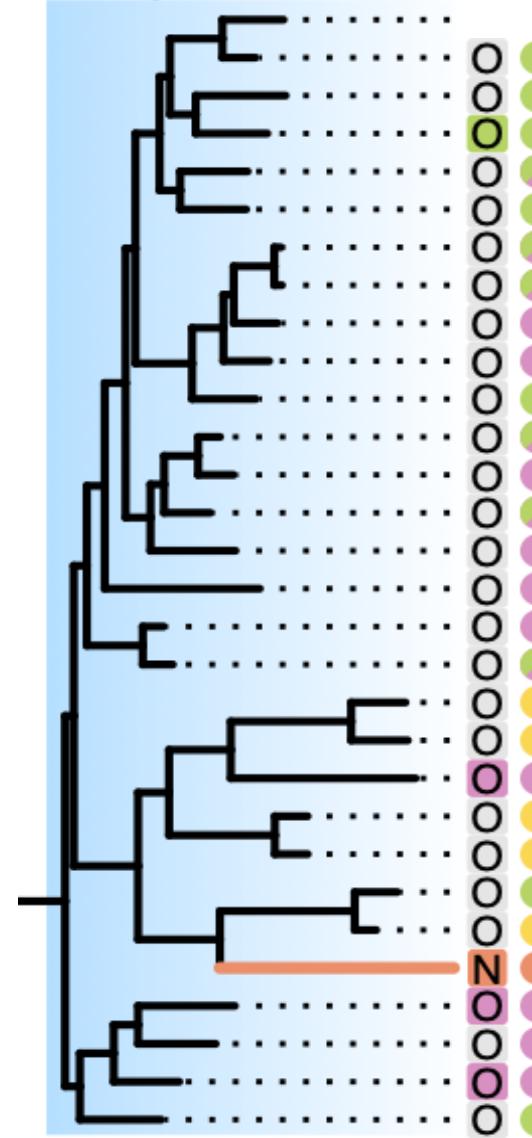
Before:



This study:

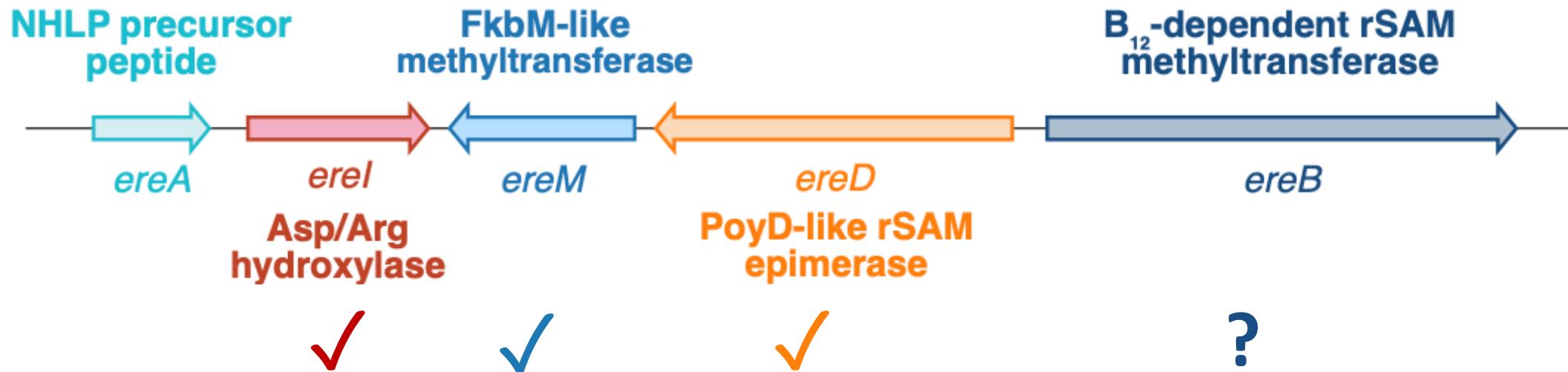


FkbM-like
methyltransferases



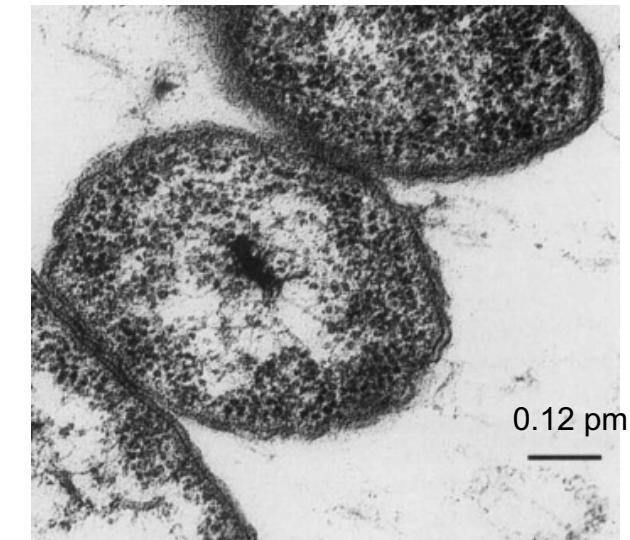
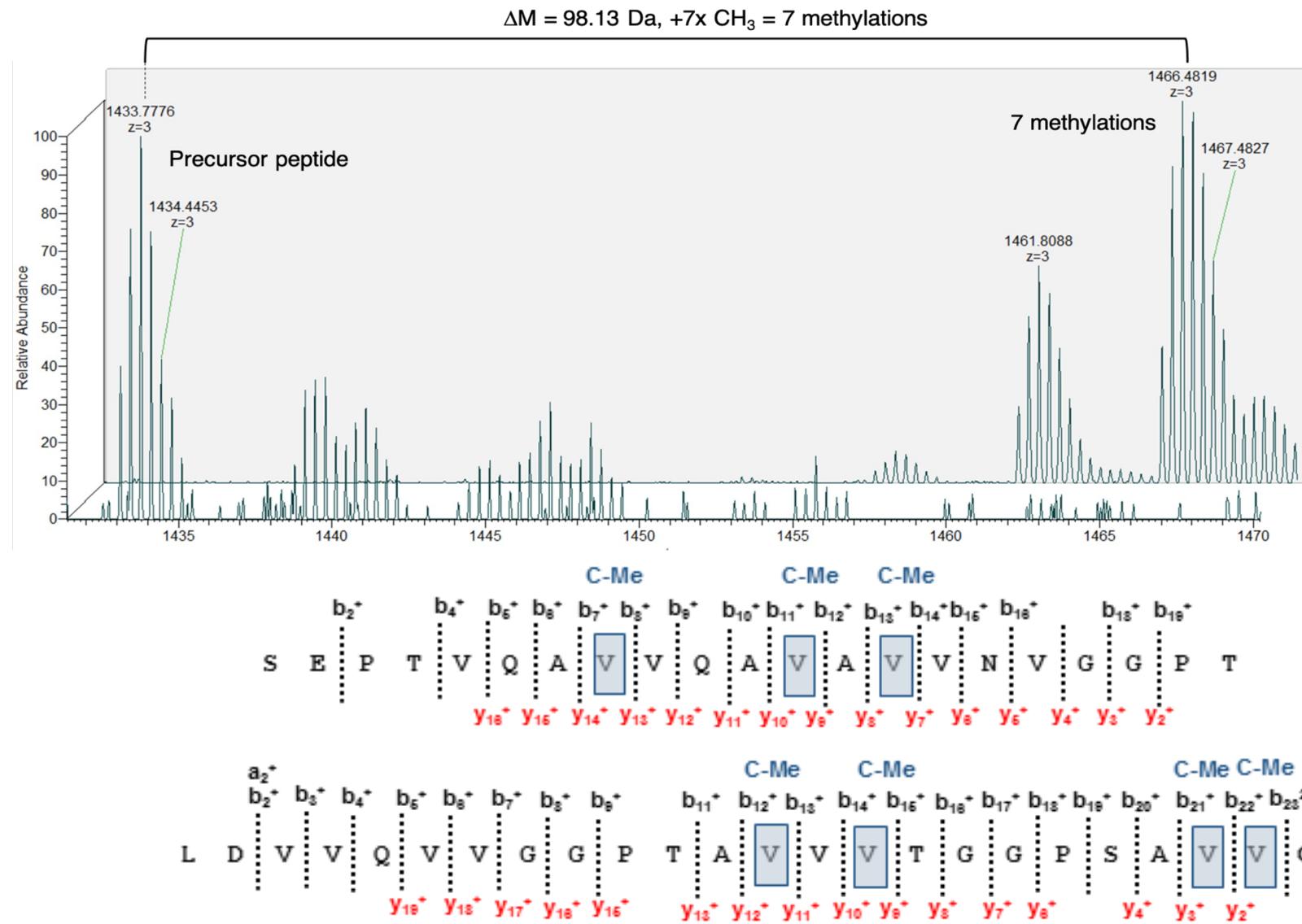
- RiPPs (Ribosomal Natural Products)
- Non-Ribosomal Peptide Synthases
- Type I Polyketide Synthases
- Type II & III Polyketide Synthases
- Terpenes
- Other

Overview of experimental results in *E. coli*



- 3 enzymes active in *E. coli* install post-translational modifications in ‘Ca. Eremiobacterota’ peptide core
 - 1 hydroxyl group
 - Up to 6 N-methylations
 - 7 D-amino acids
 - Cobalamin-dependent enzyme not active in *E. coli*

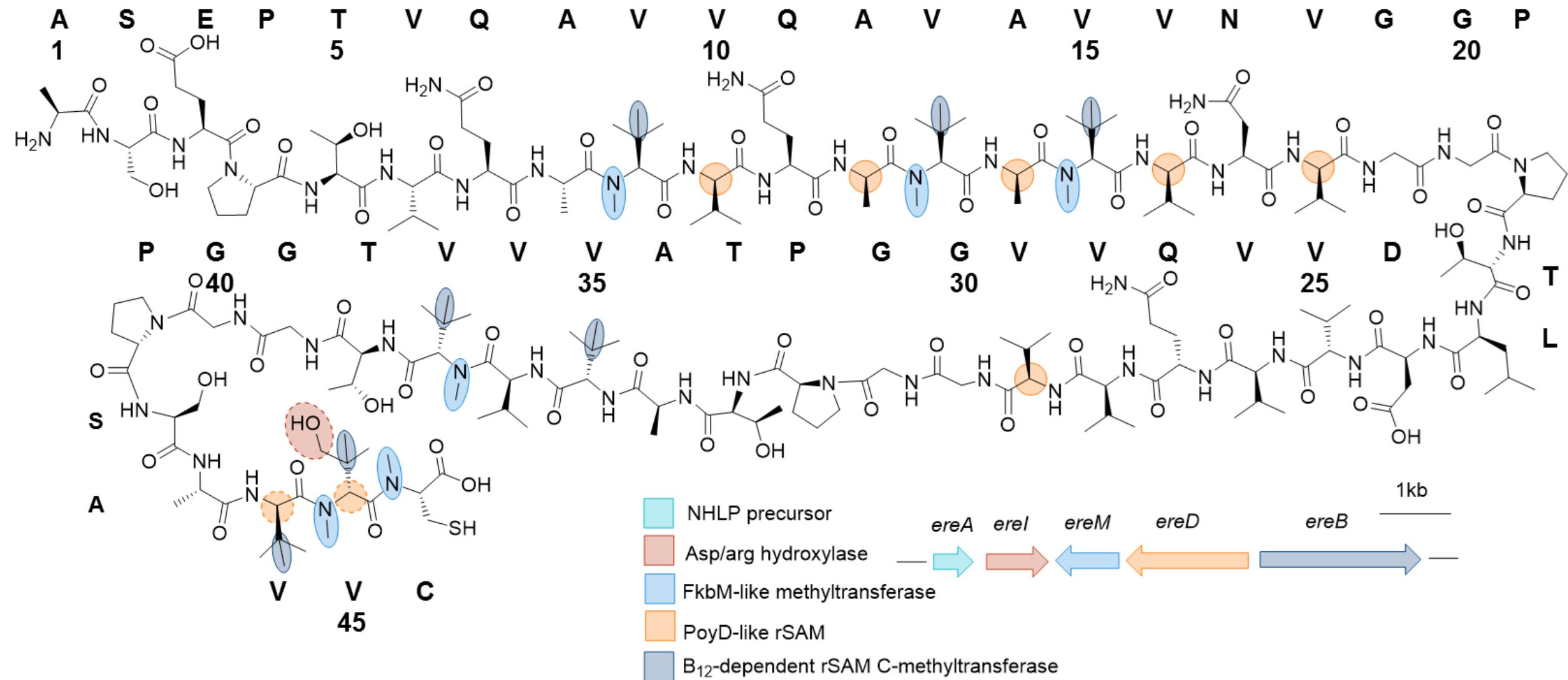
Heterologous cluster expression in wastewater denitrifier: *Microvirgula aerodenitrificans*



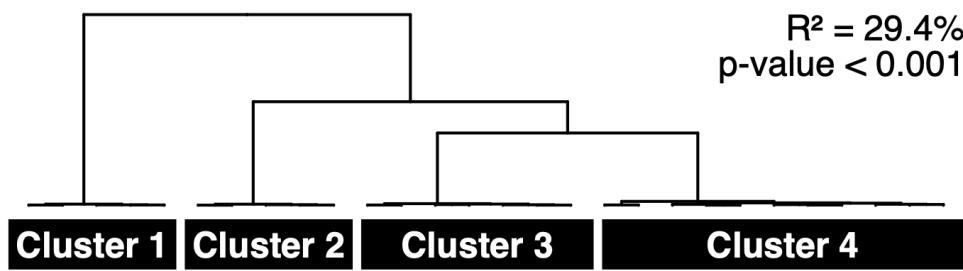
Microvirgula aerodenitrificans
Patureau et al. 1993

✓ Complete pathway for cobalamin biosynthesis

Up to 21 modifications to a 46 aa peptide natural product



Case study: *Ca. Eudoregemicrobiaceae* Ecology with *in situ* metatranscriptomics

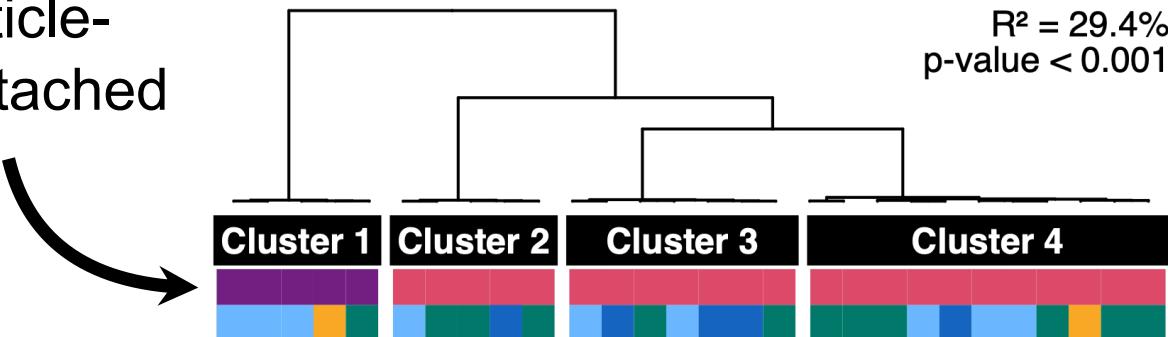


Clustering based on the expression levels of all genes

29.4% of transcriptome variance explained by 4 discrete clusters

Case study: Ca. Eudoremicrobiaceae Ecology with *in situ* metatranscriptomics

Particle-
attached



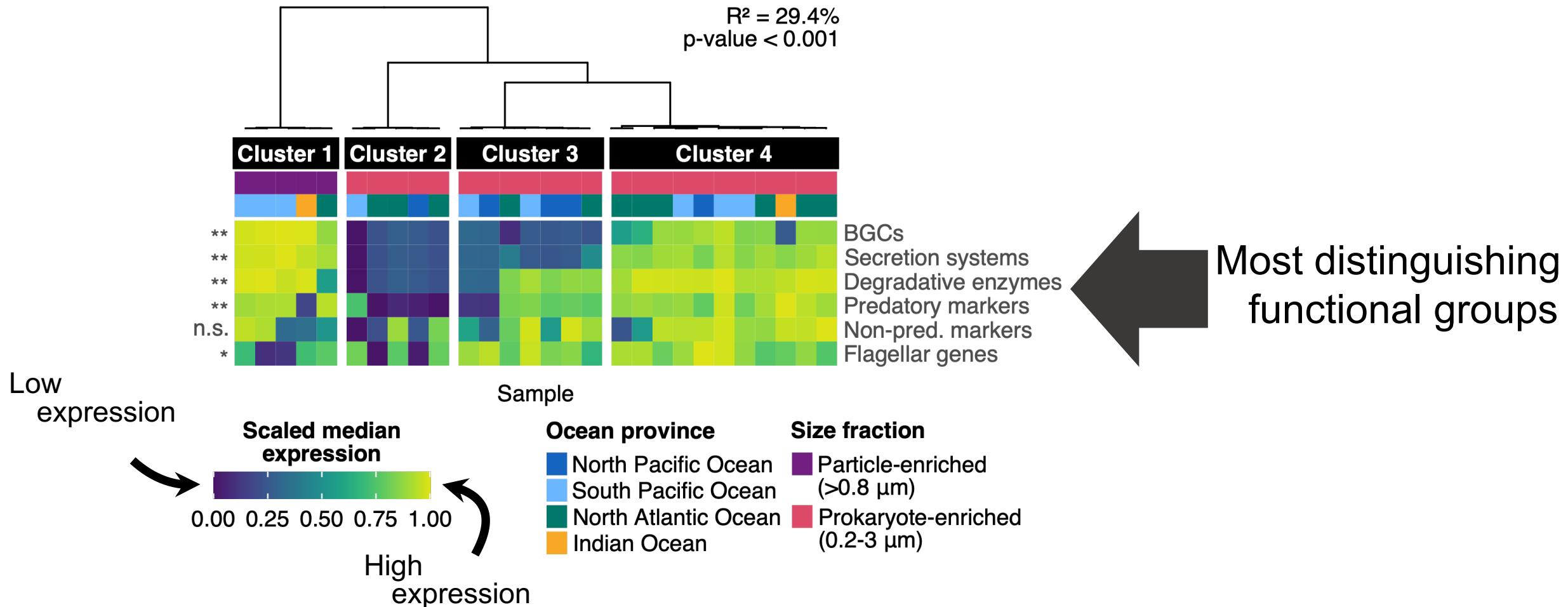
$R^2 = 29.4\%$
 $p\text{-value} < 0.001$

29.4% of transcriptome variance
explained by 4 discrete clusters
(distinct transcriptional states)

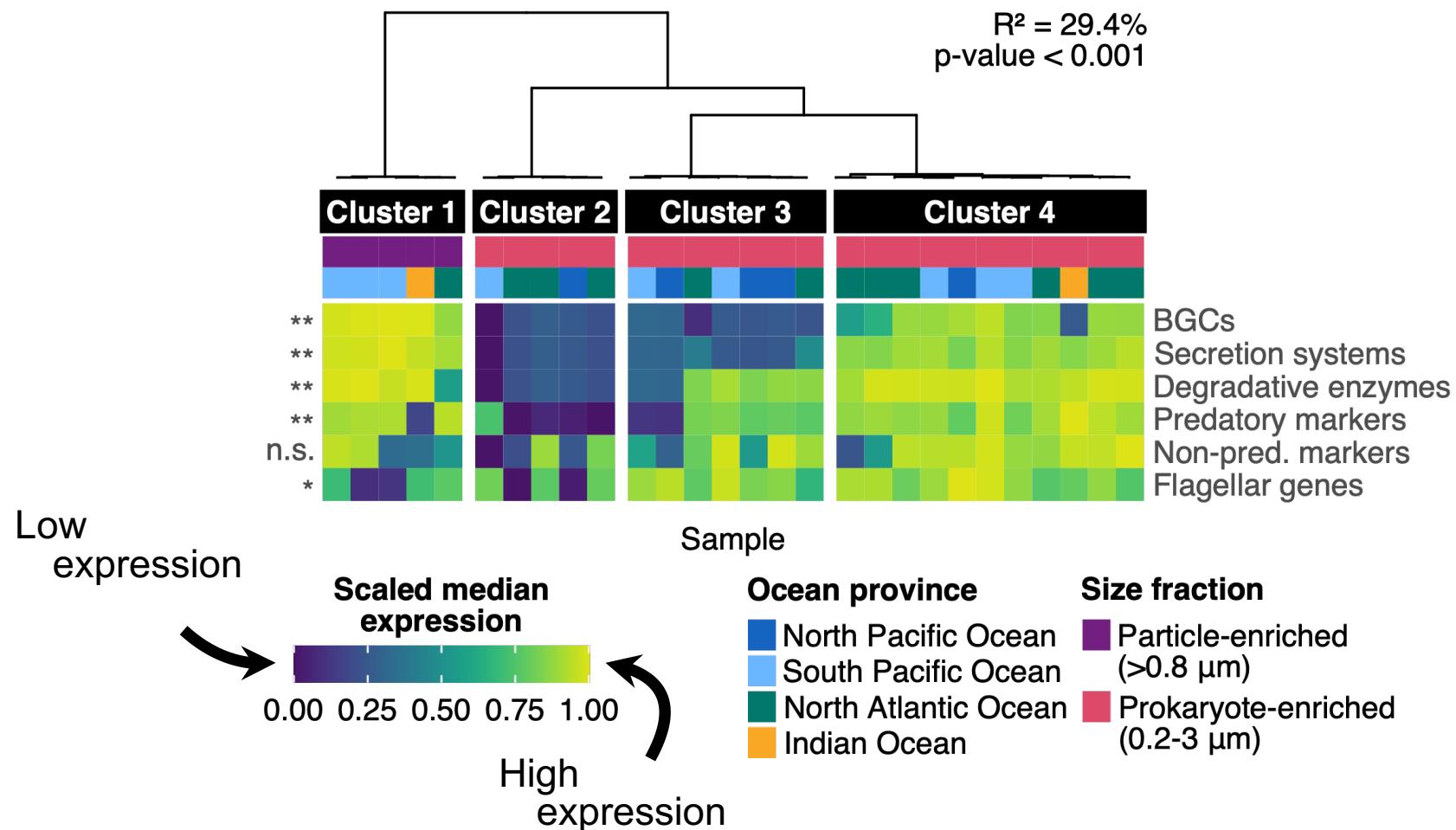
Ocean province	Size fraction
North Pacific Ocean	Particle-enriched ($>0.8 \mu\text{m}$)
South Pacific Ocean	Prokaryote-enriched (0.2-3 μm)
North Atlantic Ocean	
Indian Ocean	

Case study: 'Ca. Eremiobacterota'

Ecology with *in situ* metatranscriptomics



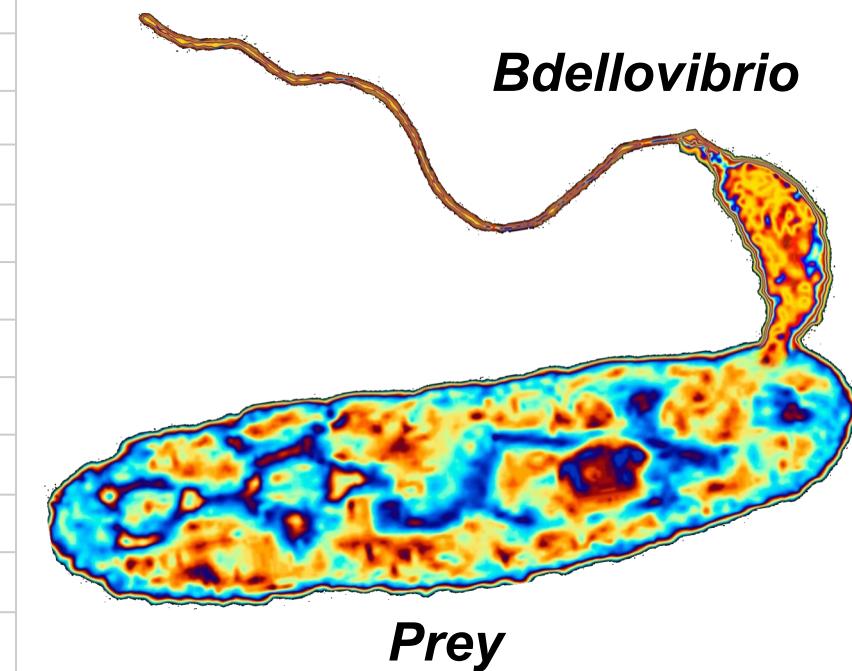
BGCs to support a predatory lifestyle?



- Large BGC diversity
- Diverse secretion systems
- Degradative enzymes
- Motility

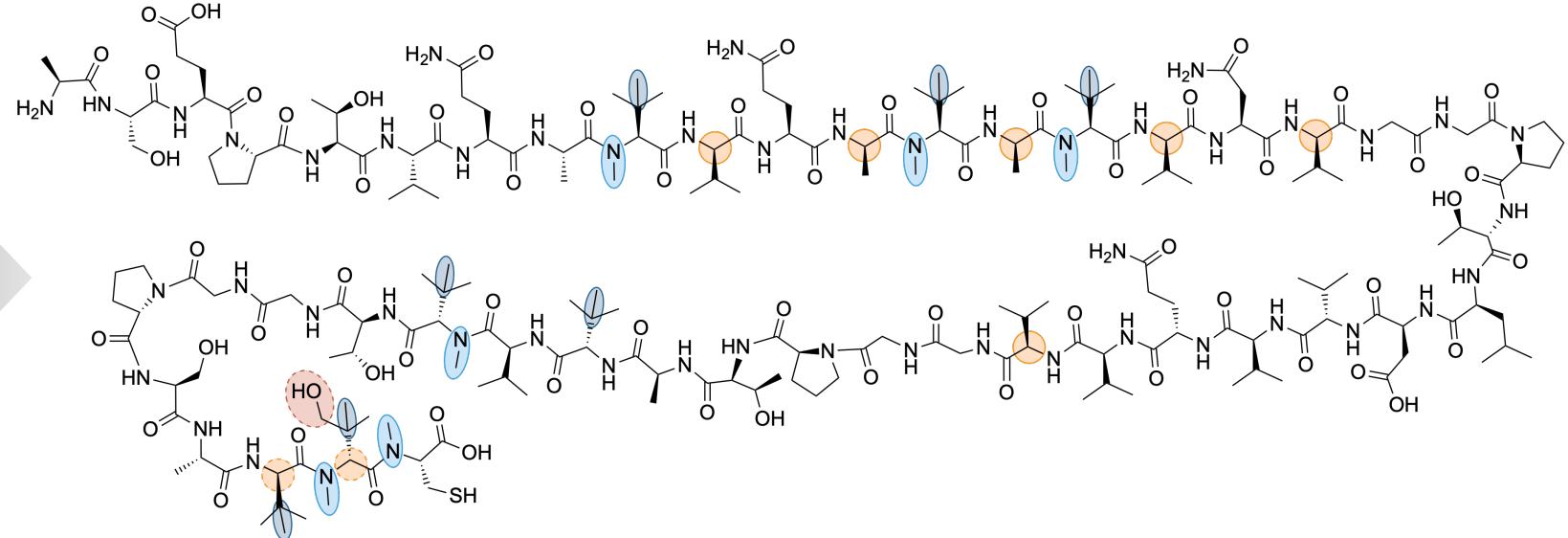
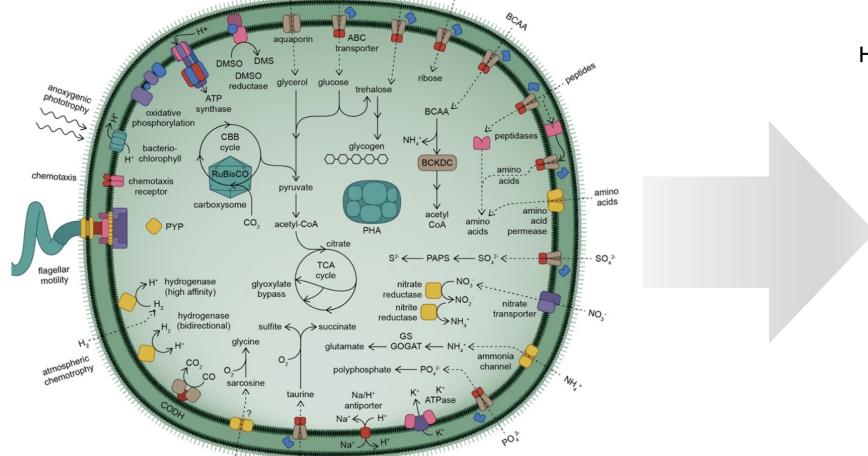
Food for thought: 'Ca. Eudoremicrobium' predatory potential?

<u>Genus</u>	<u>Taxon</u>	<u>Predatory Index</u>
'Ca. Eudoremicrobium malaspini'	Eremiobacterota	15
'Ca. Eudoremicrobium taraoceanii'	Eremiobacterota	14
<i>Bdellovibrio</i> JSS	δ-proteobacteria	13
<i>Saprosira</i>	Bacteroides	12
<i>Bdellovibrio HD100</i>	δ-proteobacteria	11
<i>Bacteriovorax</i>	δ-proteobacteria	10
'Ca. Autonoemicrobium septentrionale'	Eremiobacterota	8
'Ca. Amphitoemicrobium indianii'	Eremiobacterota	7
<i>Herpetosiphon</i>	Chloroflexi	6
<i>Flavobacterium</i>	Bacteroides	6
<i>Myxococcus</i>	δ-proteobacteria	5
'Ca. Amphitoemicrobium mesopelagicum'	Eremiobacterota	4
<i>Stigmatella</i>	δ-proteobacteria	4
<i>Micavibrio</i>	α-proteobacteria	3
<i>Sorangium</i>	δ-proteobacteria	2
<i>Cytophaga</i>	Bacteroides	2
<i>Enttheonella</i> (TSY1)	Tectomicrobia	-8



Summary: biosynthetic potential of the global ocean

- First experimental evidence for ‘Ca. Eremiobacterota’ biosynthetic potential and natural product repertoire
- What else can it do? What is the biotransformation potential?



Moving from the global ocean...to Swiss streams

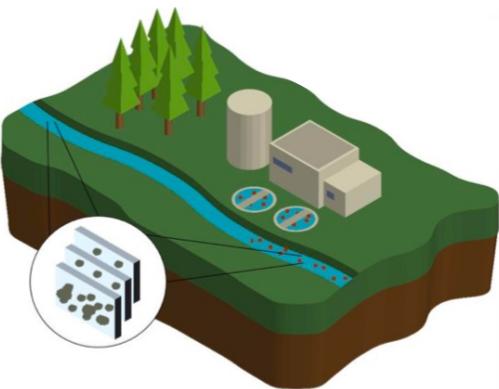
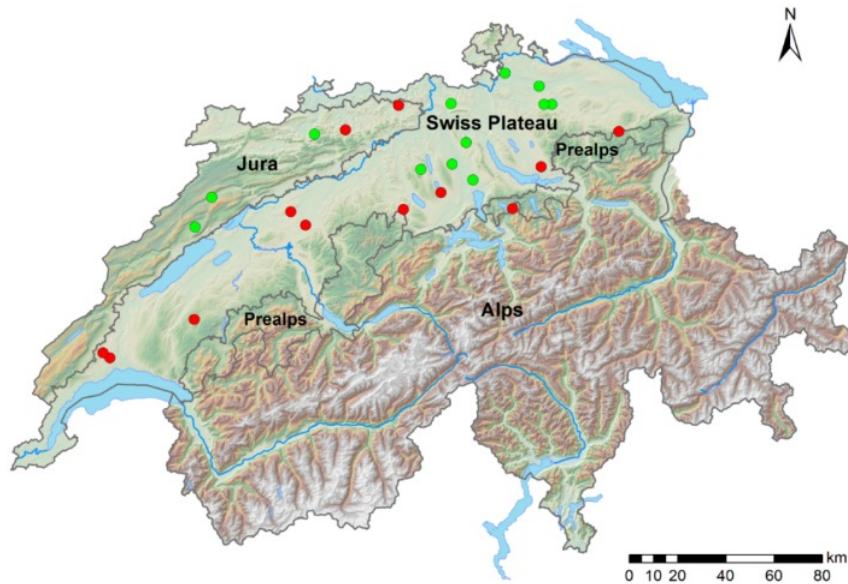


Photo Credit: Niklas Ferenc Trottmann

Kathrin Fenner



Michael Zimmermann



Research Question
How does wastewater effluent affect the biotransformation potential of stream biofilms?

Questions?

