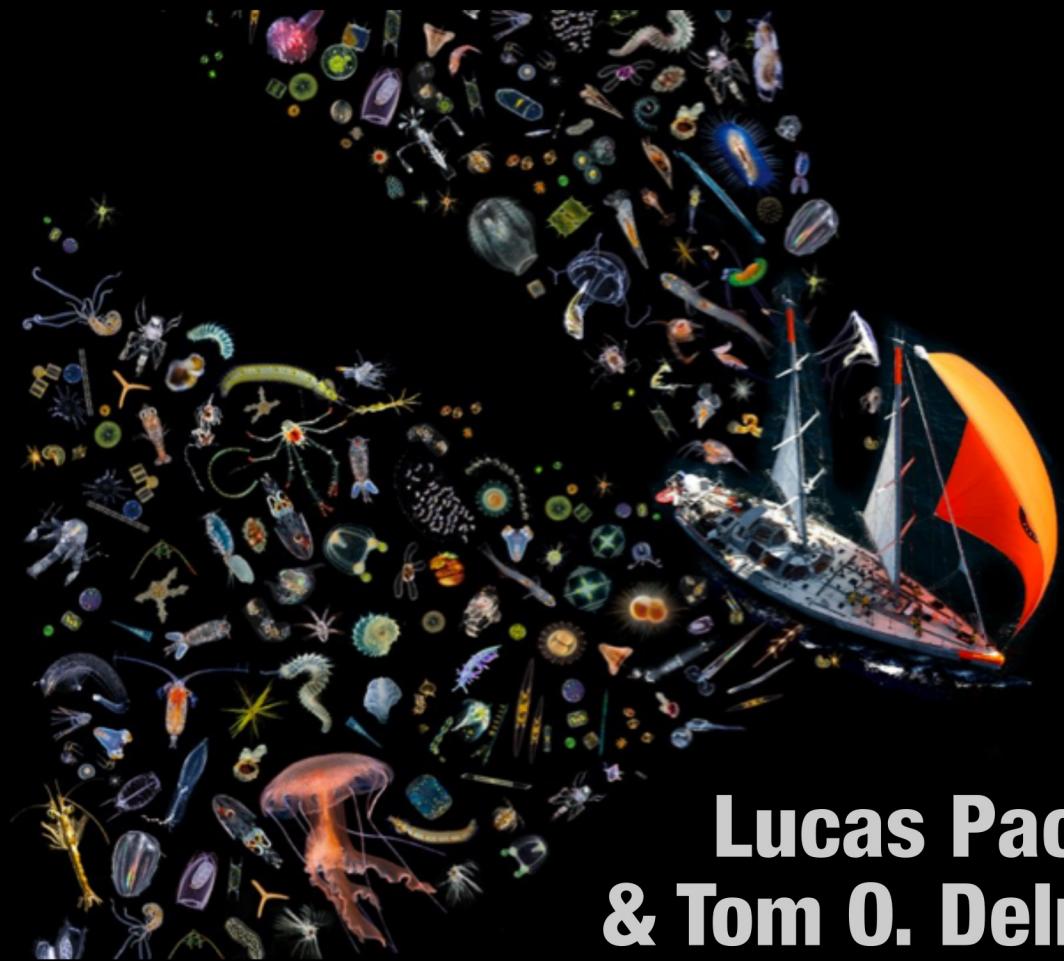
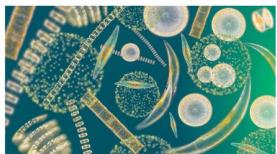
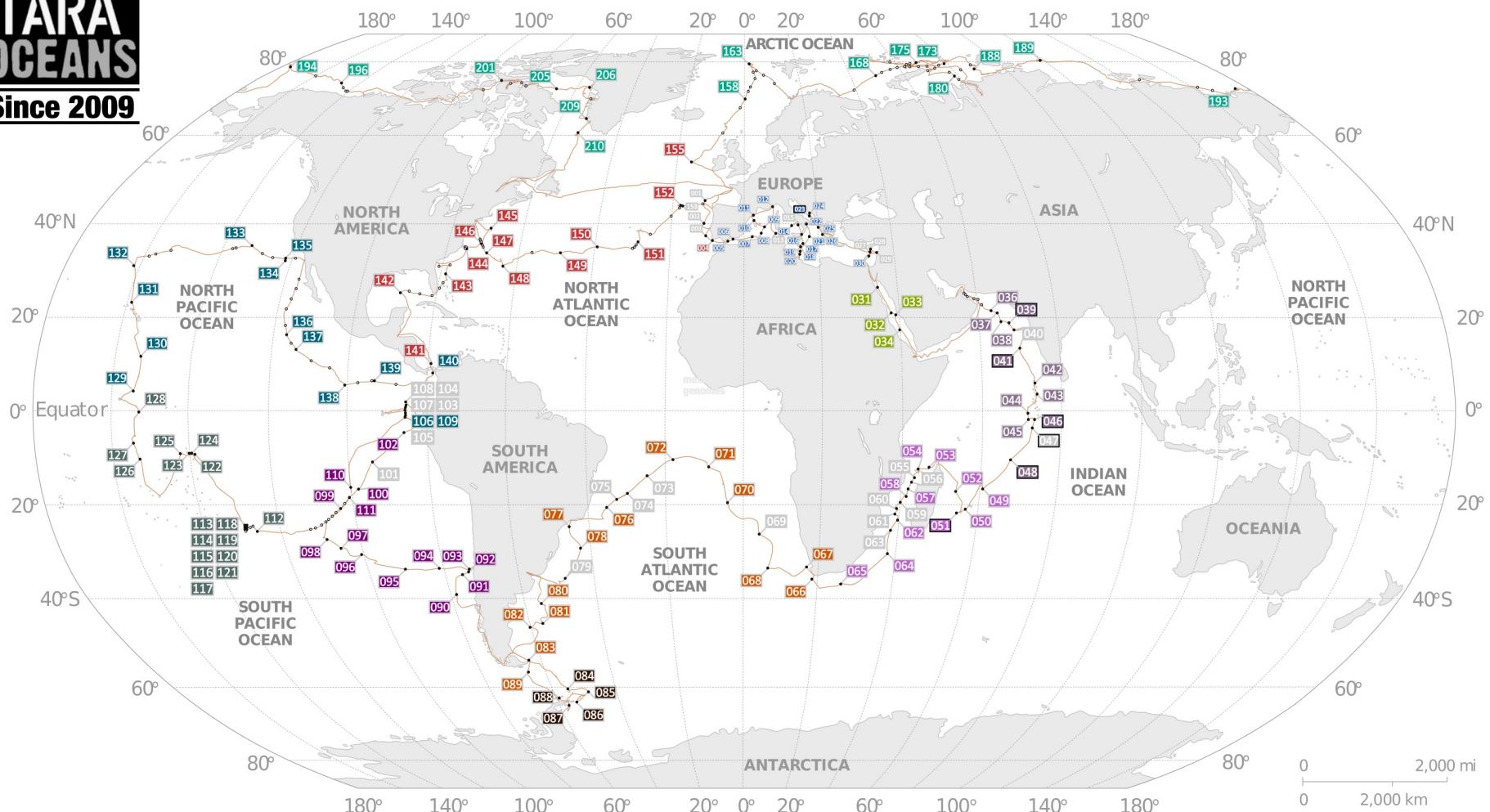


Plankton genomics with Tara Oceans



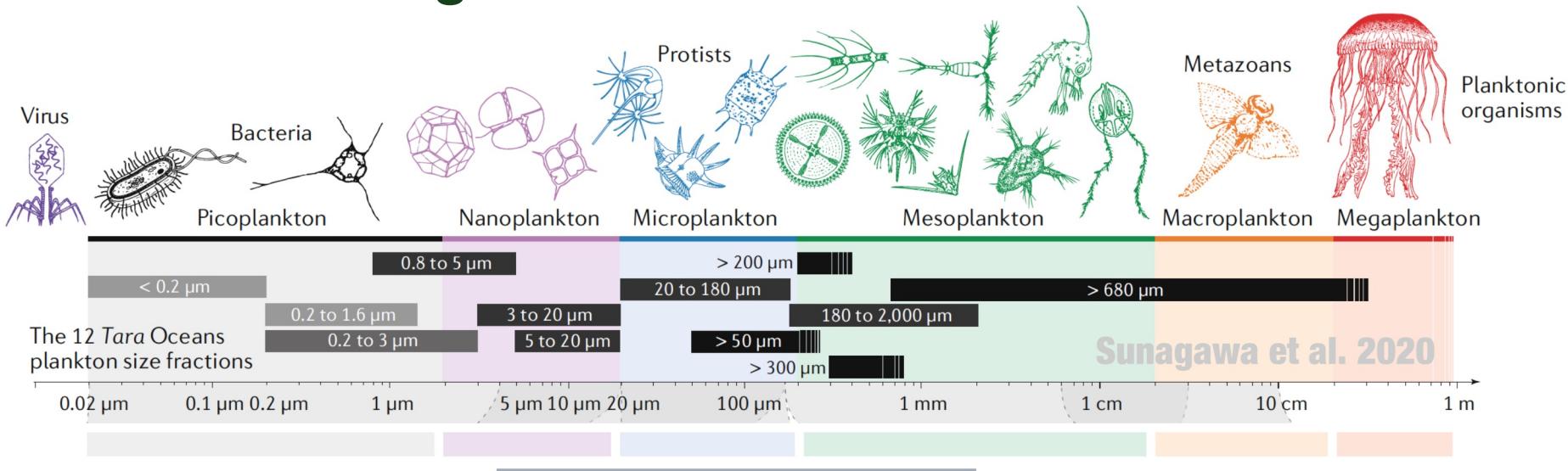
Lucas Paoli
& Tom O. Delmont

Sampling the oceans to understand plankton



Equivalent of 10,000 human genomes sequenced

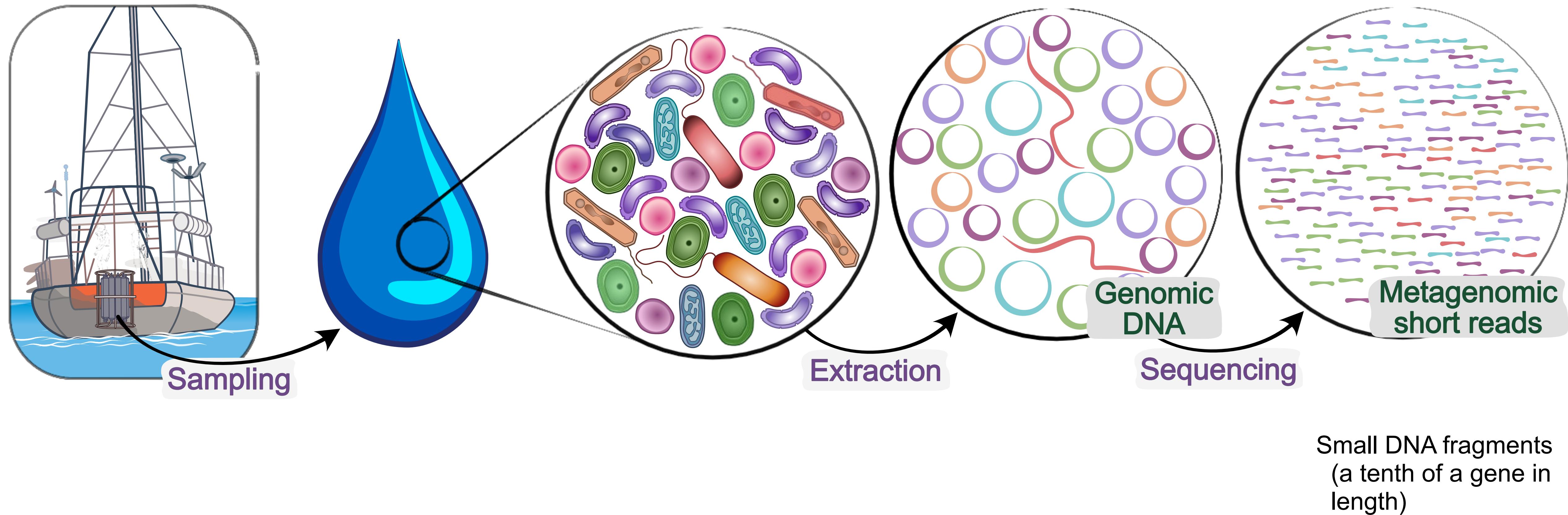
~300 billion metagenomic reads



A considerable opportunity to better understand the genomic diversity of plankton

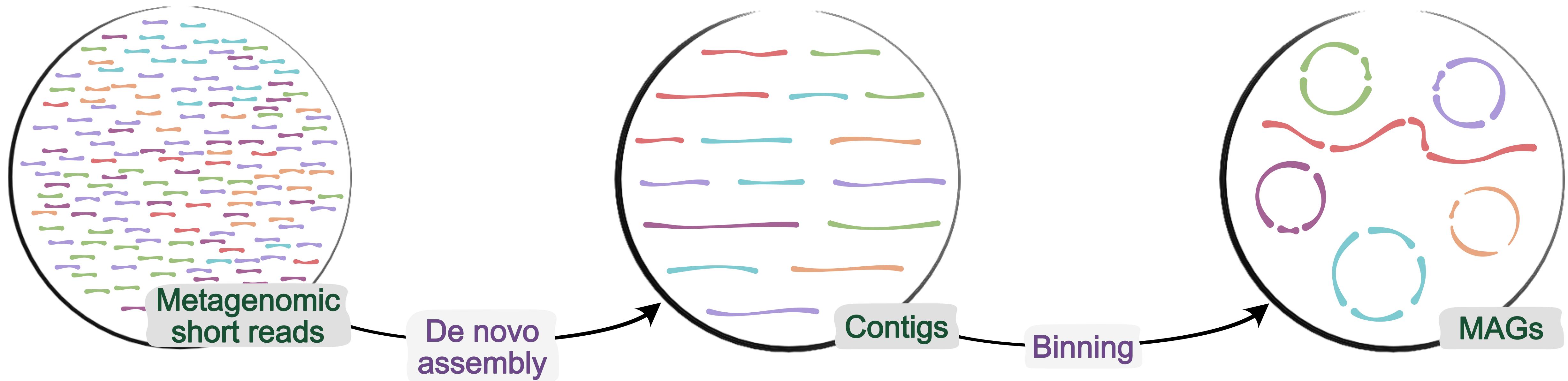
We need to solve a complex metagenomic puzzle

Metagenomics: community-level genomics

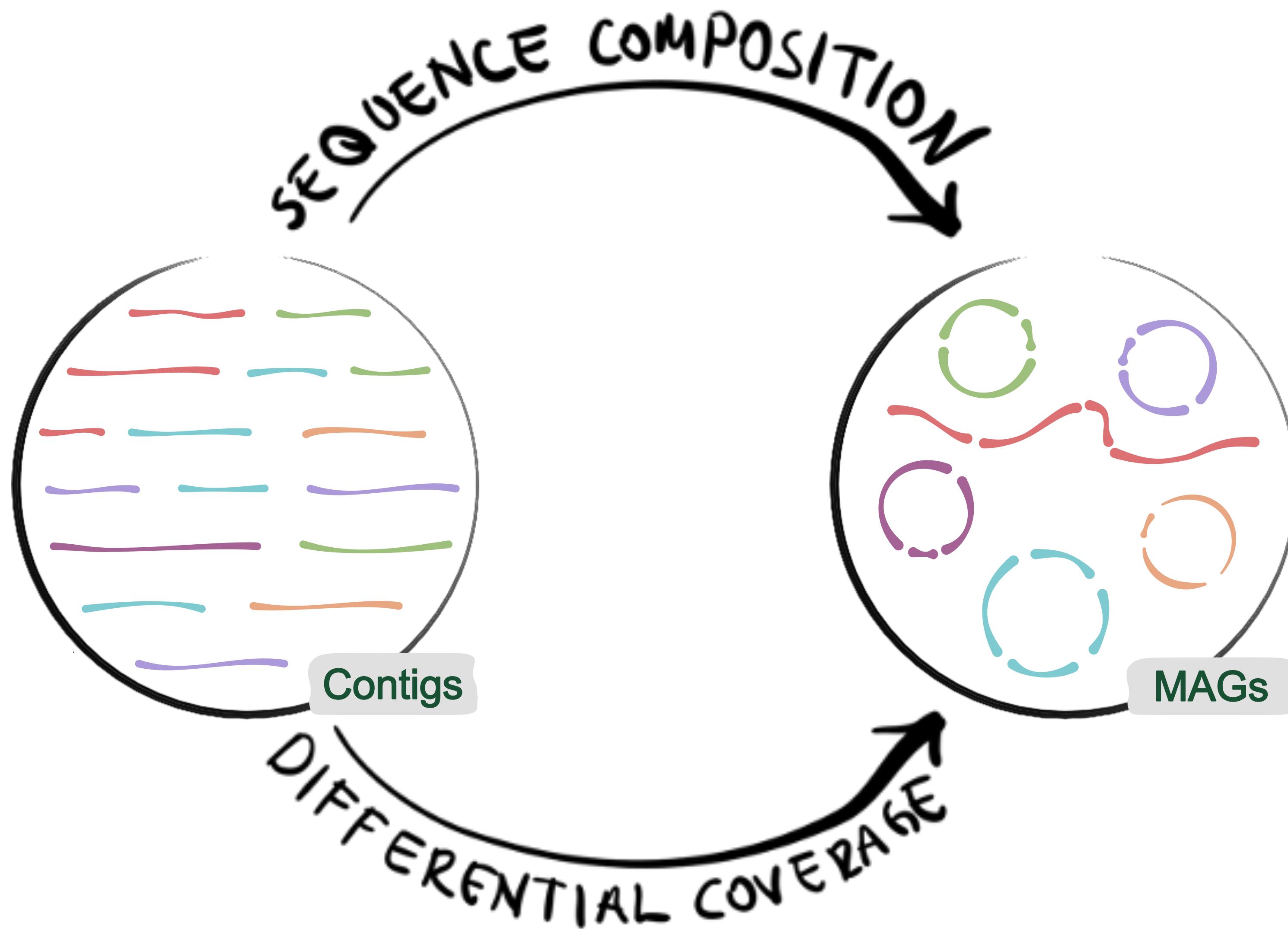


Small DNA fragments
(a tenth of a gene in
length)

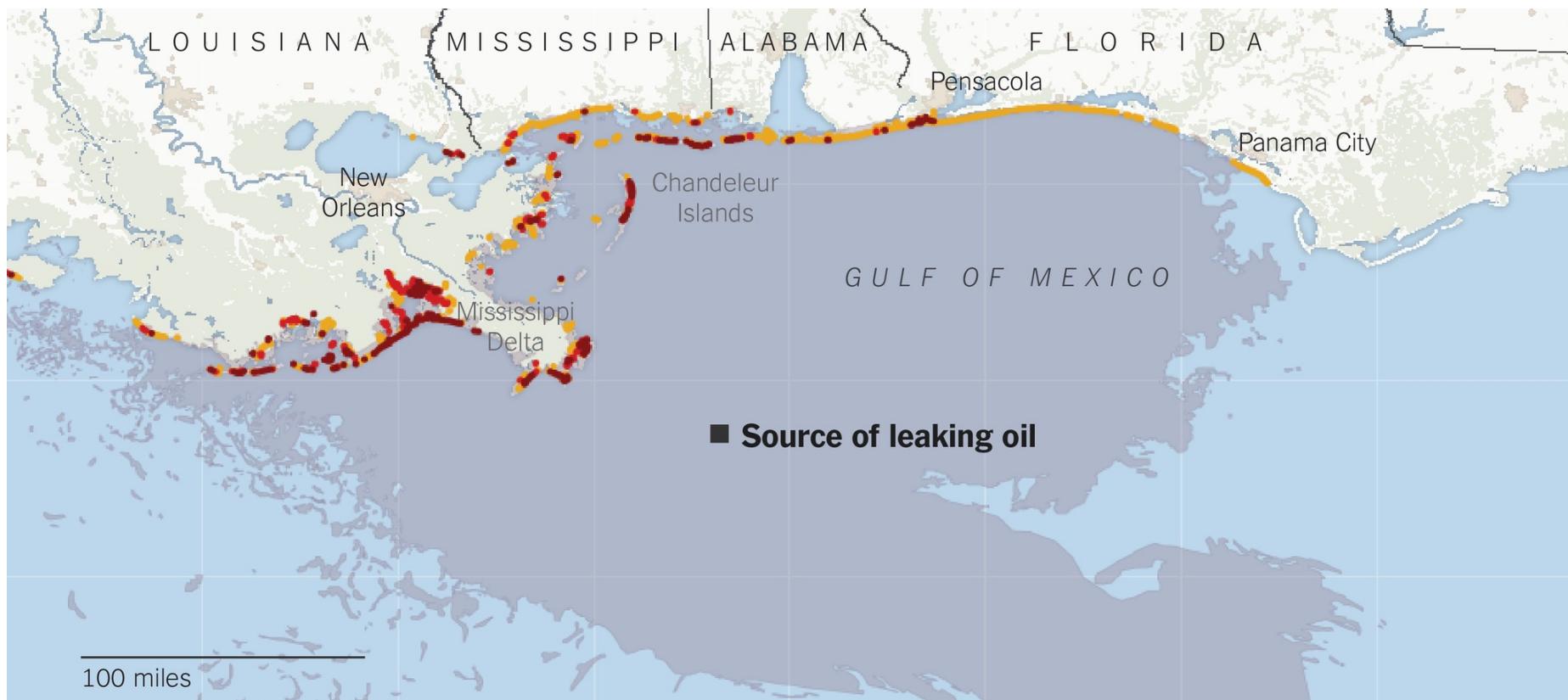
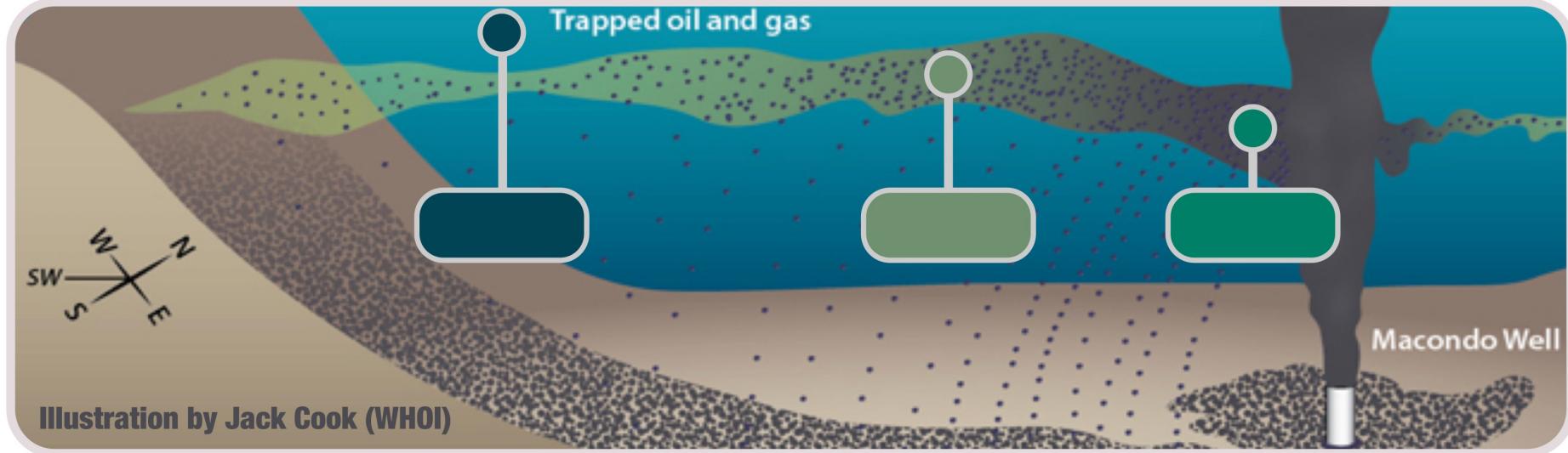
Cultivation-independent genome reconstruction

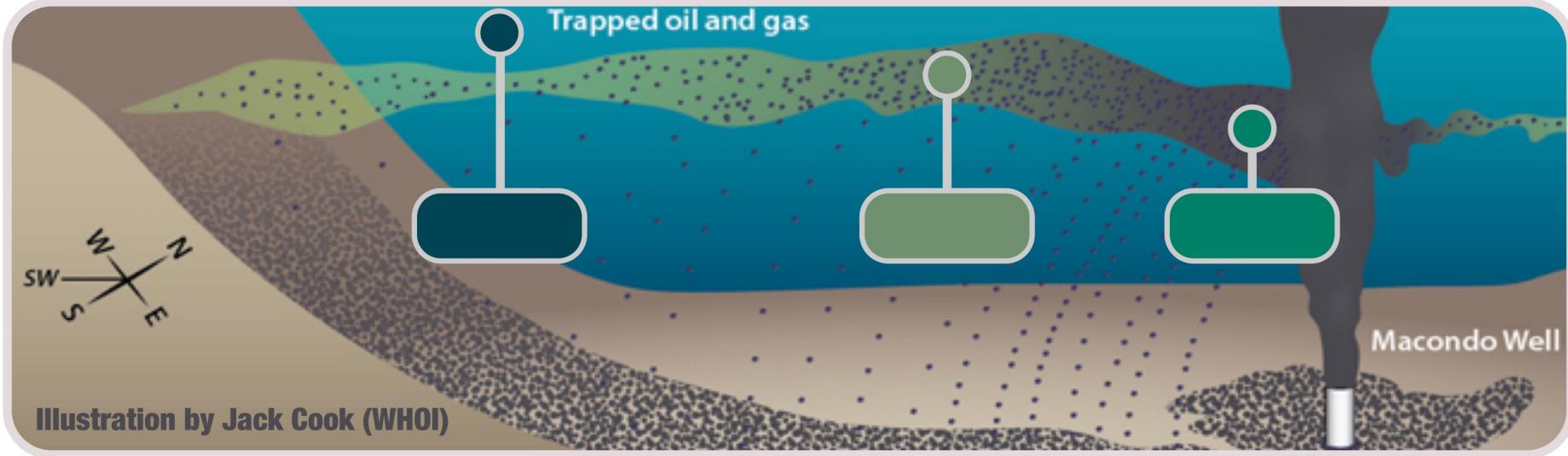


Cultivation-independent genome reconstruction









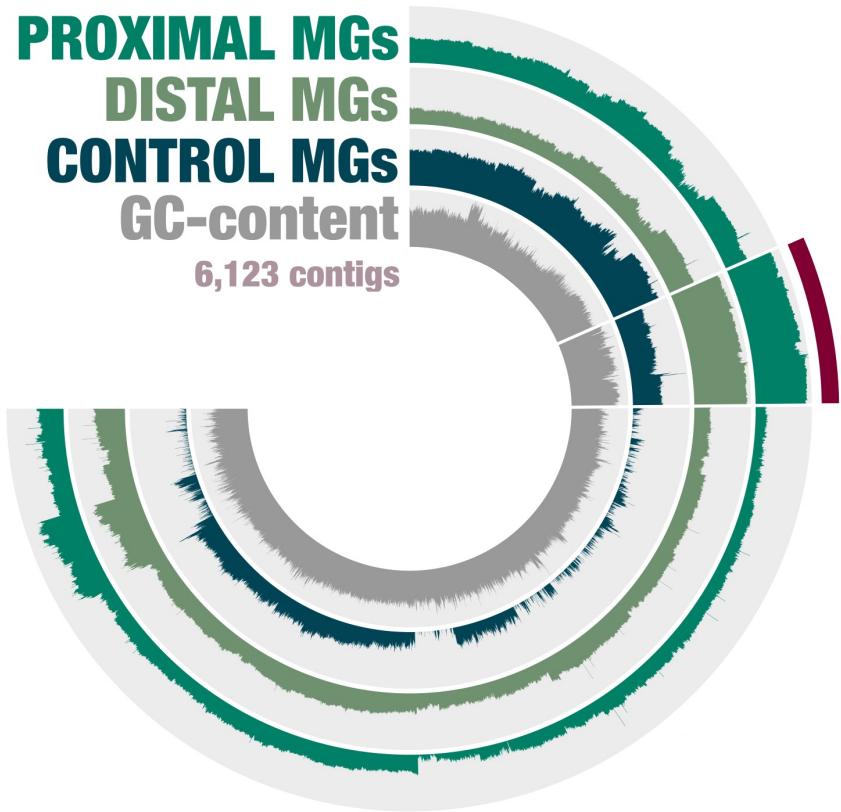
PROXIMAL MGs

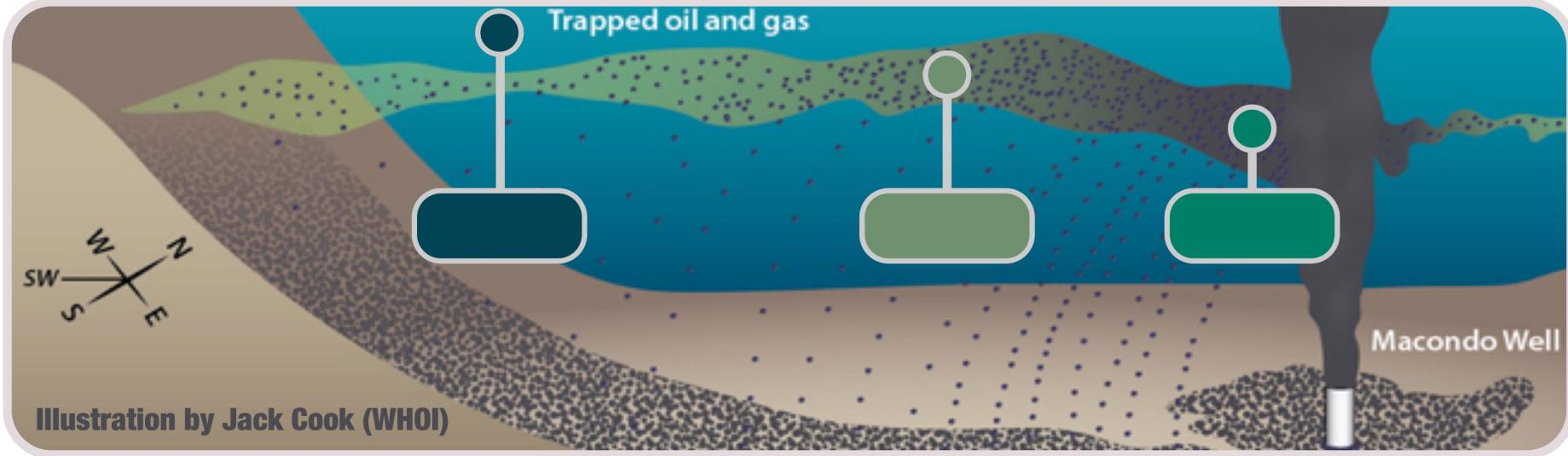
DISTAL MGs

CONTROL MGs

GC-content

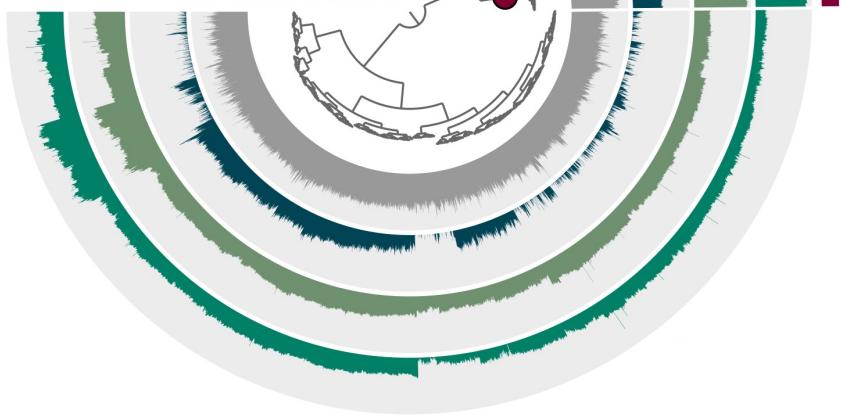
6,123 contigs

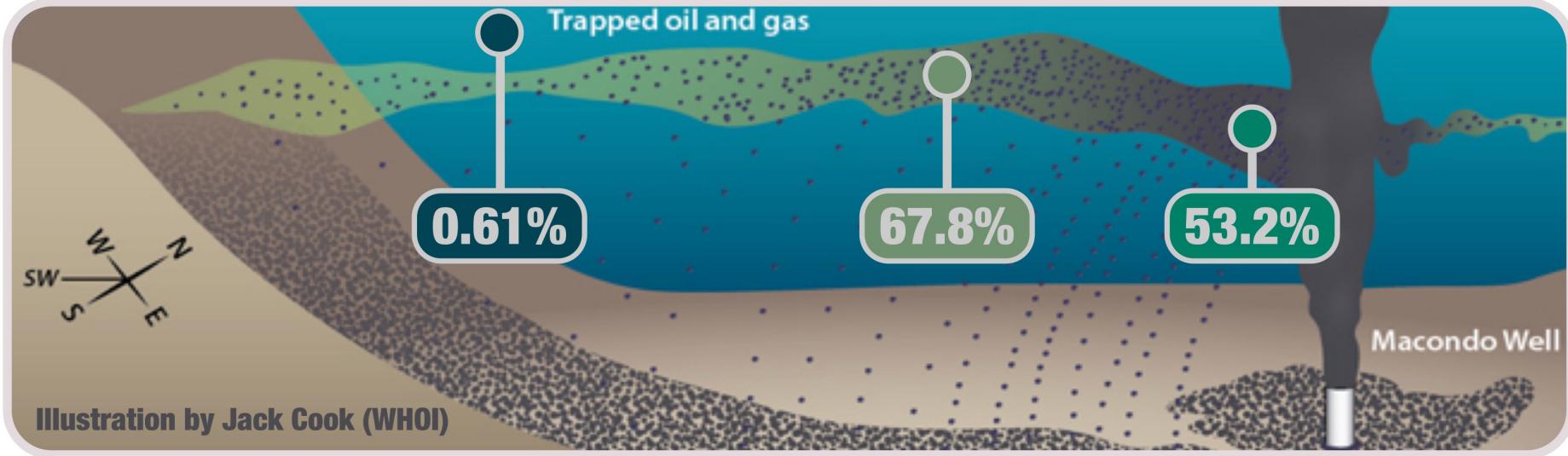




PROXIMAL MGs DISTAL MGs CONTROL MGs GC-content

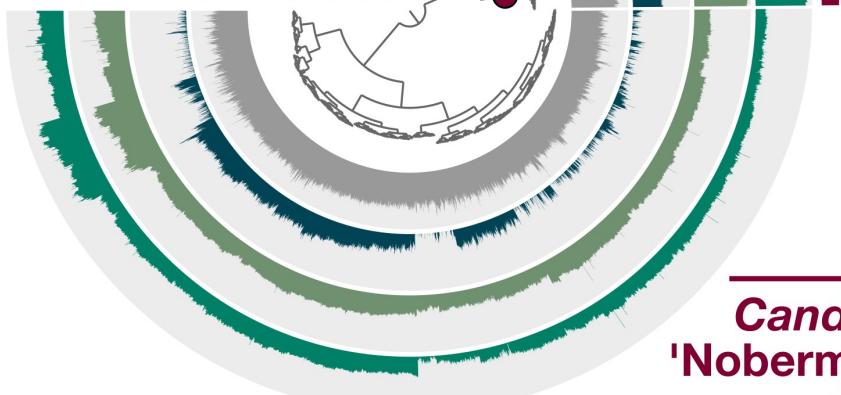
6,123 contigs
Binning métagénomique
Composition de séquence
et distribution





PROXIMAL MGs DISTAL MGs CONTROL MGs GC-content

6,123 contigs
Binning métagénomique
Composition de séquence et distribution



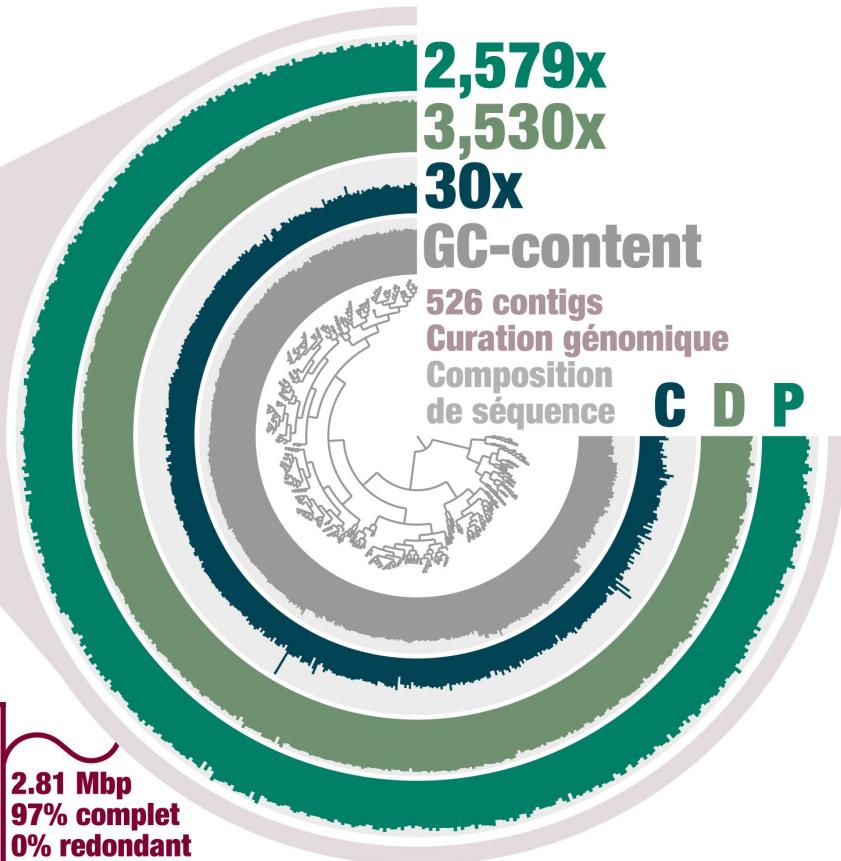
*Candidatus
'Nobermanella
desum'*

2.81 Mbp
97% complet
0% redondant

2,579x
3,530x
30x
GC-content

526 contigs
Curation génomique
Composition de séquence

C D P



Nearly ten years of environmental genomics

Methodological developments

Genome-resolved metagenomics

Manual binning & automatic binning

Fantastic diversity of planktonic genomes were characterized

Still a lot to discover within plankton!

Environmental genomics of plankton in the sunlit oceans

2015

2016

2017

2018

2019

2020

2021

2022

2023

2024

2025

Tara Oceans metagenomes are made public



Nearly ten years of environmental genomics

Methodological developments



Platform dedicated
to genome-resolved
metagenomics

Environmental genomics of plankton in the sunlit oceans

2015

Tara Oceans
metagenomes
are made public

2016

2017

2018

2019

2020

2021

2022

2023

2024

2025



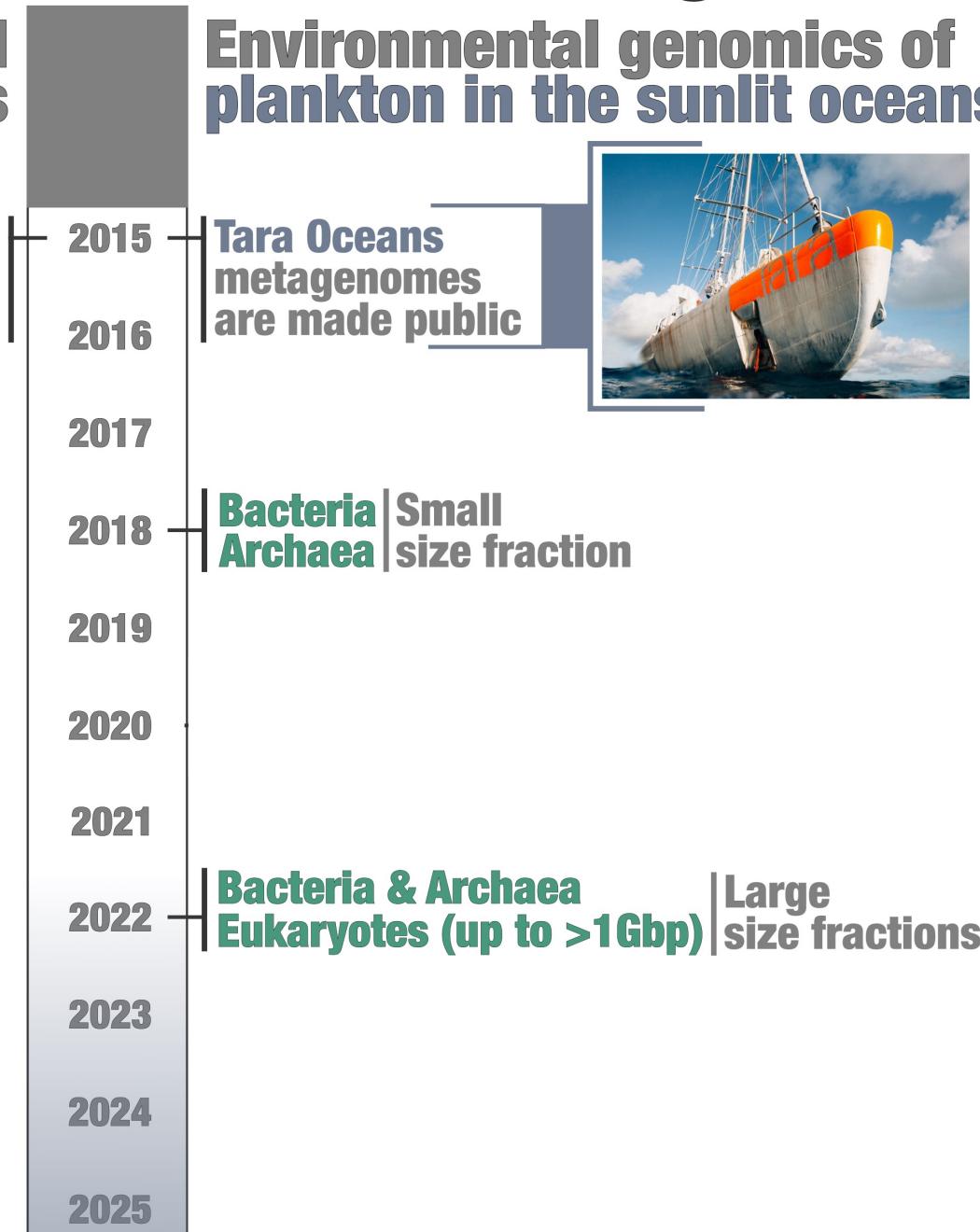
Nearly ten years of environmental genomics

Methodological developments



Platform dedicated
to genome-resolved
metagenomics

Environmental genomics of plankton in the sunlit oceans



Trichodesmium

Paradigm: N-fixing genus

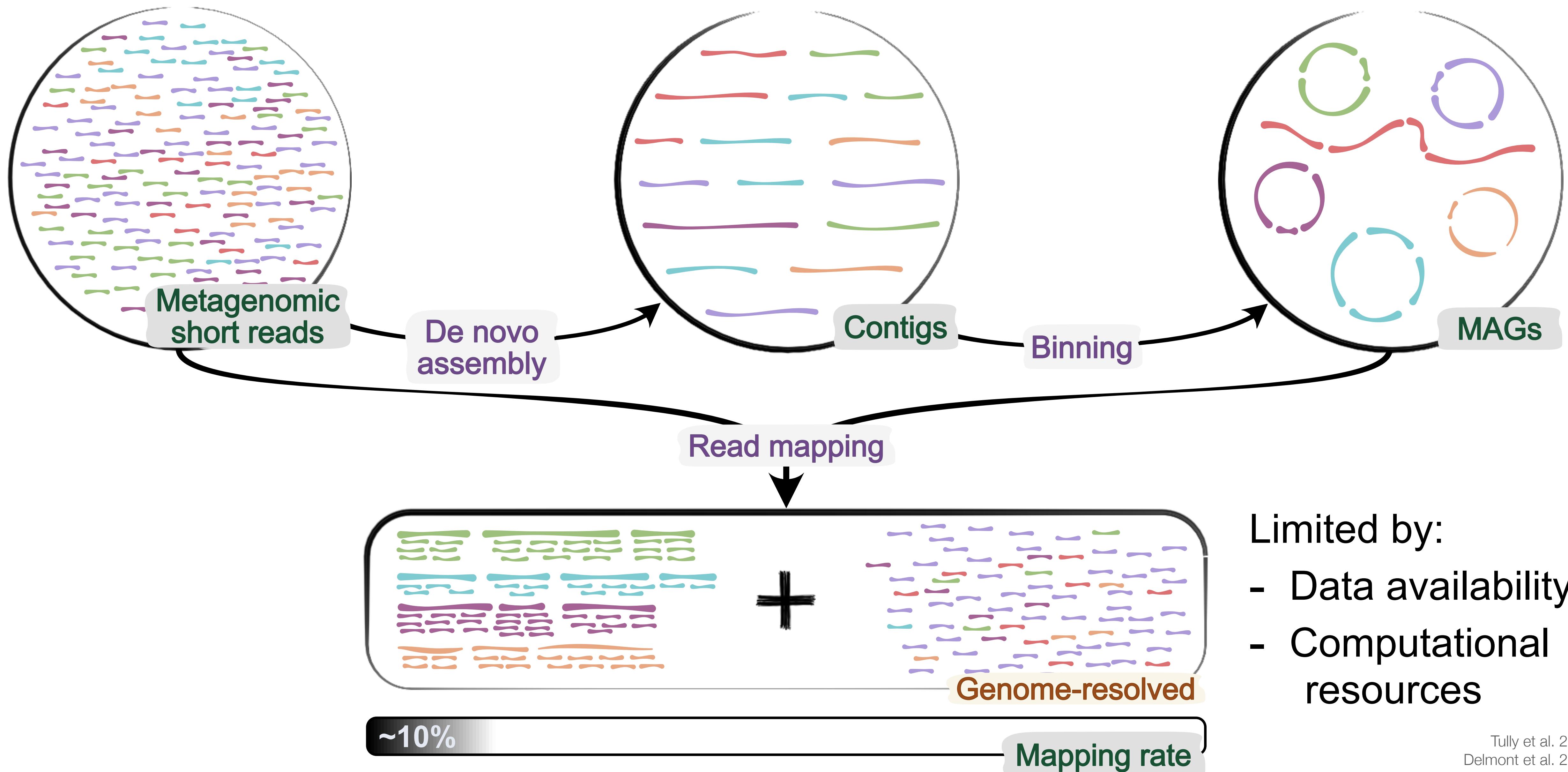
Substantial N-fixing rates (large blooms)

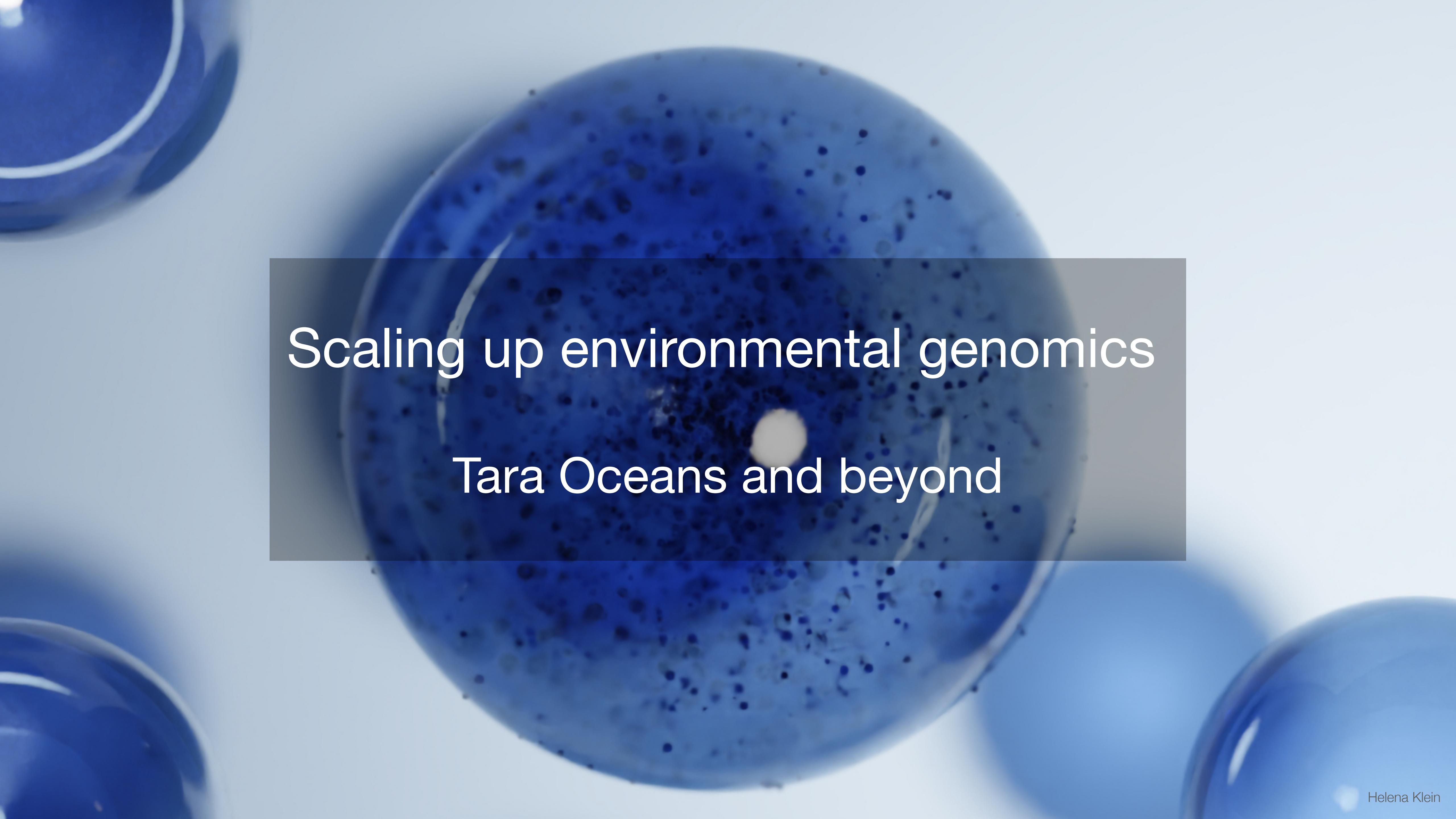


Nurture their own microbiome (colonies)

A well-known lineage (discovery: 1830)

Cultivation-independent genome reconstruction

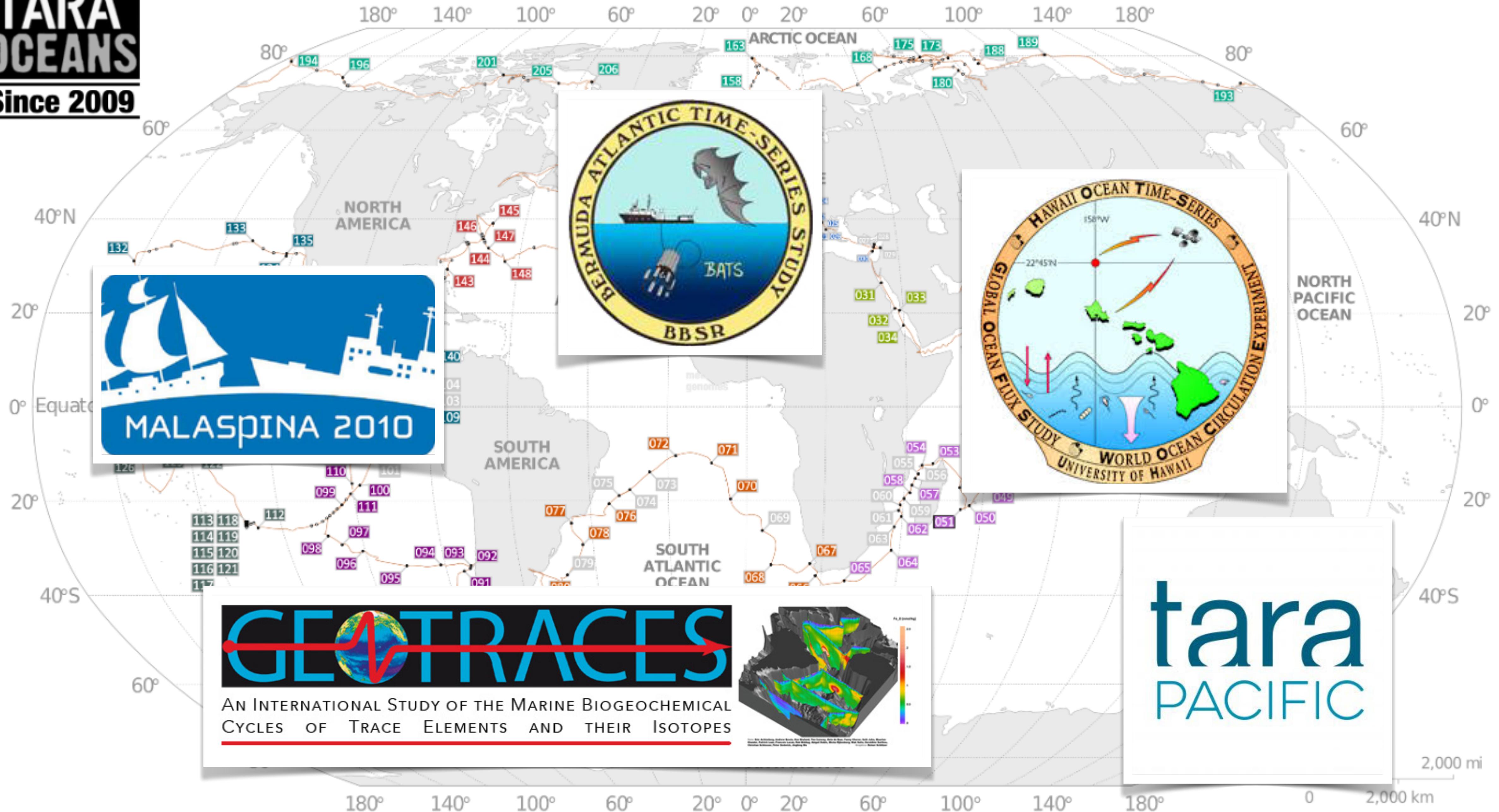




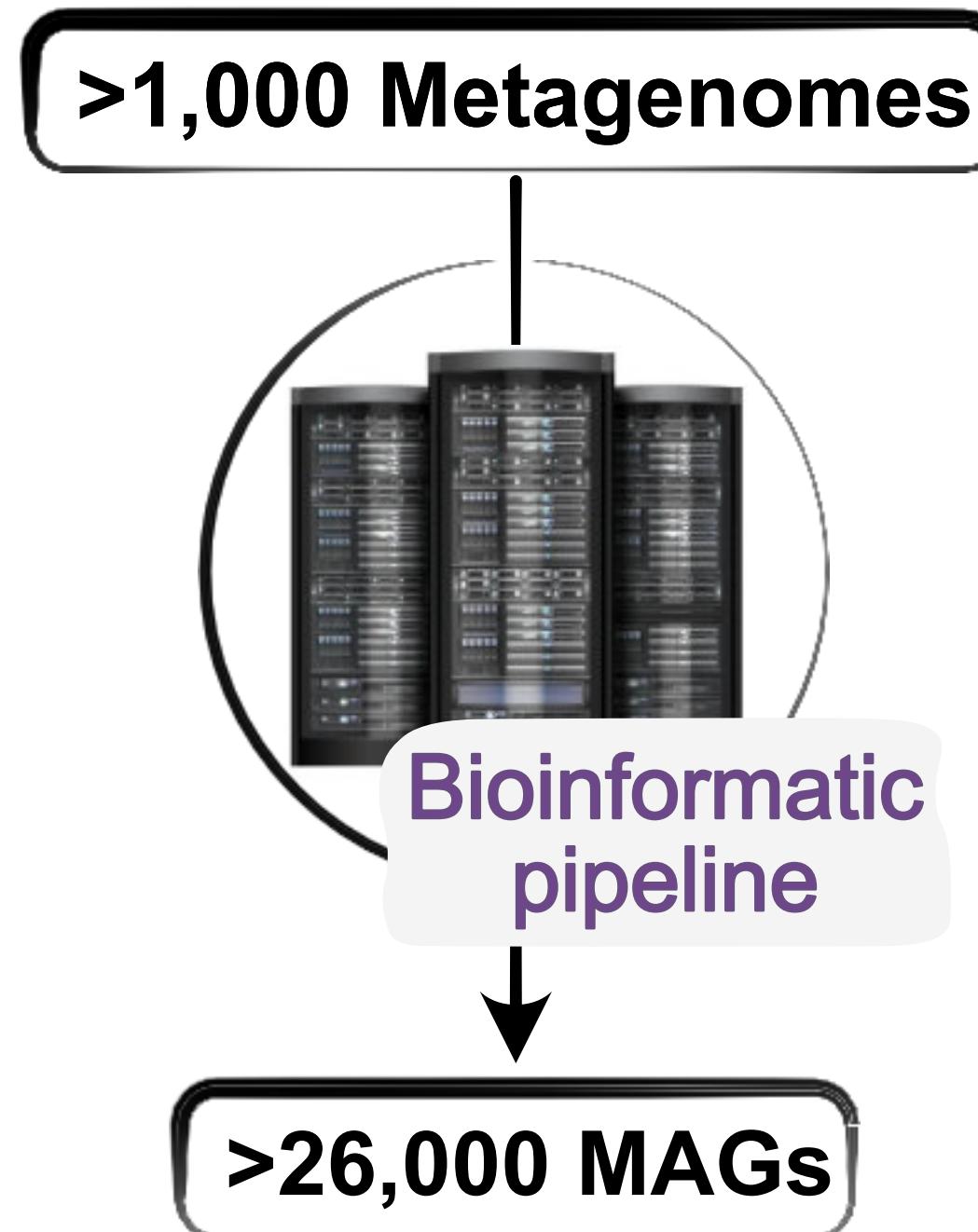
Scaling up environmental genomics

Tara Oceans and beyond

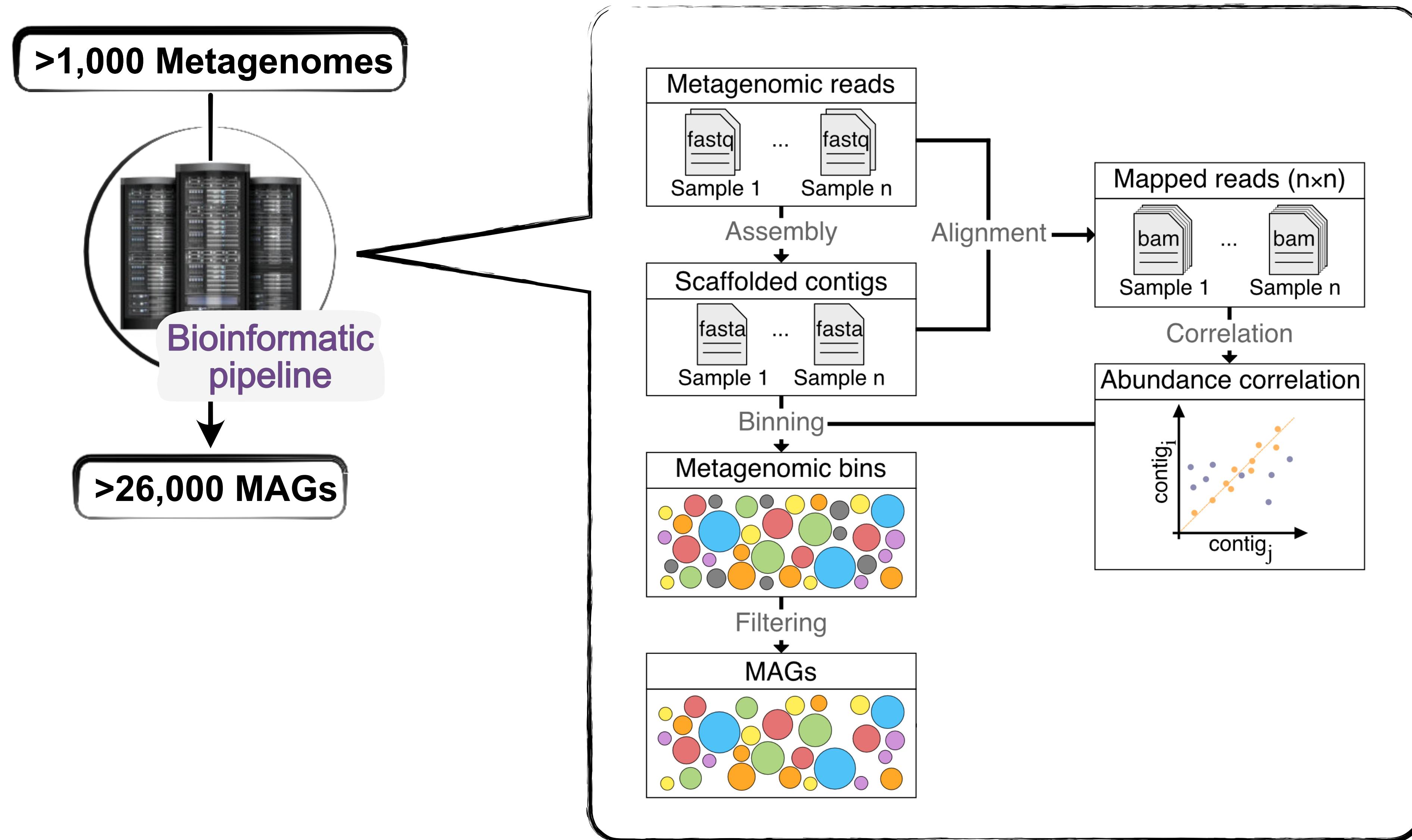
**TARA
OCEANS**
Since 2009



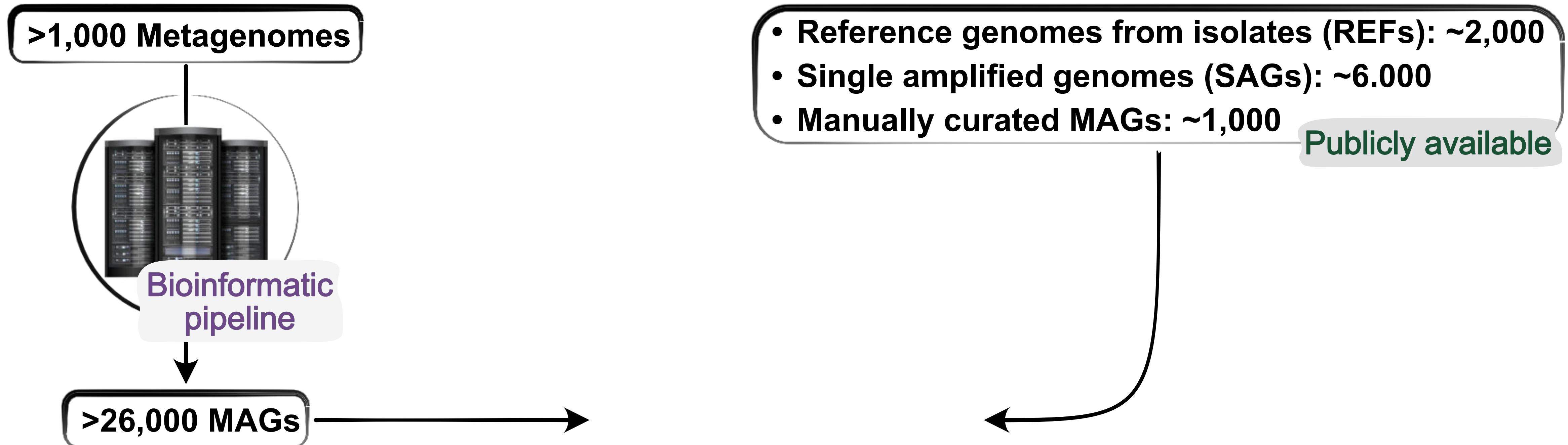
Establishing the Ocean Microbiomics Database



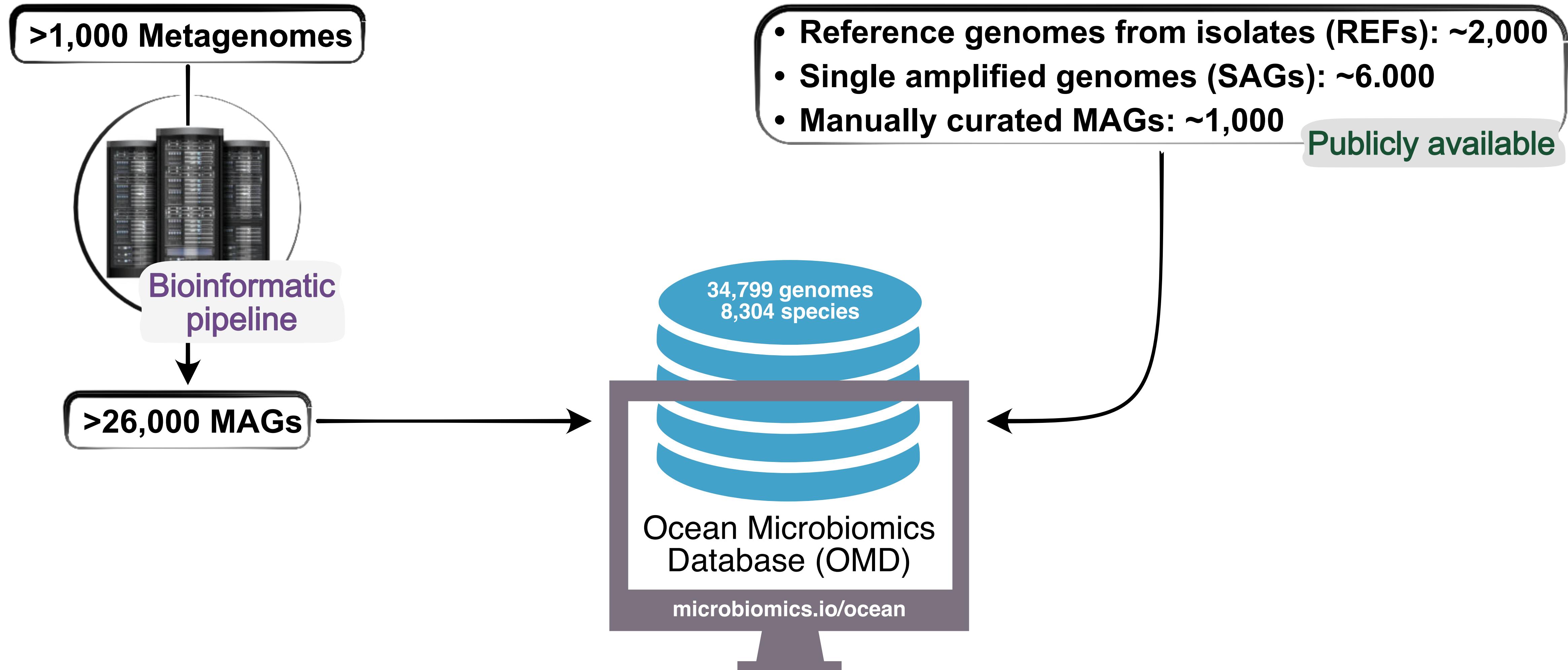
Establishing the Ocean Microbiomics Database



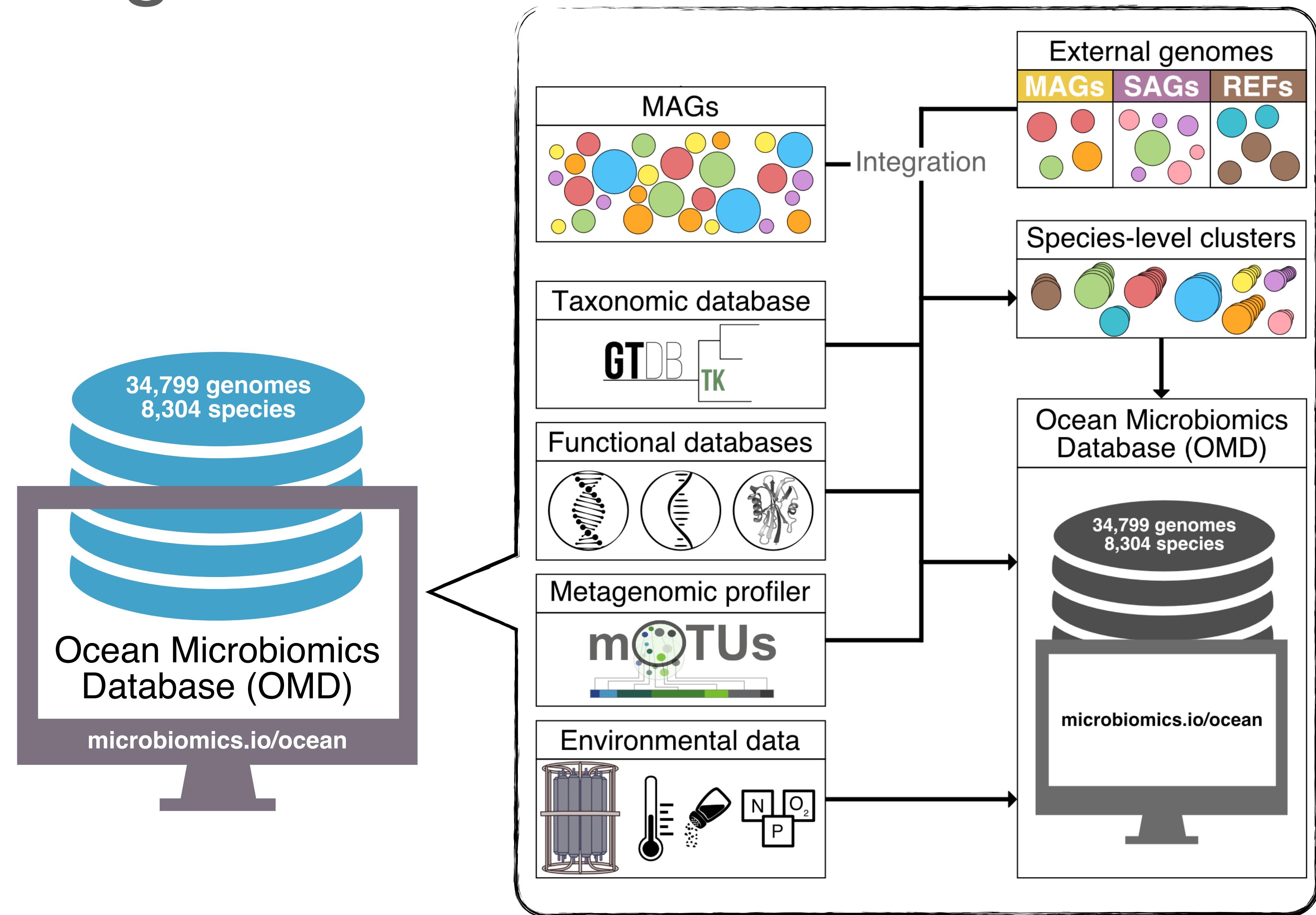
Establishing the Ocean Microbiomics Database



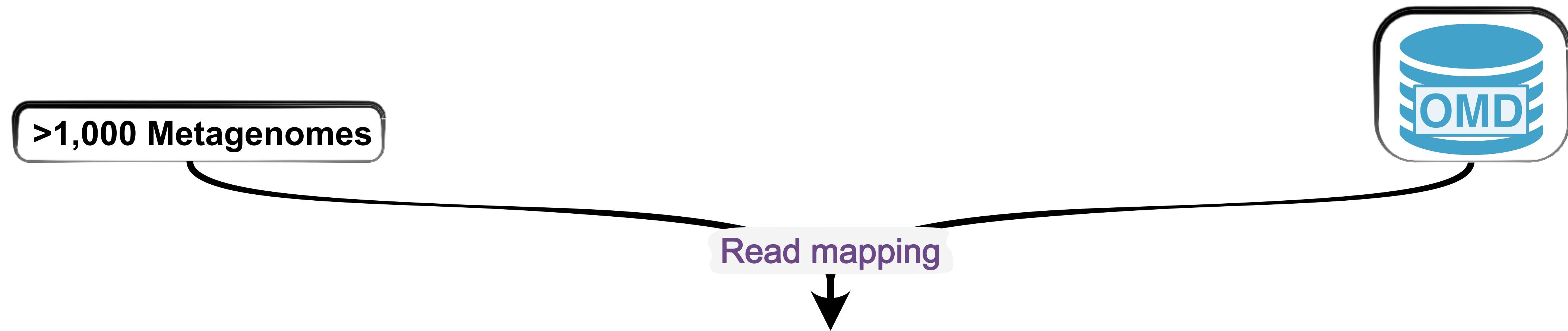
Establishing the Ocean Microbiomics Database



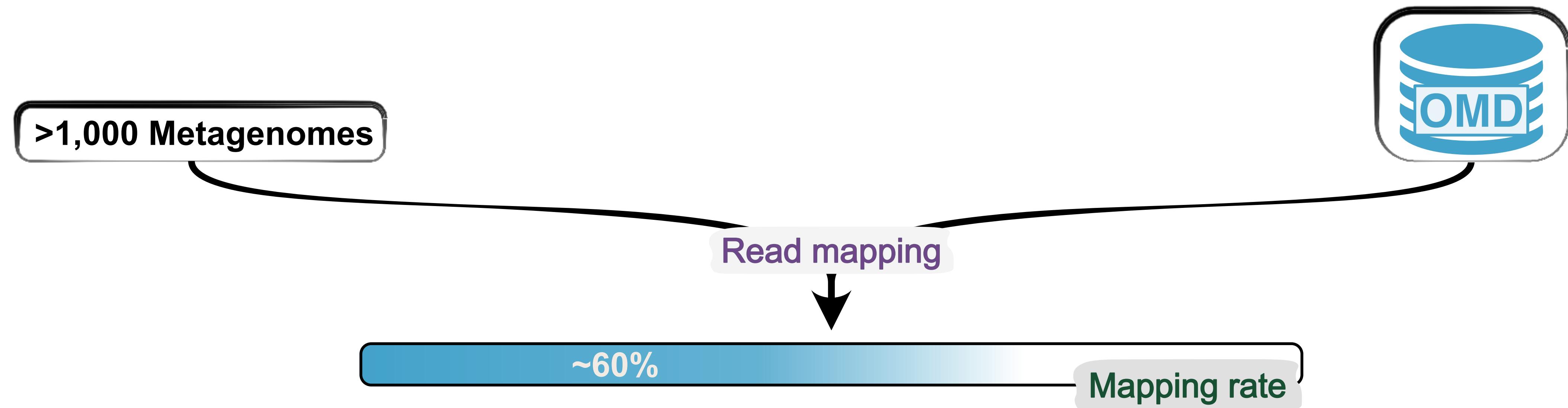
Establishing the Ocean Microbiomics Database



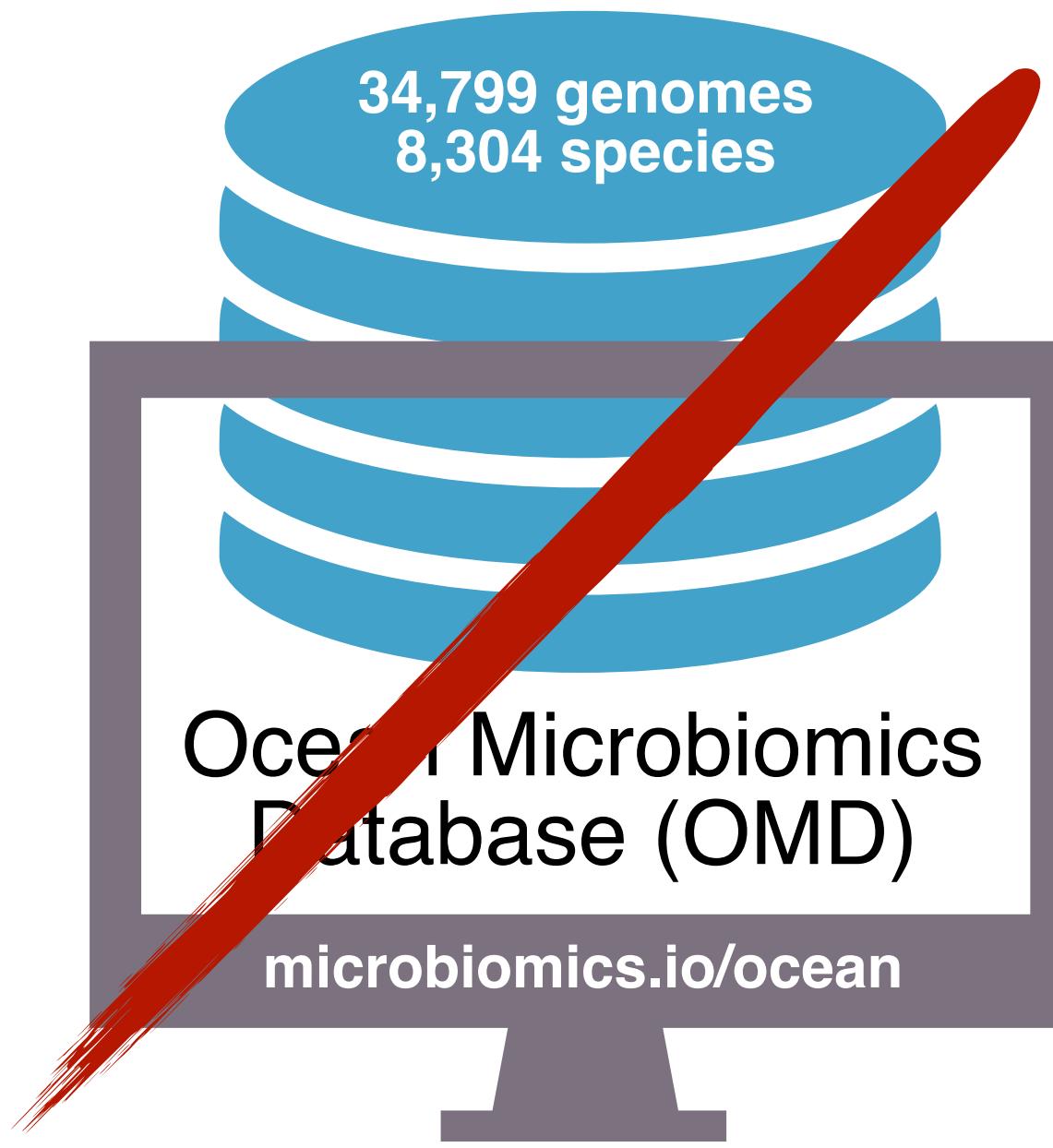
The OMD captures most of the ocean microbiome



The OMD captures most of the ocean microbiome



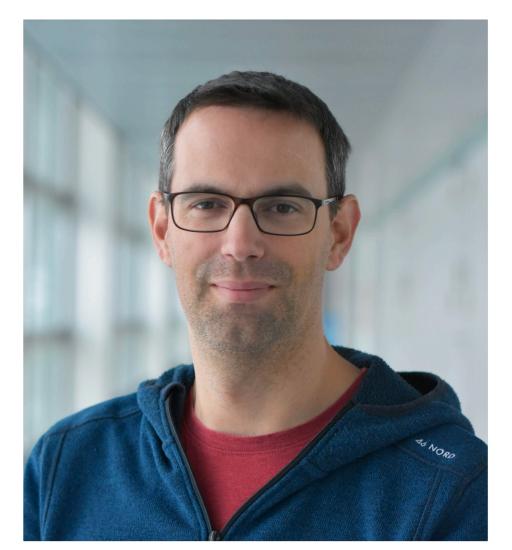
OMD early 2024 update

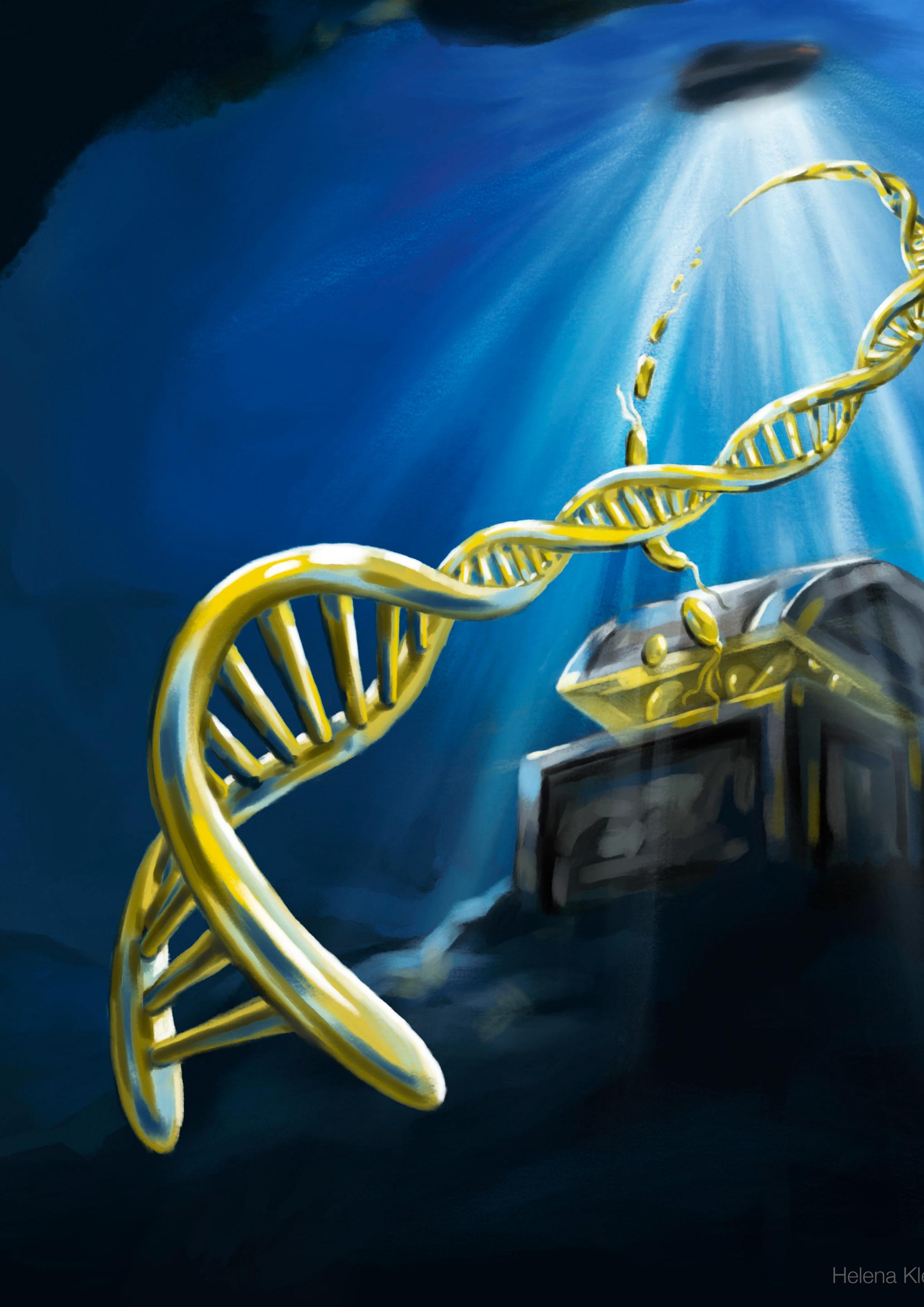


- **200 Studies**
- **12k samples**
- **270k MAGs**

→ **microbiomics.io/ocean2**

Please do let us know if you have
feedback and/or feature requests!





Helena Klein

Article

Biosynthetic potential of the global ocean microbiome

<https://doi.org/10.1038/s41586-022-04862-3>

Received: 21 May 2021

Accepted: 12 May 2022

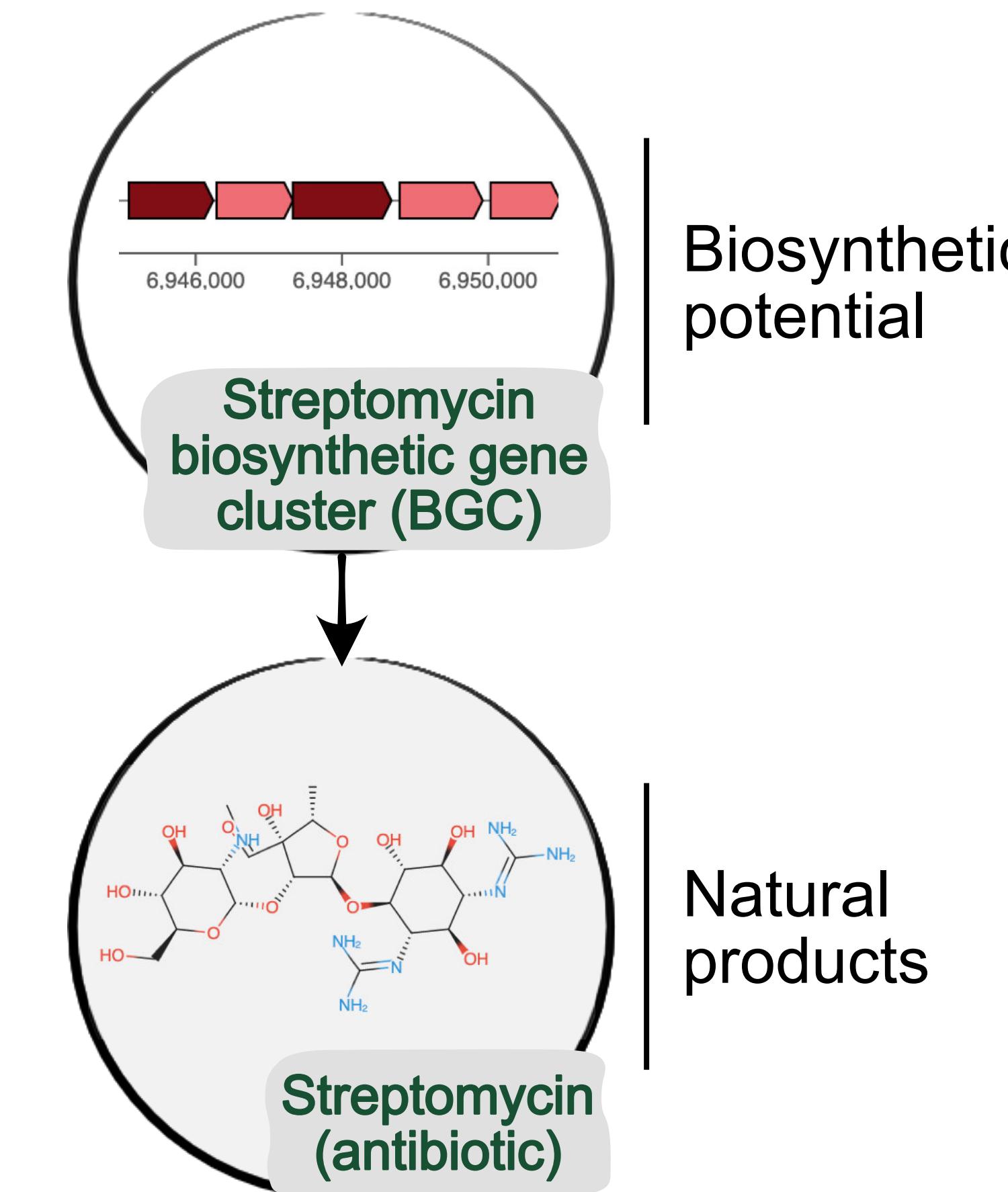
Published online: 22 June 2022

Open access

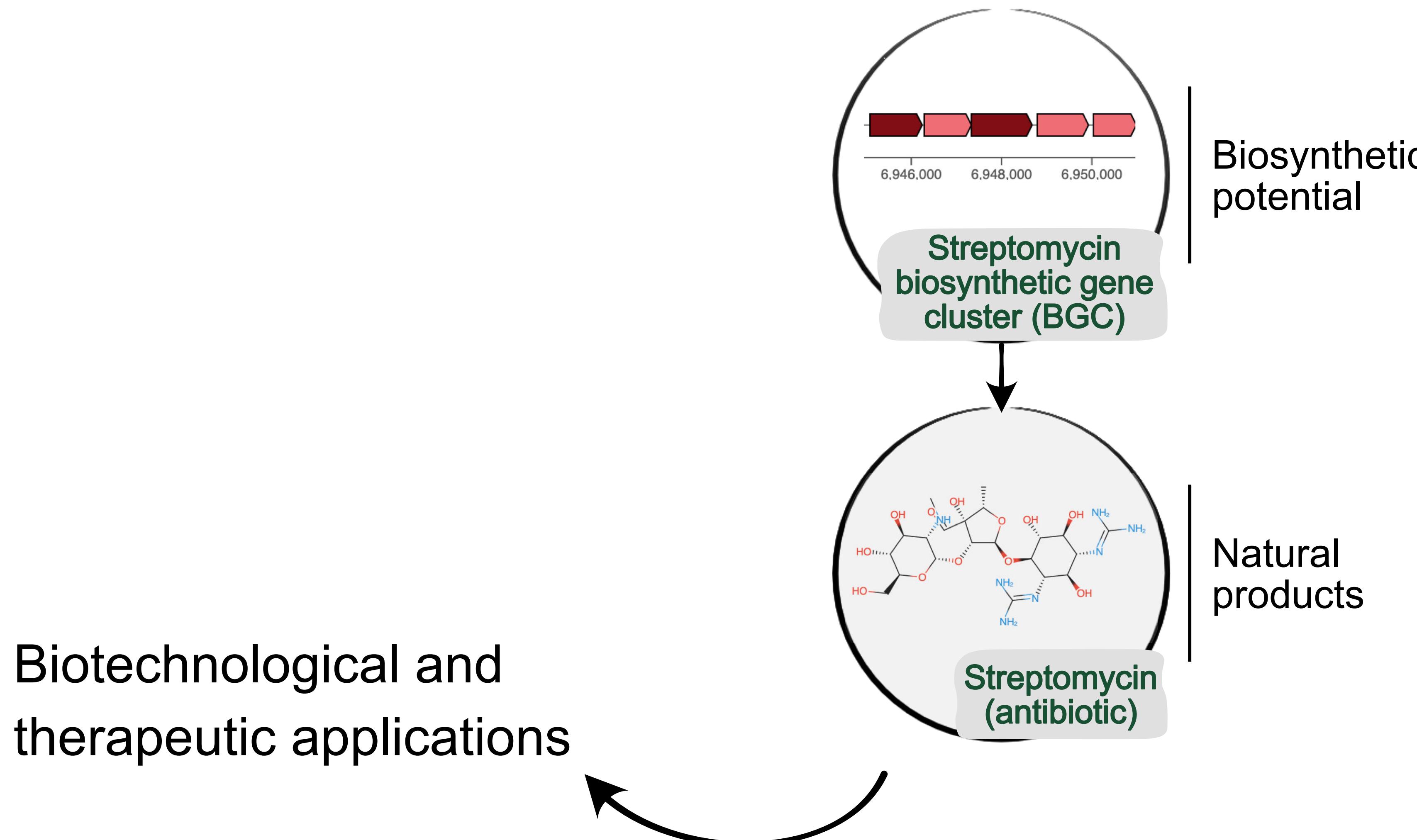
 Check for updates

Lucas Paoli¹, Hans-Joachim Ruscheweyh^{1,18}, Clarissa C. Forneris^{2,18}, Florian Hubrich^{2,18}, Satria Kautsar³, Agneya Bhushan², Alessandro Lotti², Quentin Clayssen¹, Guillem Salazar¹, Alessio Milanese¹, Charlotte I. Carlström¹, Chrysa Papadopoulou¹, Daniel Gehrig¹, Mikhail Karasikov^{4,5,6}, Harun Mustafa^{4,5,6}, Martin Larralde⁷, Laura M. Carroll⁷, Pablo Sánchez⁸, Ahmed A. Zayed⁹, Dylan R. Cronin⁹, Silvia G. Acinas⁸, Peer Bork^{7,10,11}, Chris Bowler^{12,13}, Tom O. Delmont^{13,14}, Josep M. Gasol⁸, Alvar D. Gossert¹⁵, André Kahles^{4,5,6}, Matthew B. Sullivan^{8,16}, Patrick Wincker^{13,14}, Georg Zeller⁷, Serina L. Robinson^{2,17}, Jörn Piel² & Shinichi Sunagawa¹

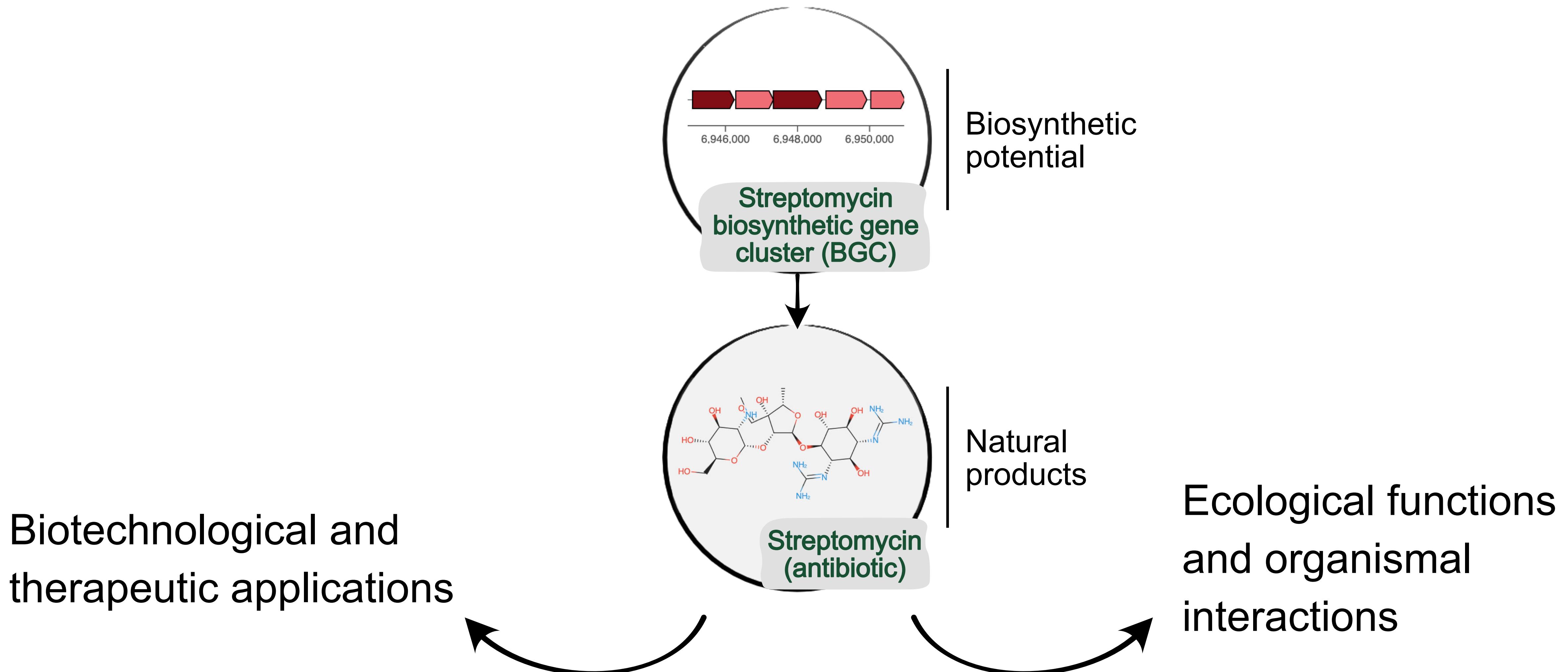
Biosynthetic potential



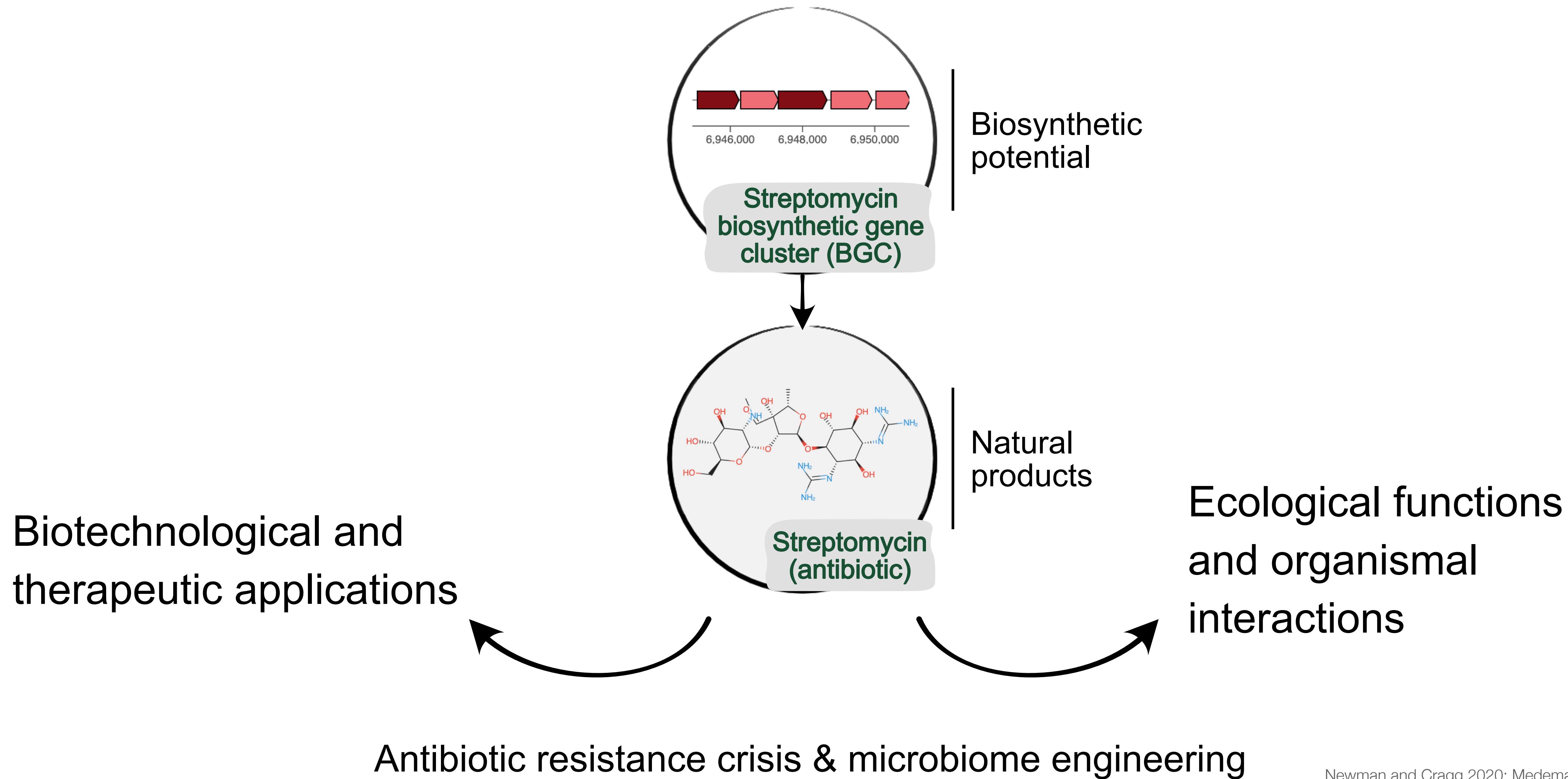
Biosynthetic potential: fuelling applications



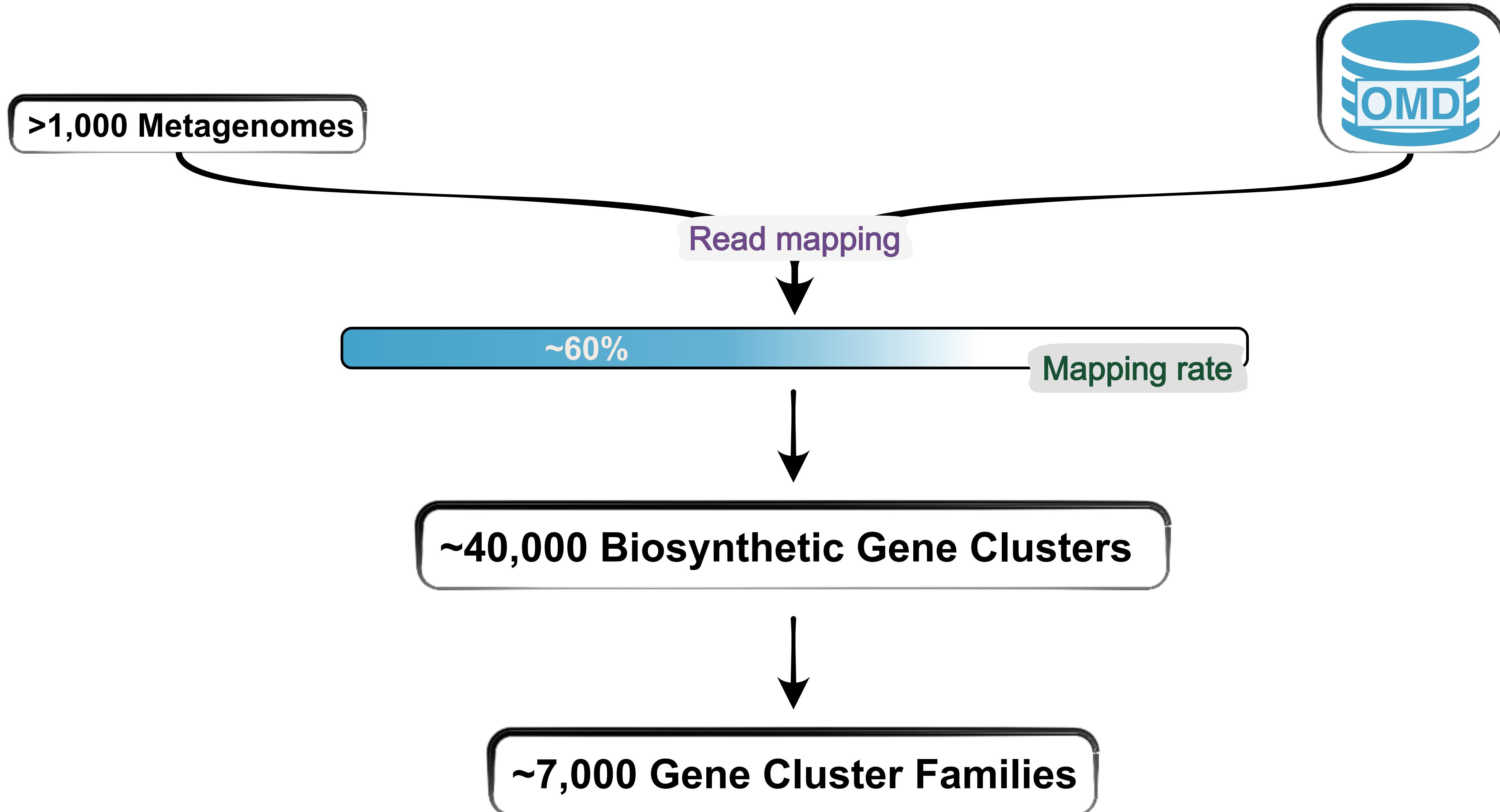
Biosynthetic potential: shaping interactions

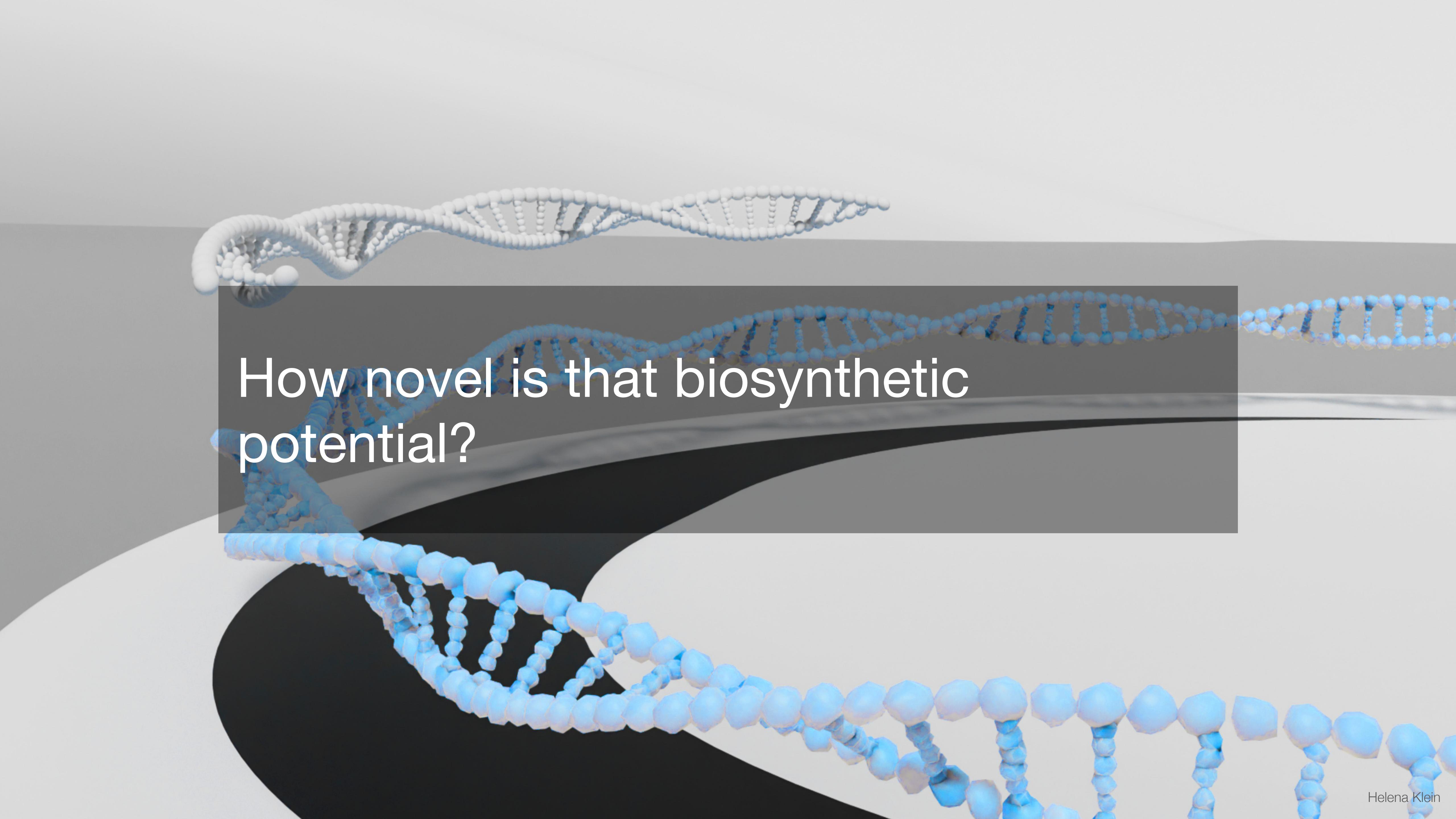


Biosynthetic potential: fundamental & applied



Mining the OMD for its biosynthetic potential





How novel is that biosynthetic potential?

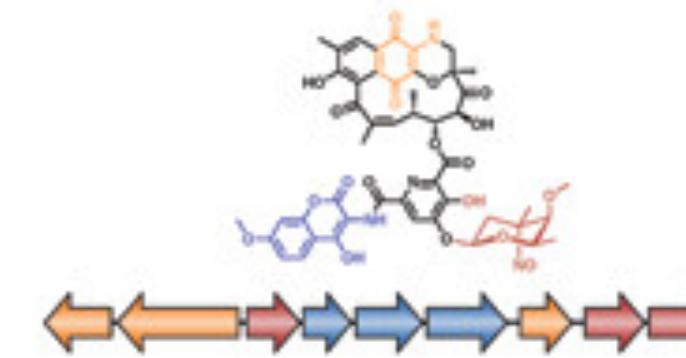
Comparison to sequenced or characterised BGCs



RefSeq
~200,000 genomes



1.2 M BGCs



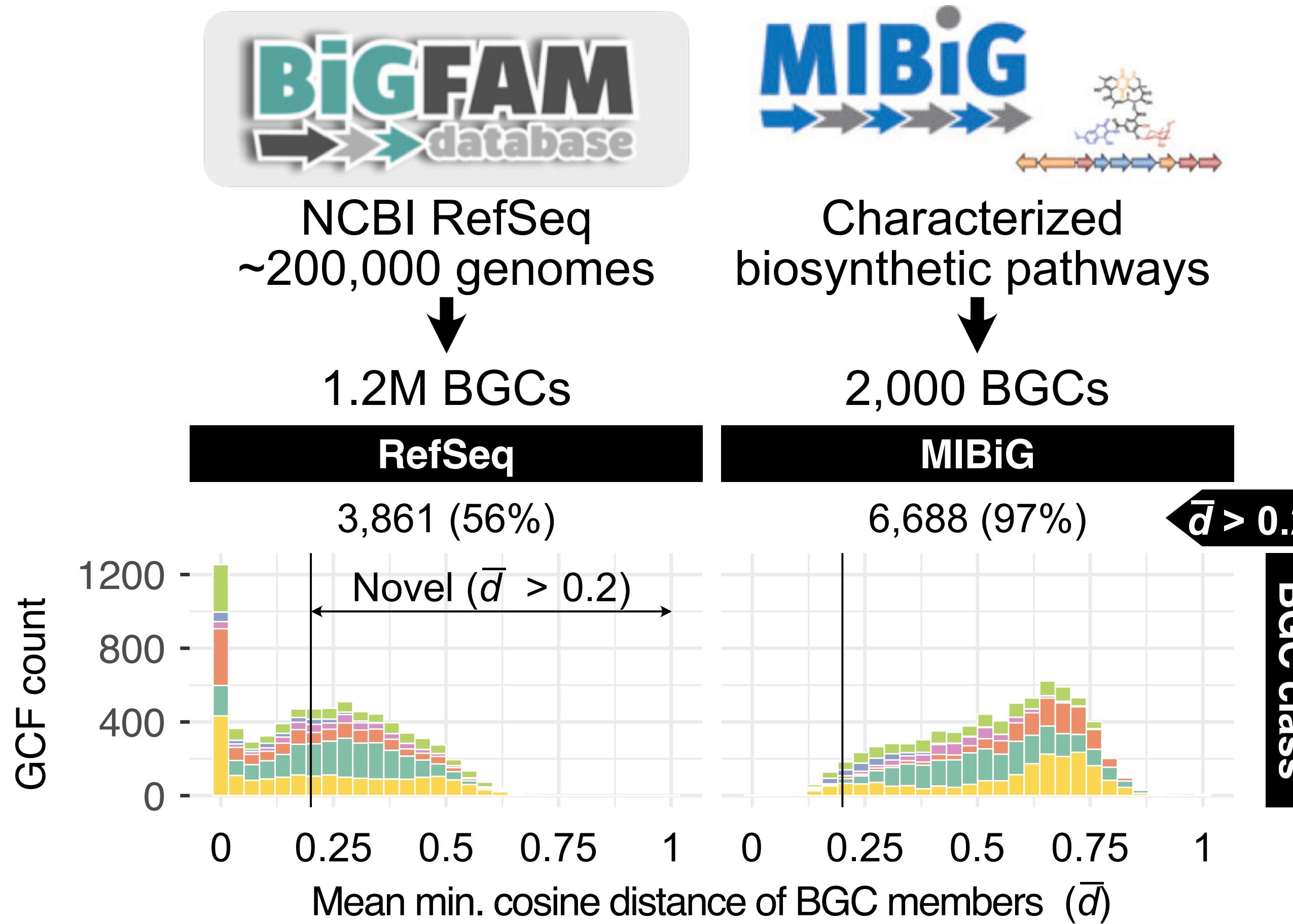
Characterised
biosynthetic pathways

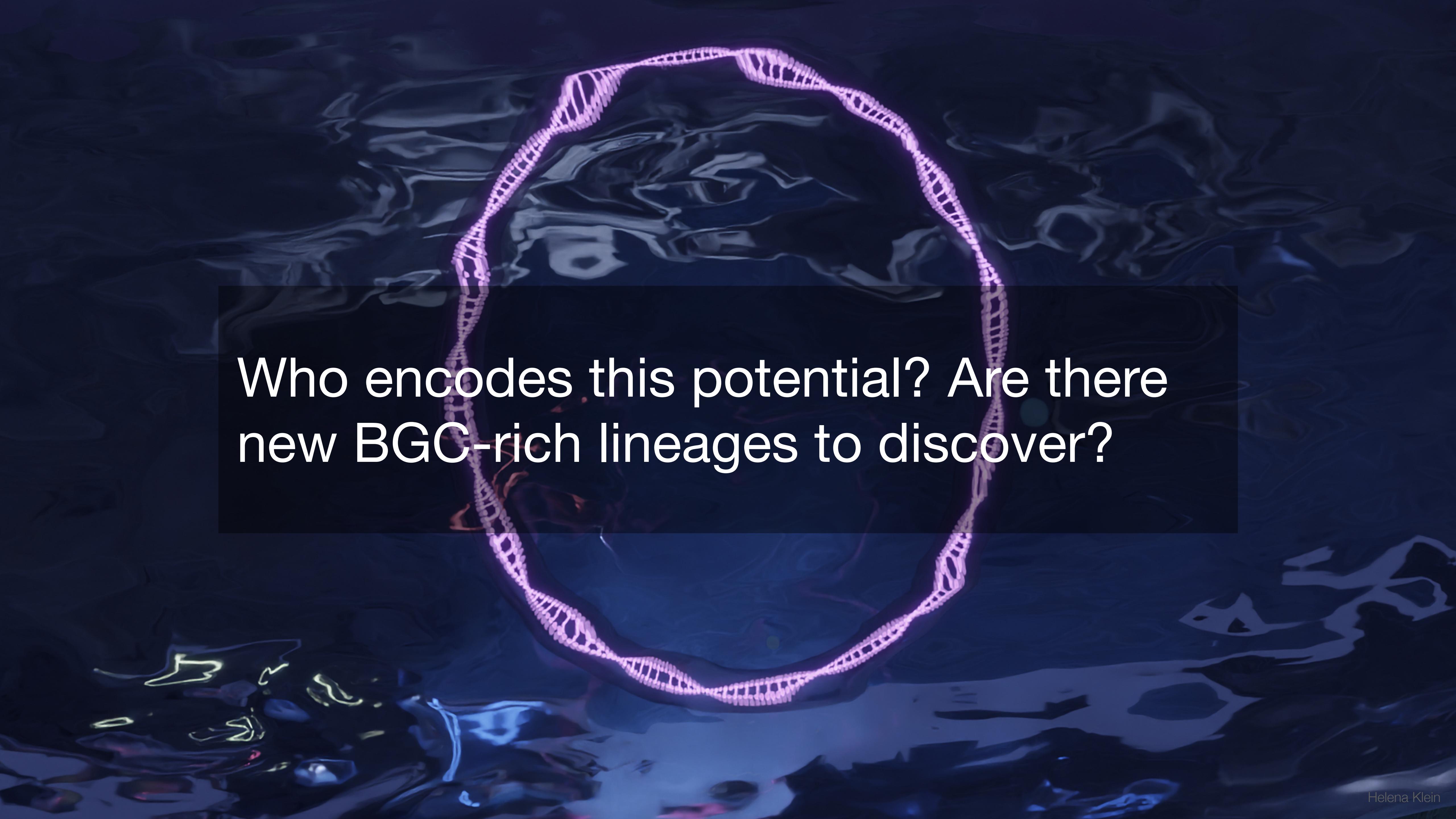


2,000 BGCs



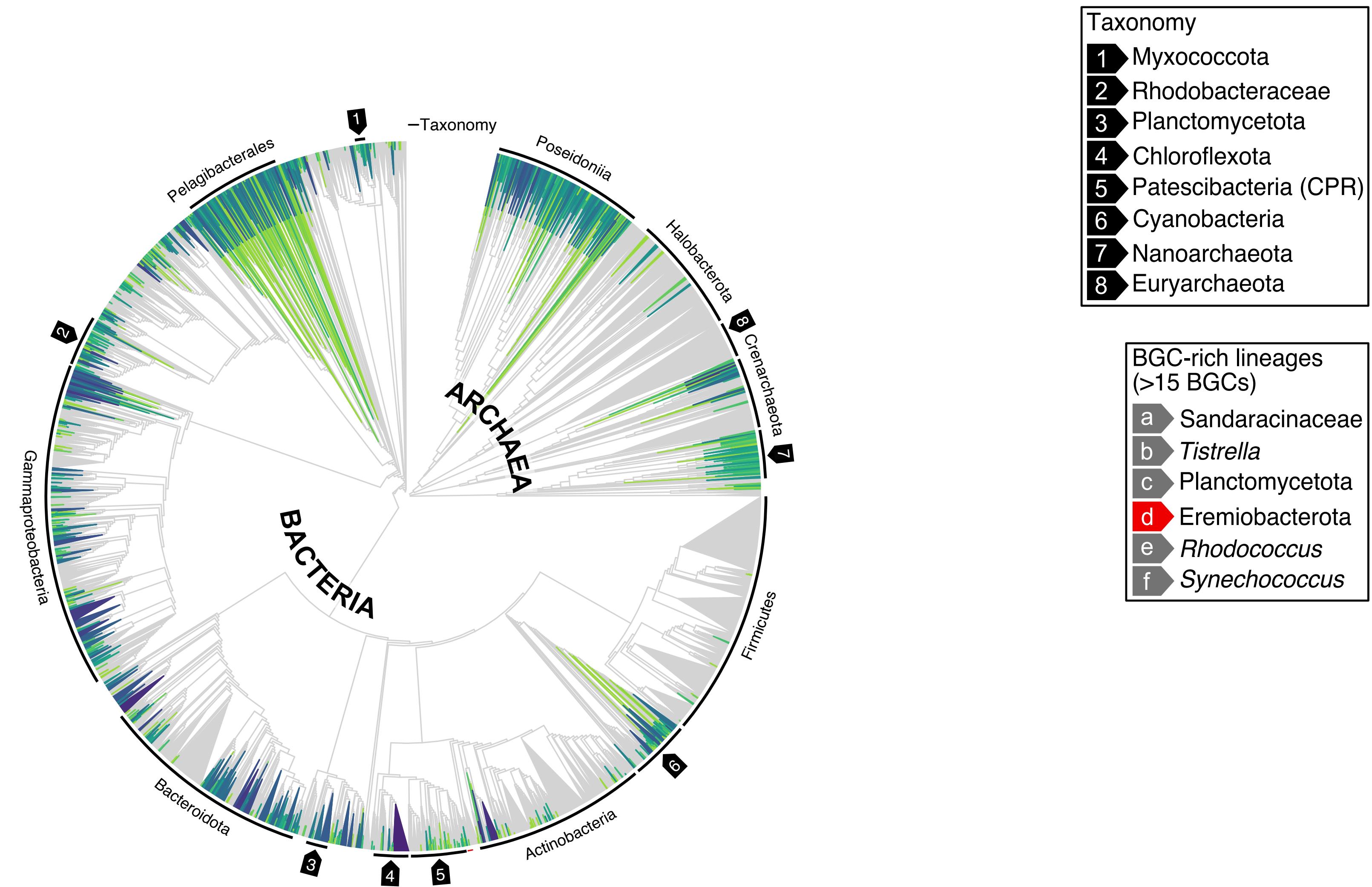
Reveals large potential for new compounds



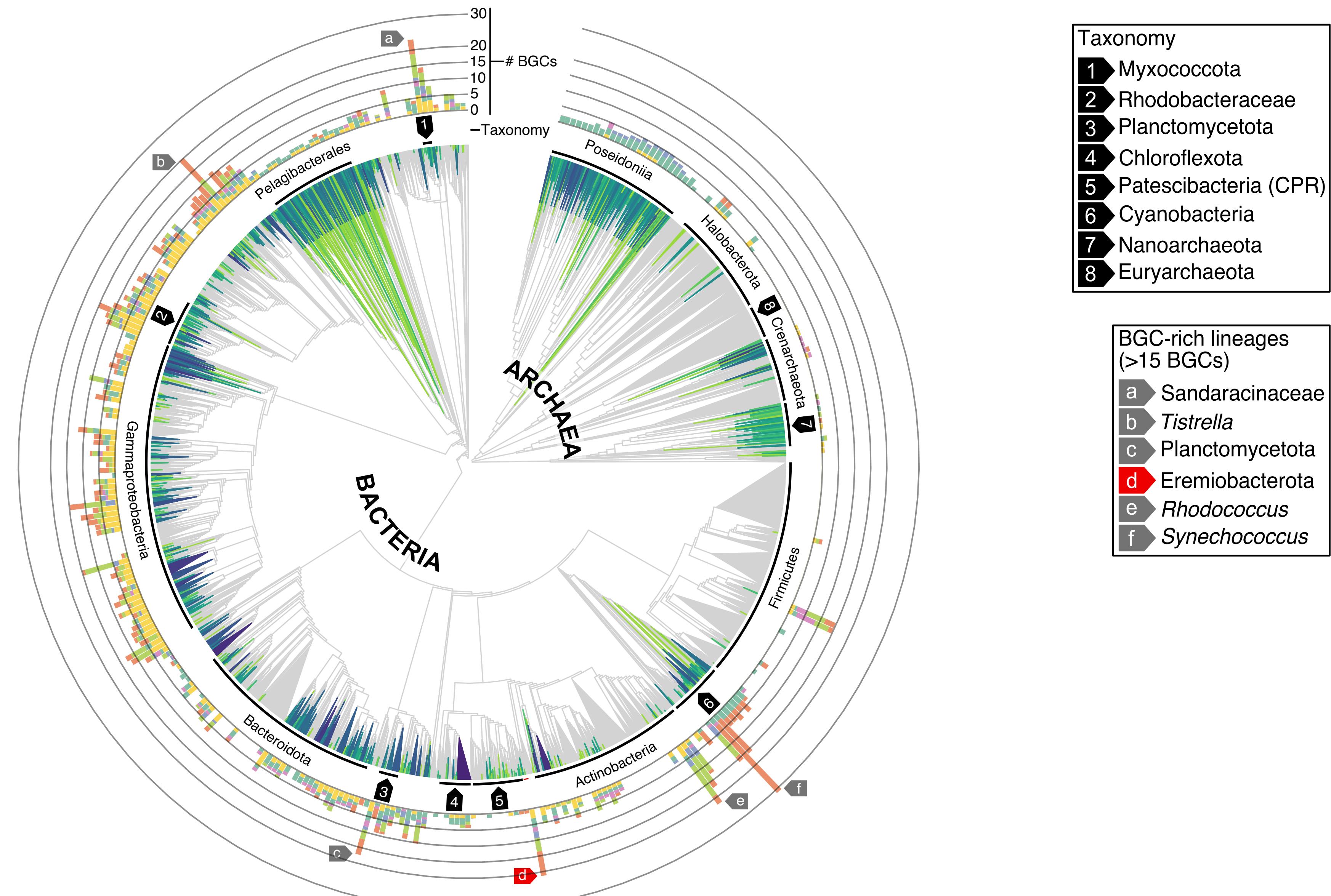
The background of the slide features a dark, abstract design. A glowing, light-blue DNA double helix is positioned in the center, oriented diagonally from the top-left towards the bottom-right. The background is filled with dark, swirling, liquid-like patterns in shades of blue and purple, with some bright, glowing spots and small, translucent, glowing droplets scattered throughout.

Who encodes this potential? Are there
new BGC-rich lineages to discover?

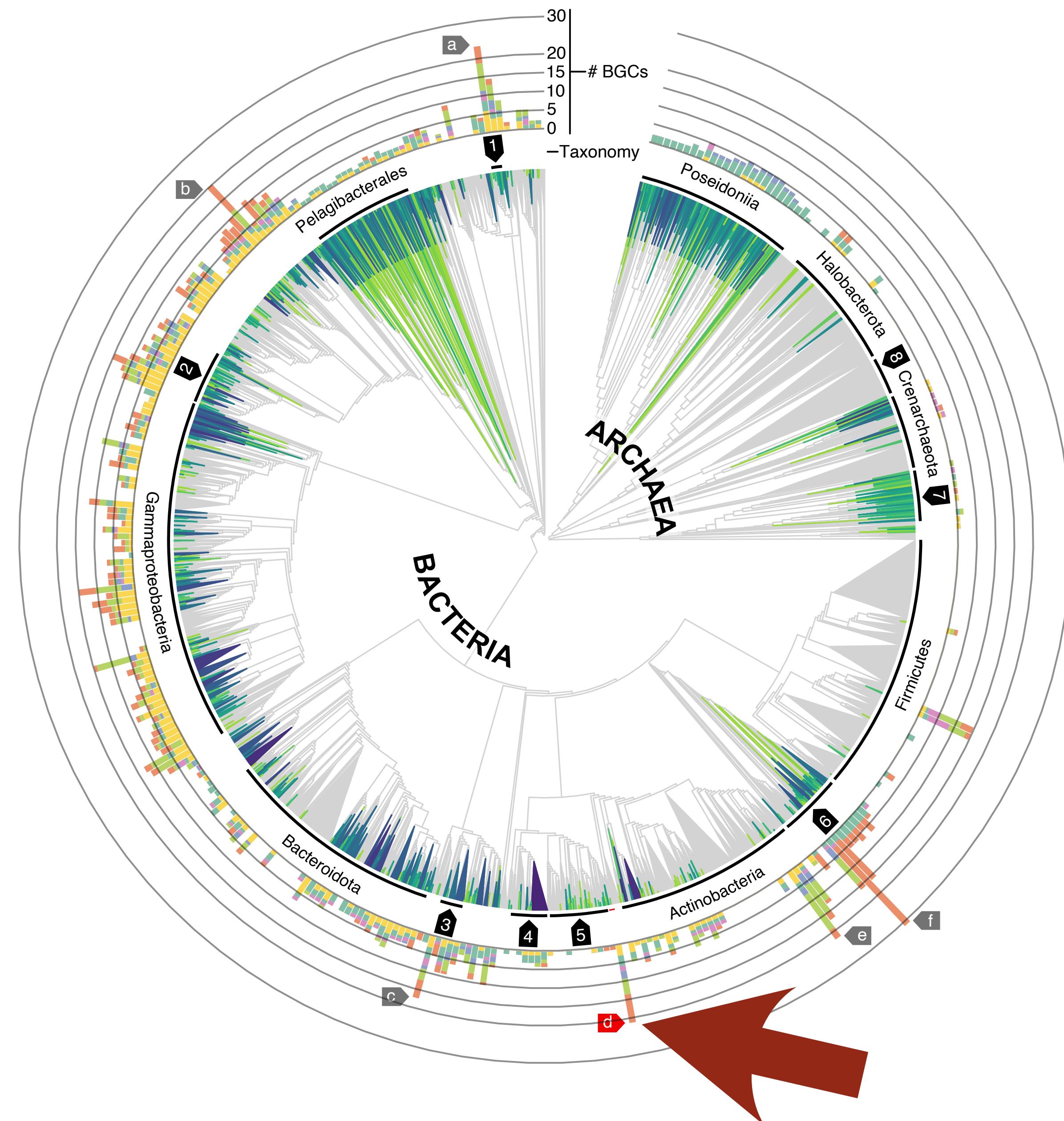
Distribution of the ocean's biosynthetic potential



Distribution of the ocean's biosynthetic potential



Eremiobacterota: unexpected BGC-richness



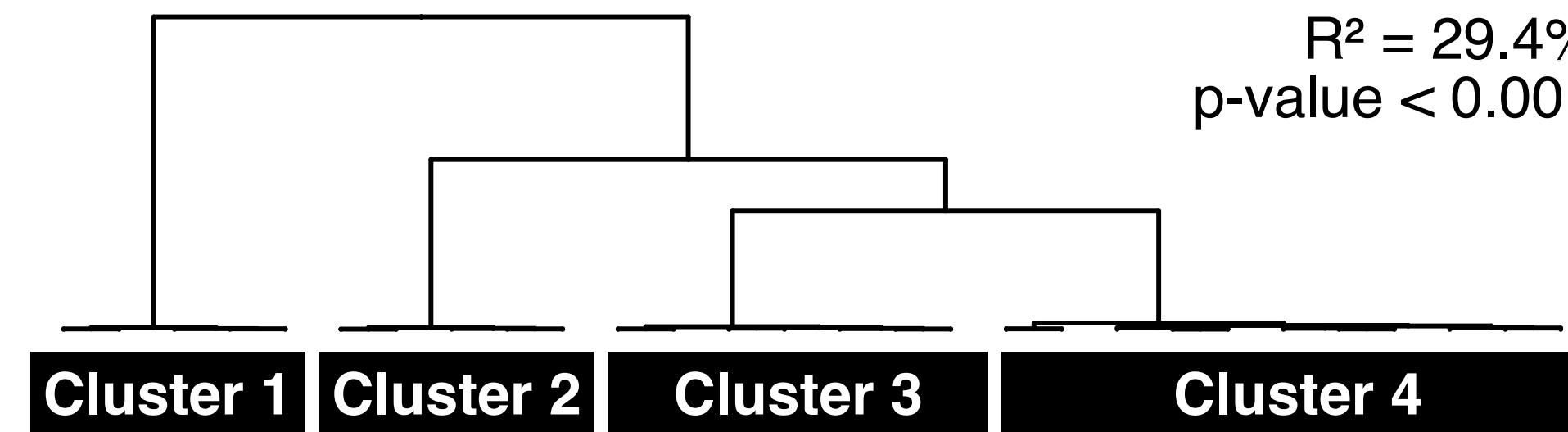
Taxonomy
1 Myxococcota
2 Rhodobacteraceae
3 Planctomycetota
4 Chloroflexota
5 Patescibacteria (CPR)
6 Cyanobacteria
7 Nanoarchaeota
8 Euryarchaeota

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



Ecology of a new biosynthetically 'talented' lineage?

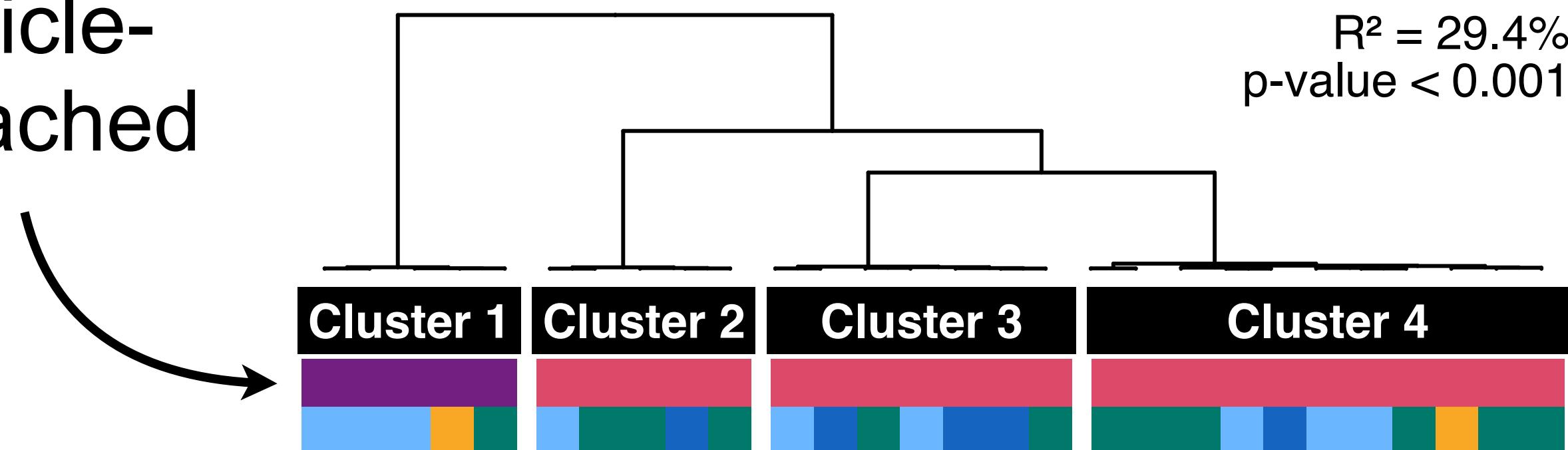
Ecology with *in situ* metatranscriptomics



Clustering based on the expression levels of all genes

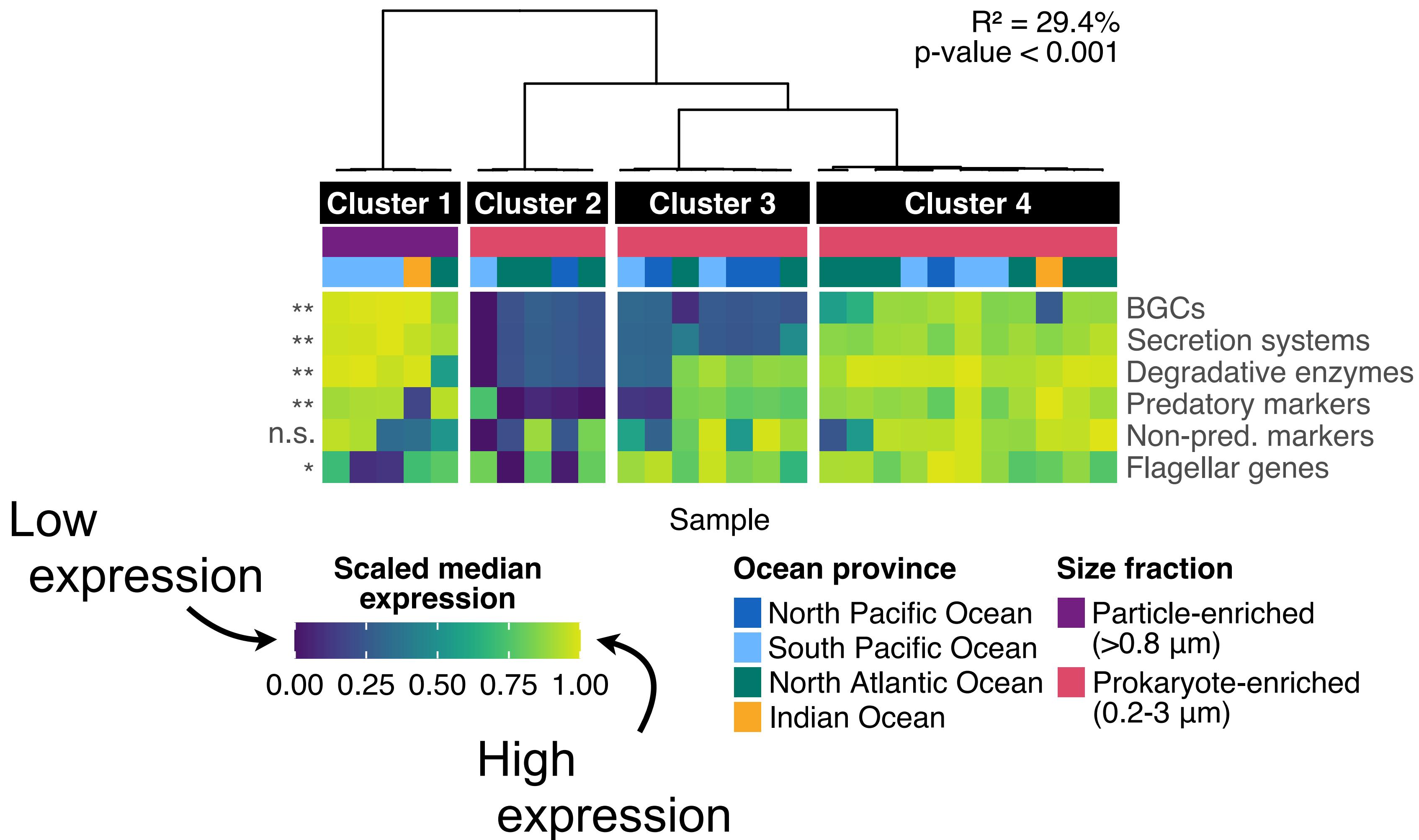
Ecology with *in situ* metatranscriptomics

Particle-
attached

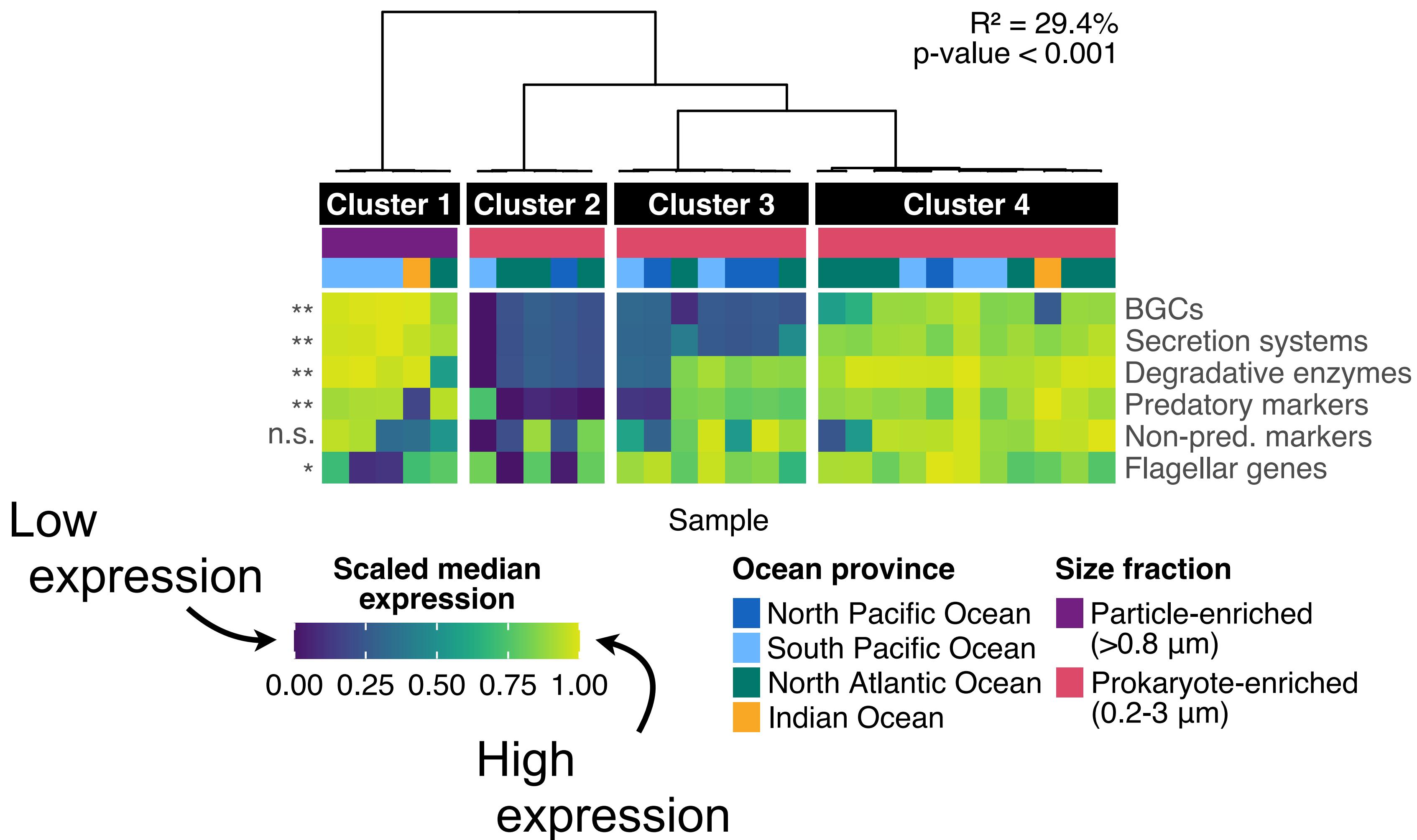


Ocean province	Size fraction
North Pacific Ocean	Particle-enriched (>0.8 μm)
South Pacific Ocean	Prokaryote-enriched (0.2-3 μm)
North Atlantic Ocean	
Indian Ocean	

Ecology with *in situ* metatranscriptomics



BGCs to support a predatory lifestyle?

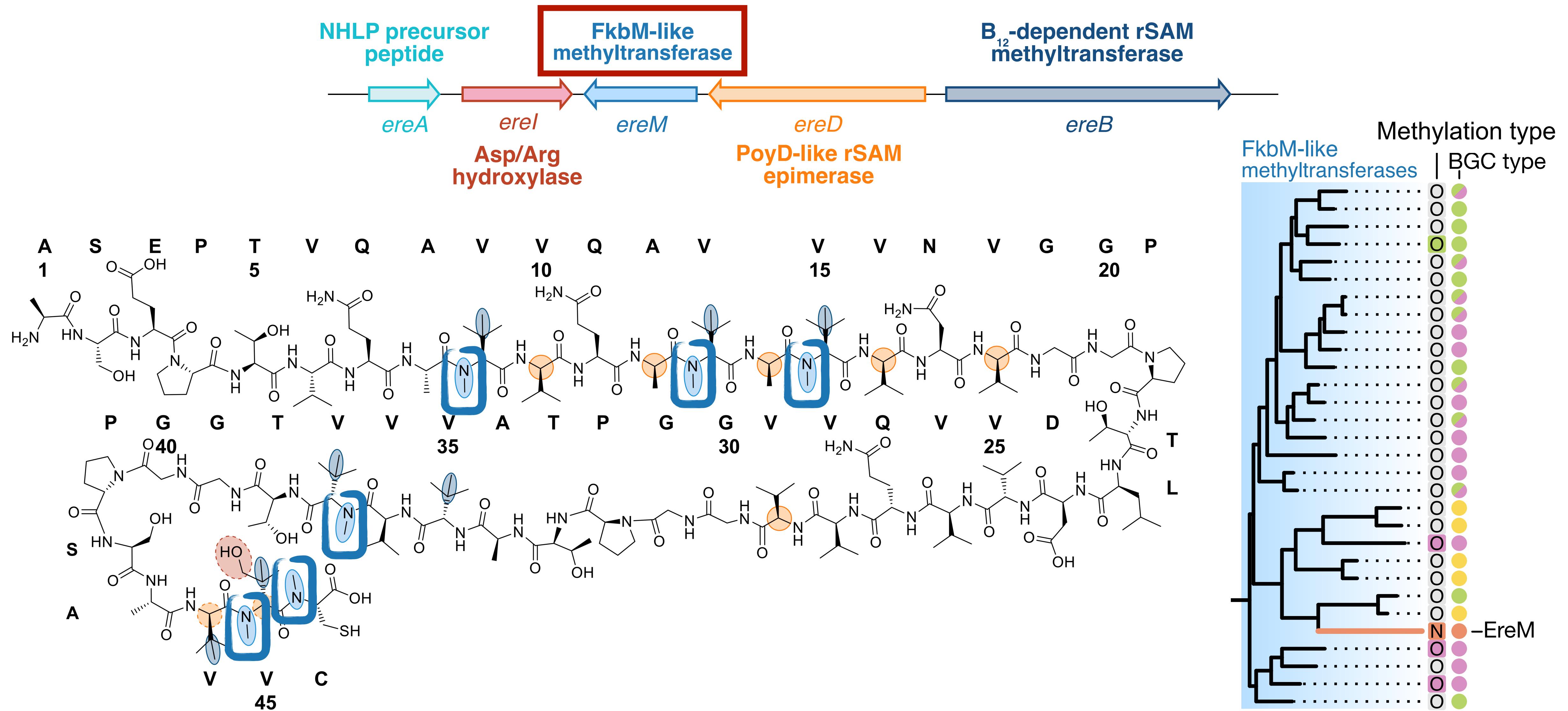


- High predatory scores
- Large BGC diversity
- Diverse secretion systems
- Degradative enzymes
- Motility

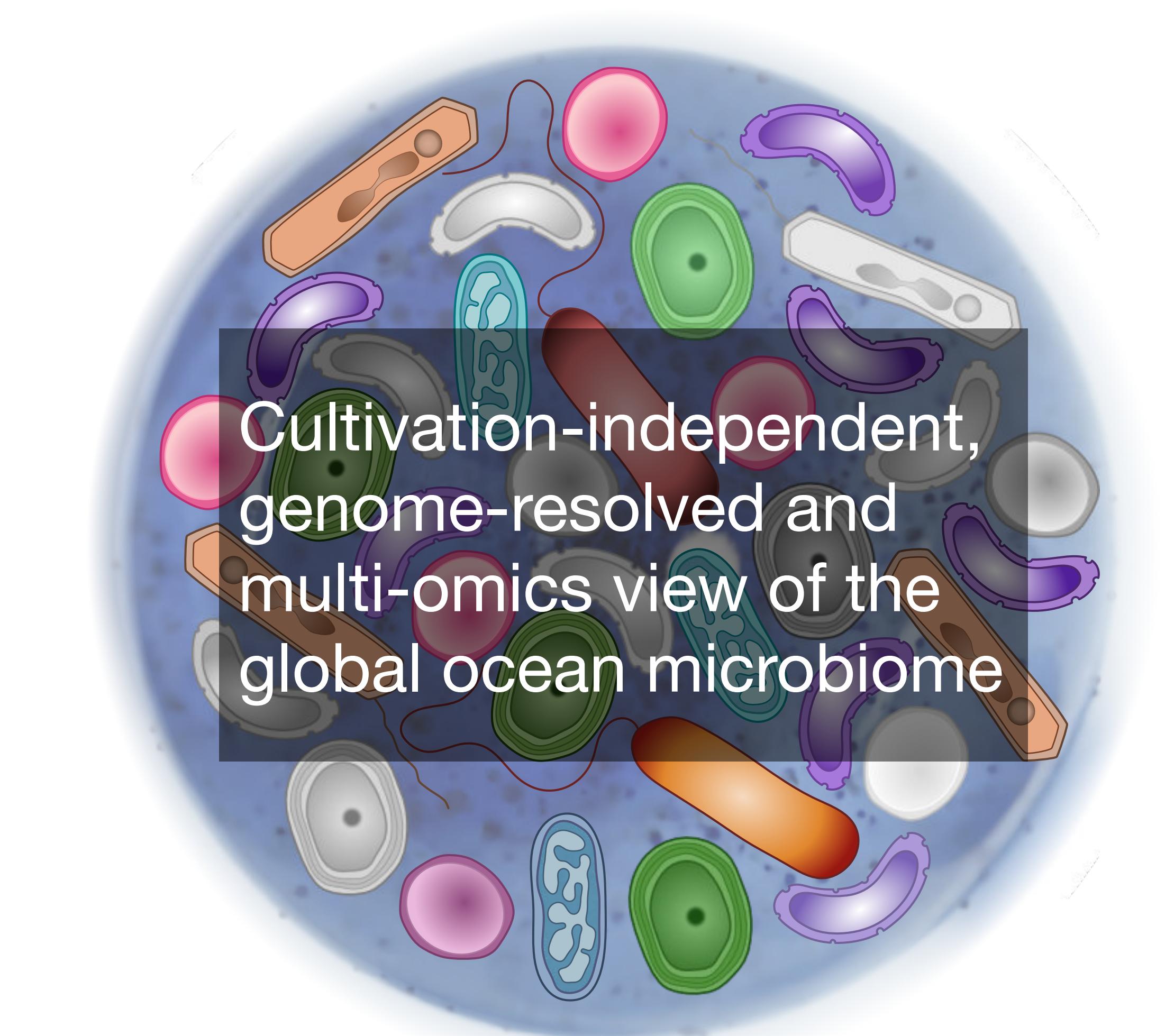


Is this computational approach
sufficiently powerful to predict new
enzymology and natural products?

With unique trans amide-N-methylation

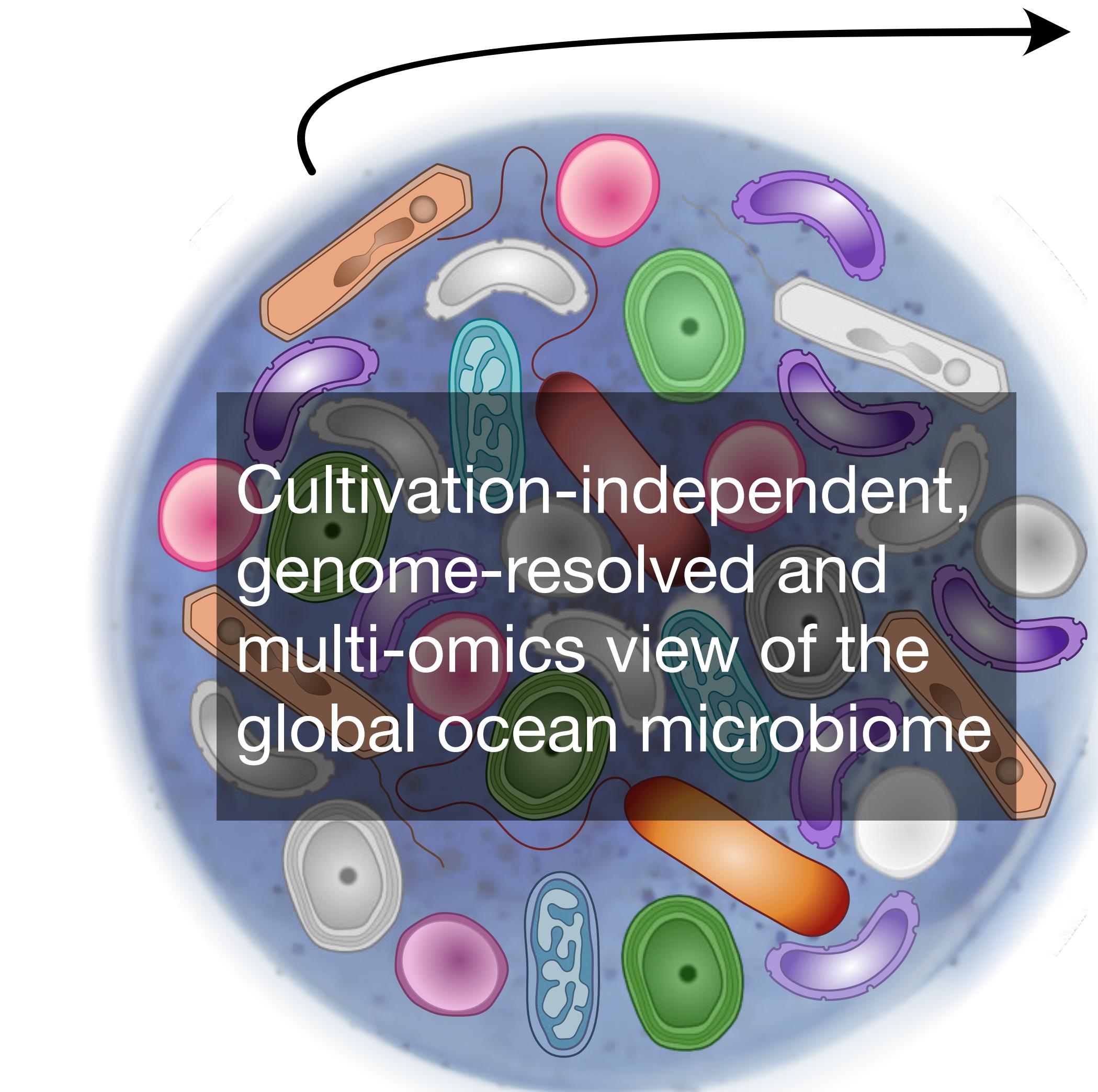


Conclusion



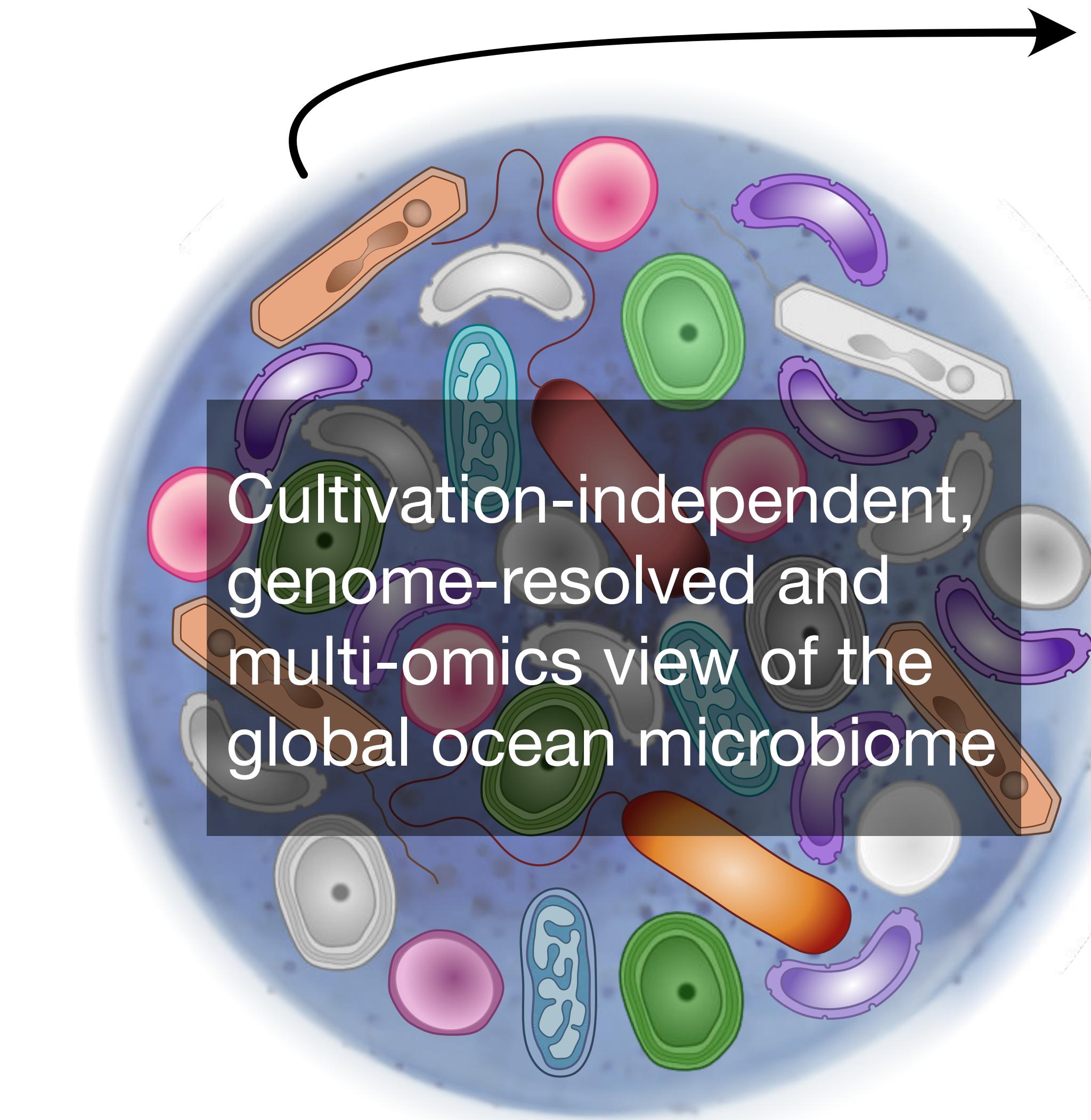
Cultivation-independent,
genome-resolved and
multi-omics view of the
global ocean microbiome

Conclusion

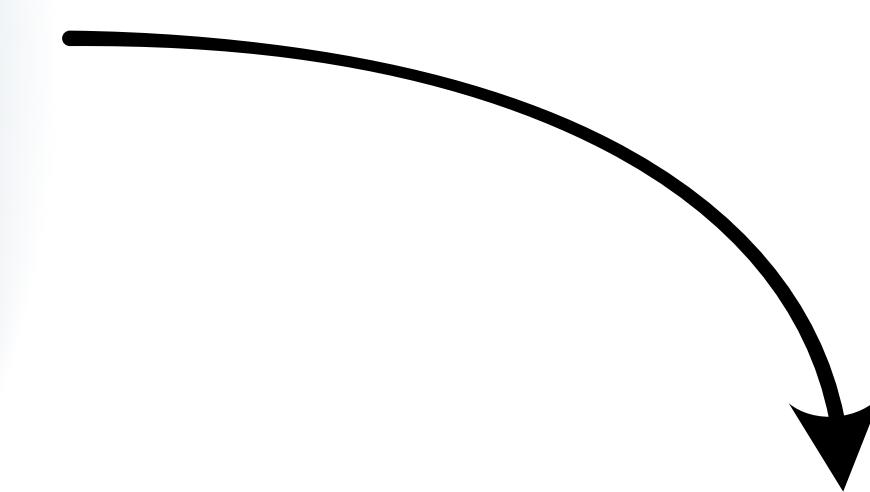


**Discovery of new
biosynthetic pathways,
enzymes and natural
products**

Conclusion



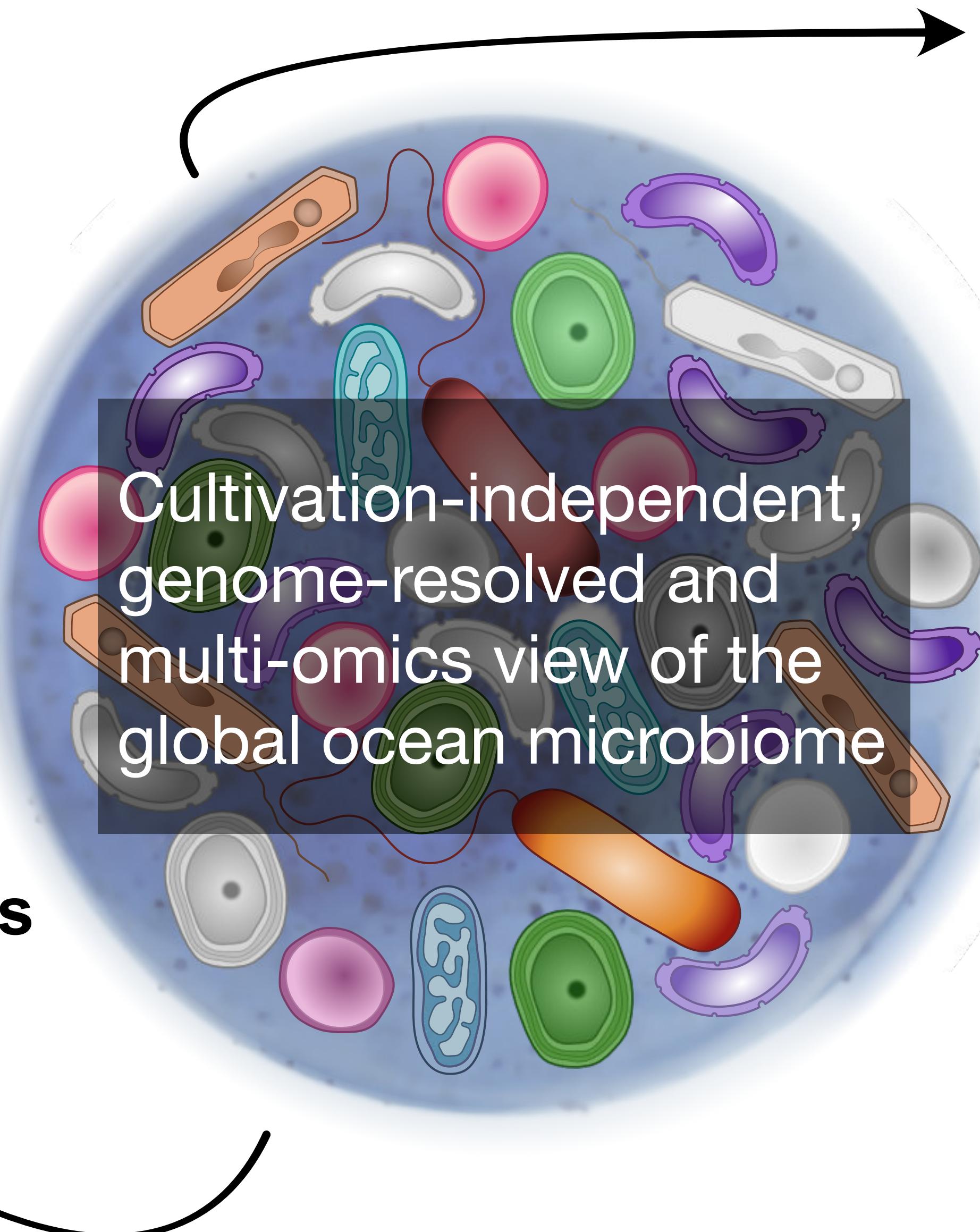
**Discovery of new
biosynthetic pathways,
enzymes and natural
products**



**Provide evolutionary and
ecological context to the
biosynthetic potential**

Conclusion

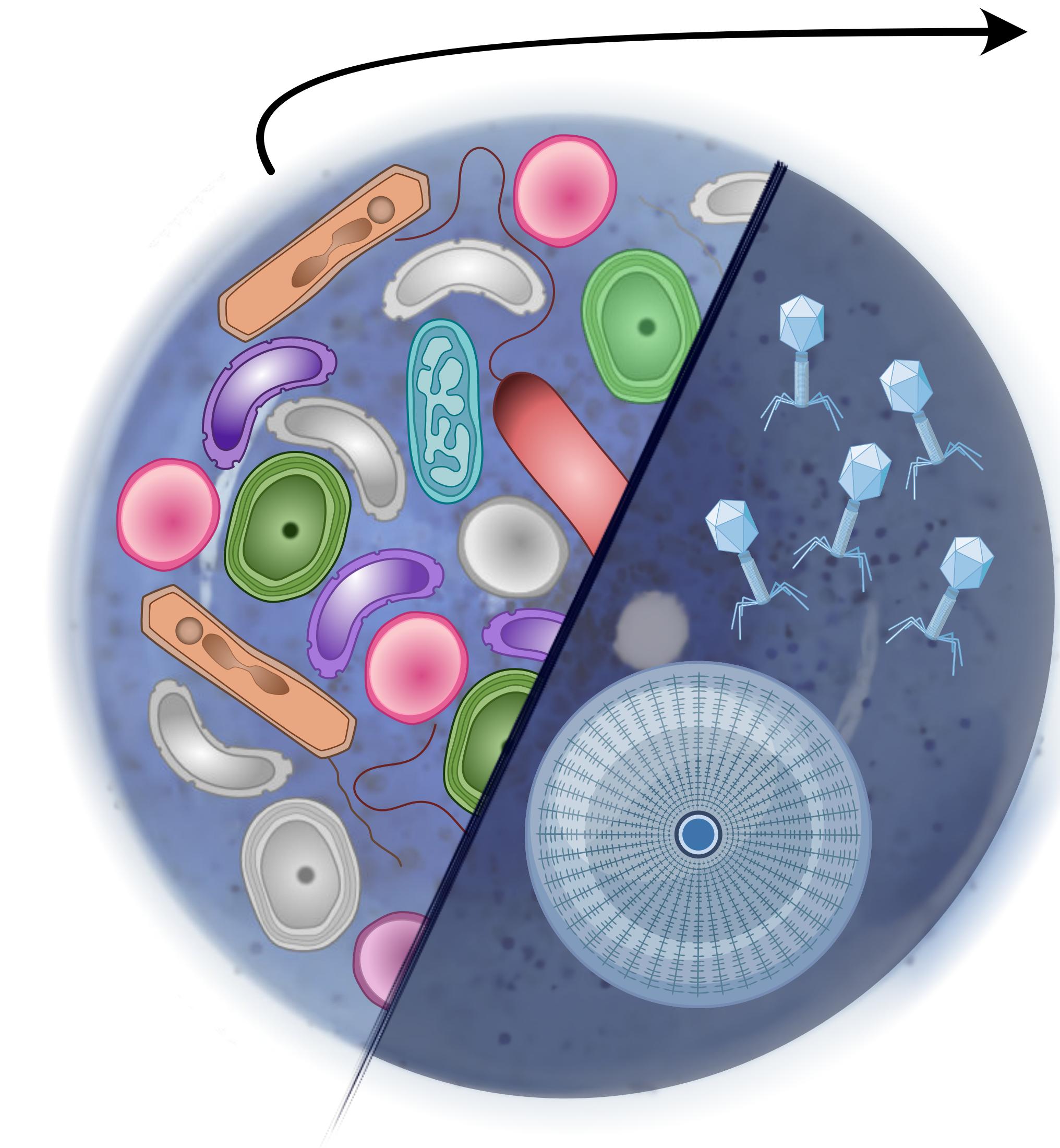
Bioinformatics-guided
experimental
characterization still leads
to unpredicted discoveries



Discovery of new biosynthetic pathways, enzymes and natural products

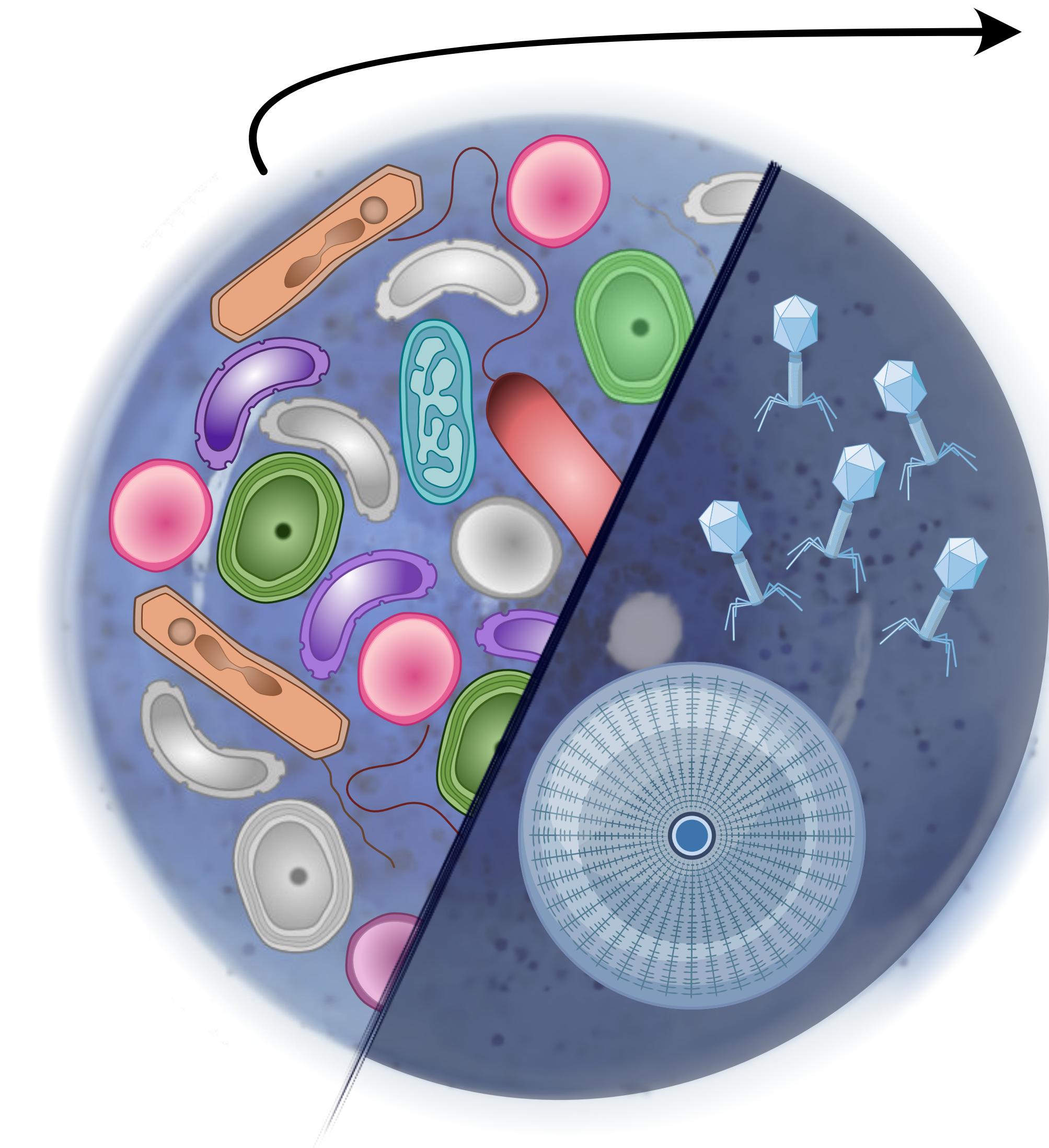
Provide evolutionary and ecological context to the biosynthetic potential

Outlook

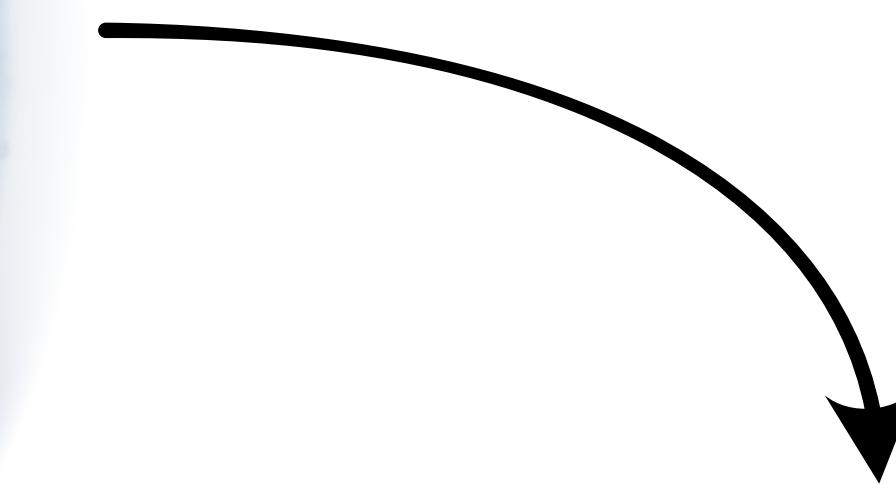


Only **two** out of **thousands** of pathways have been characterized

Outlook



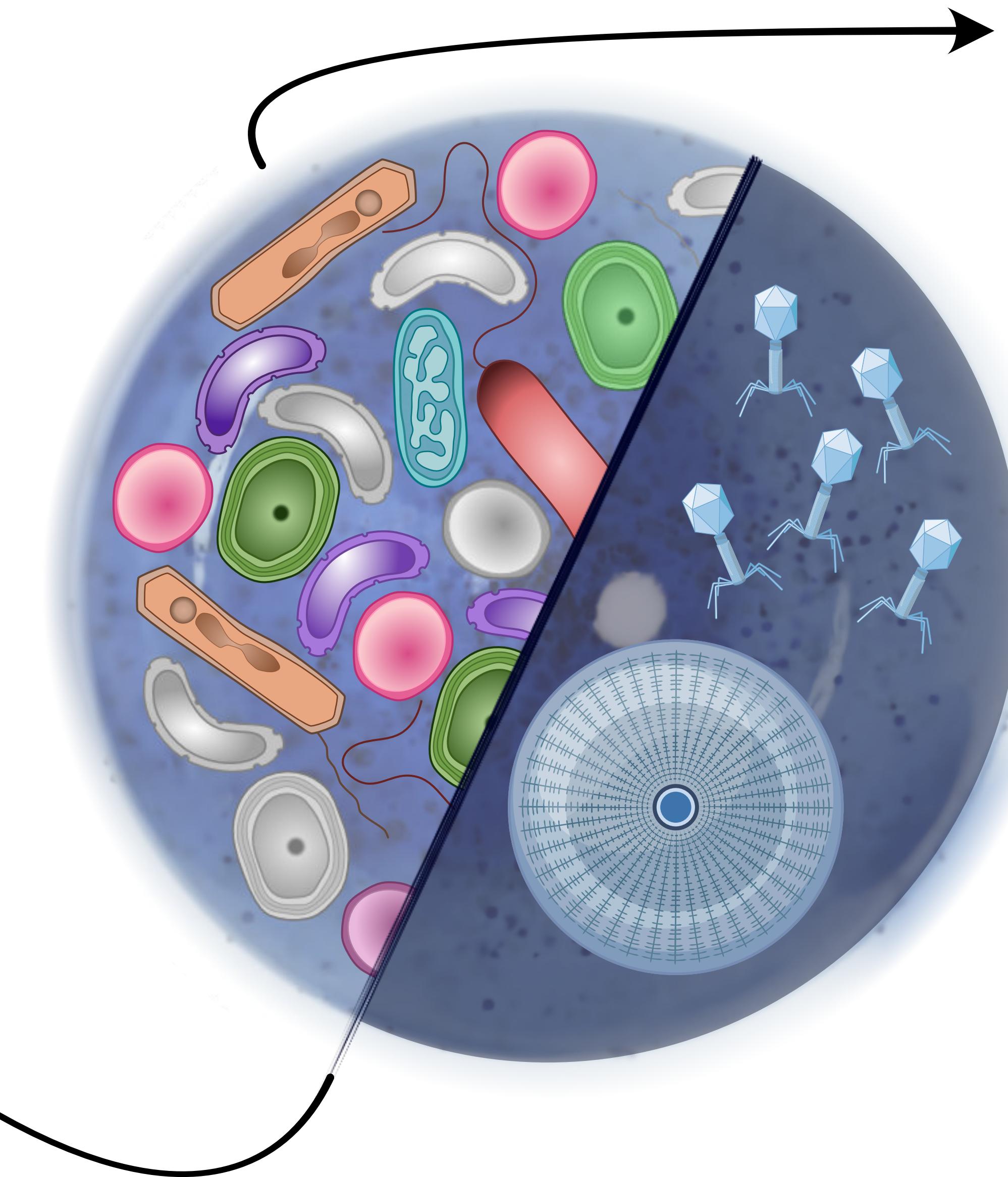
Only **two** out of **thousands** of pathways have been characterized



Integrate the biosynthetic potential of **other members** of the microbiome

Outlook

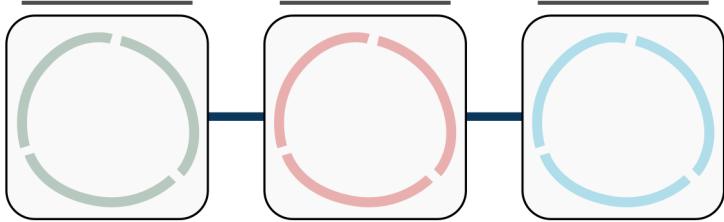
**Ecological functions
(biotic interactions)
and applications
of these natural products**



**Only two out of thousands
of pathways have been
characterized**

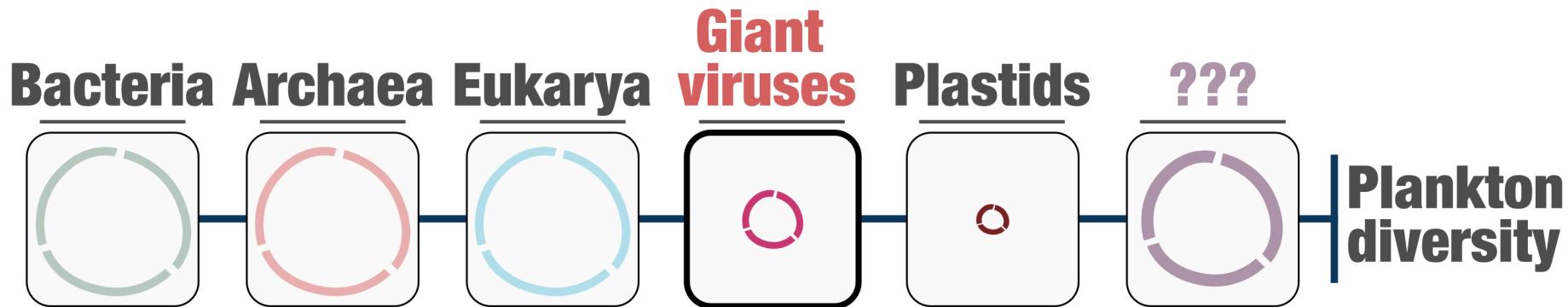
**Integrate the biosynthetic
potential of other members
of the microbiome**

Bacteria Archaea Eukarya



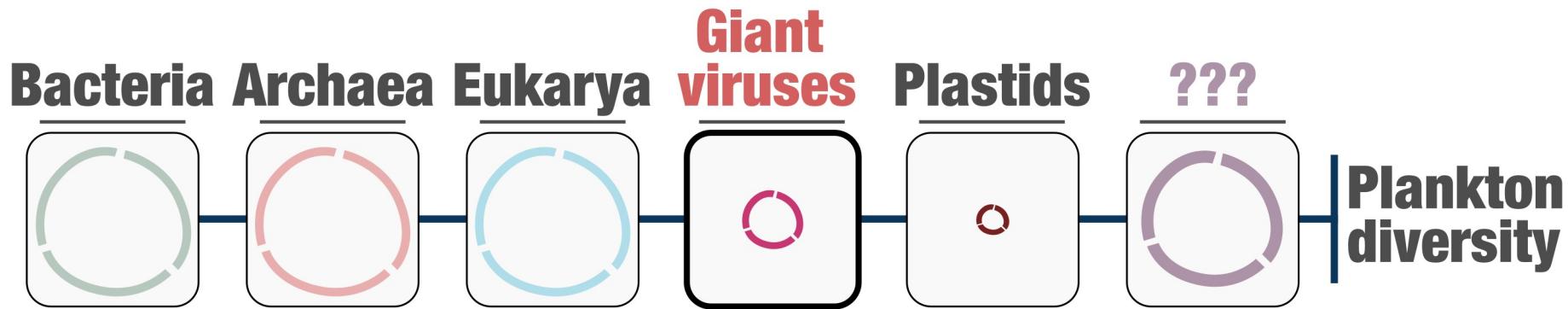
From 2015 to 2019

**Single copy core gene
collections for the
three domains
of life**



From 2015 to 2019

Single copy core gene collections for the three domains of life



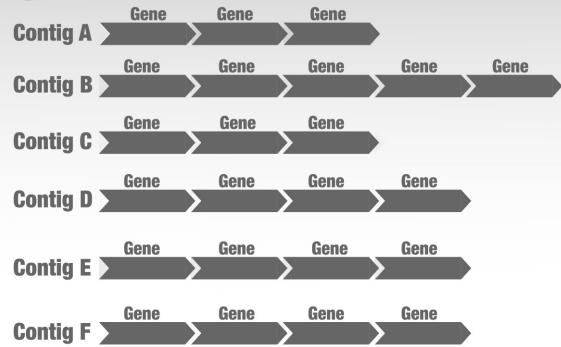
Starting
in 2019

The remarkable **RNA polymerase genes**

Patrick Forterre
Morgan Gaïa

A compass for
targeted binning

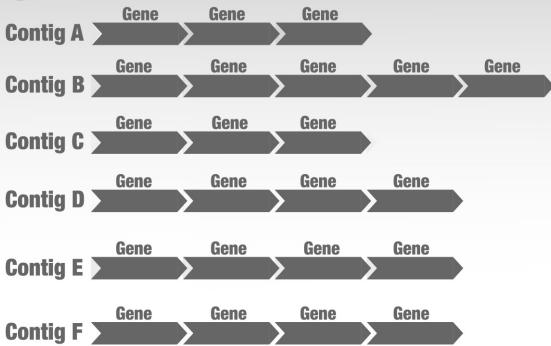
A Metagenomic assemblies



~300 billion metagenomic reads
11 large co-assemblies by region
~12 million contigs >2.5kb

Tara Oceans

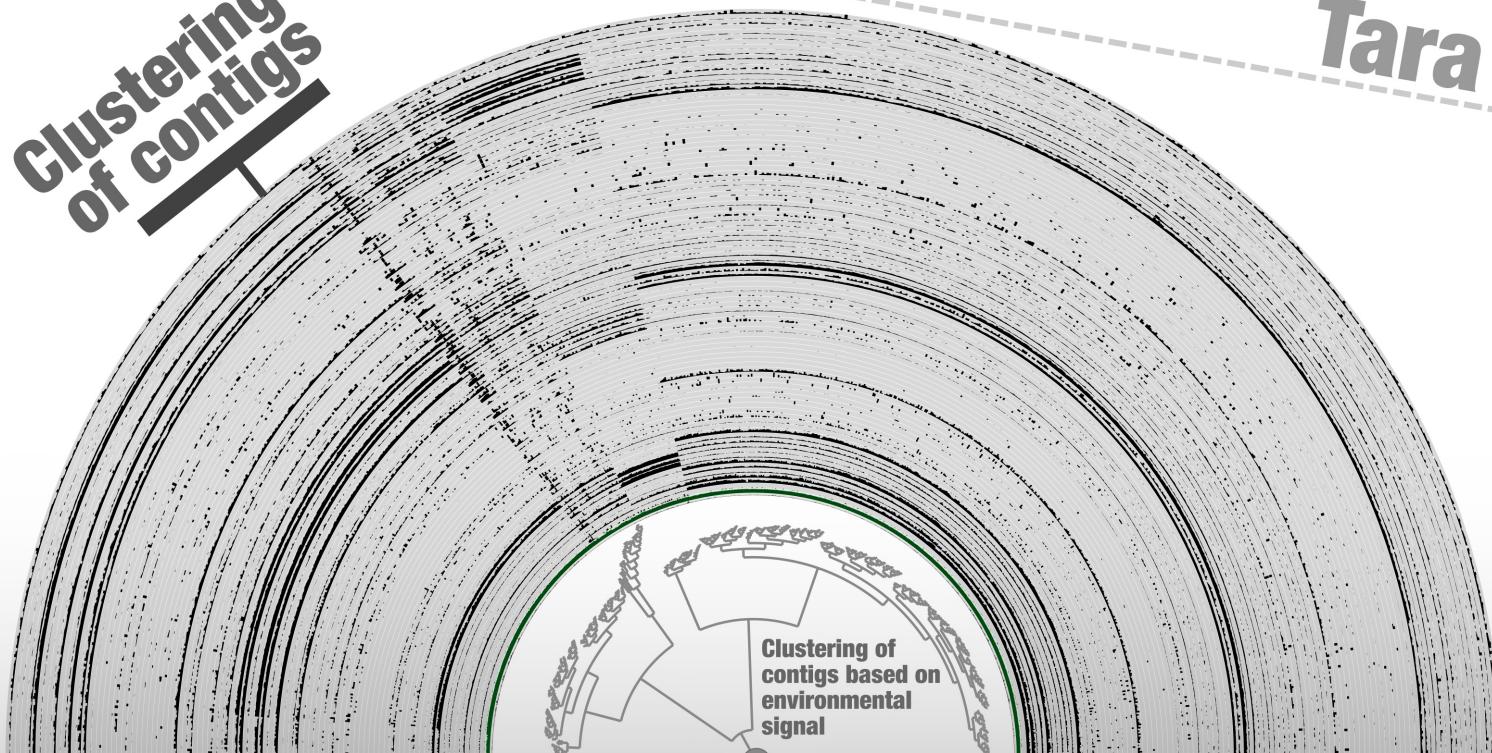
A Metagenomic assemblies



~300 billion metagenomic reads
11 large co-assemblies by region
~12 million contigs >2.5kb

Clustering
of contigs

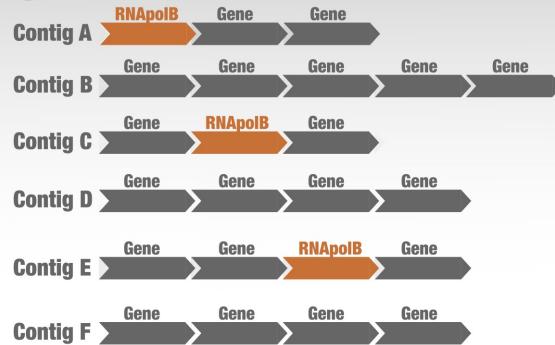
Tara Oceans
Anvi'o



C

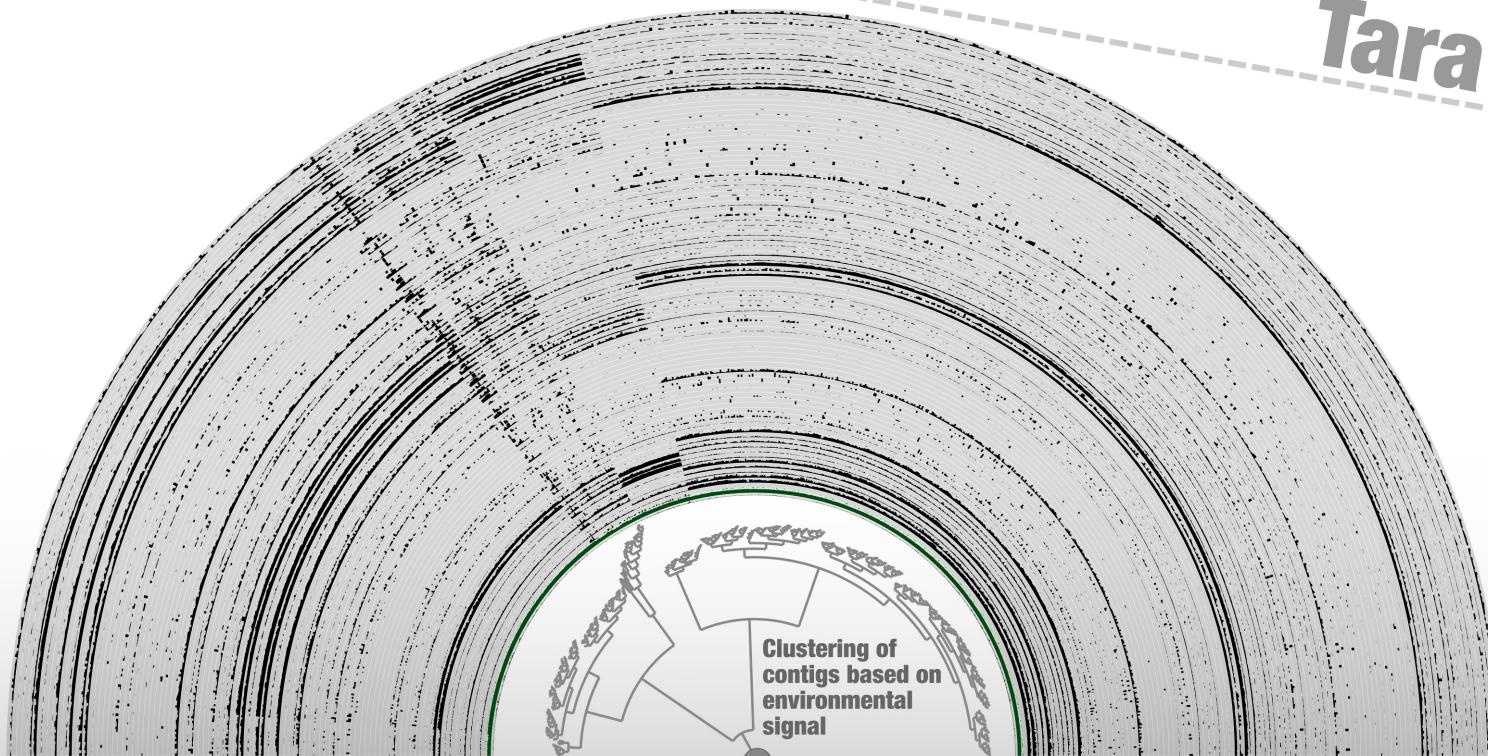
Genome-resolved metagenomics

A Metagenomic assemblies



Search for the DNA-dependent RNA polymerase B (**RNApolB**) using a global HMM

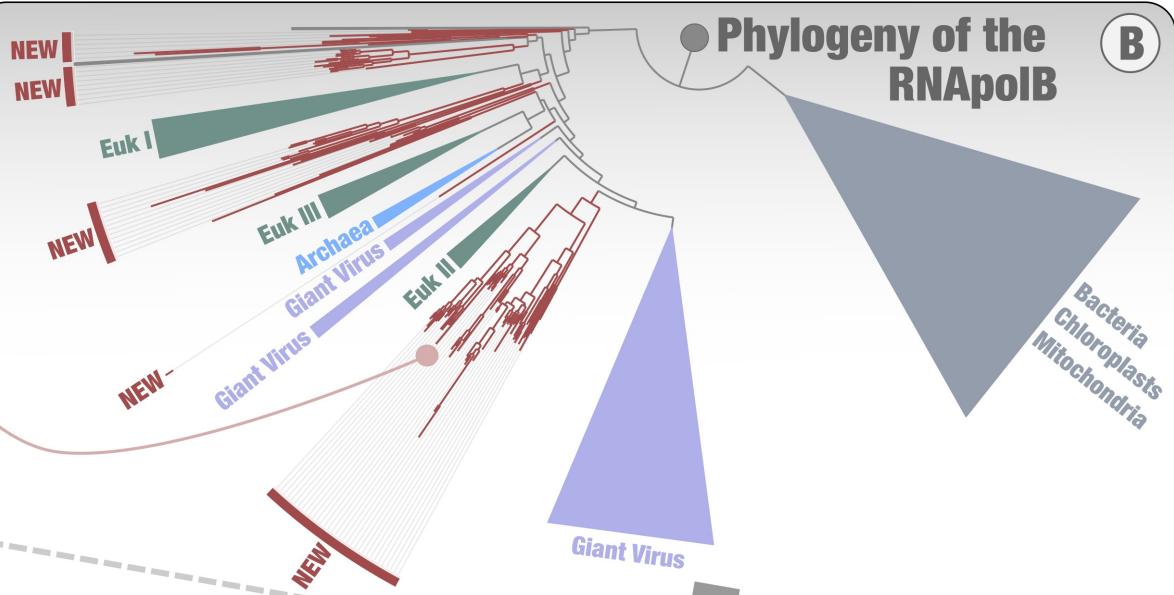
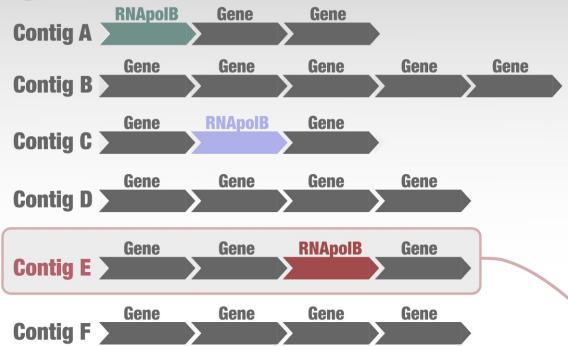
Tara Oceans
Anvi'o



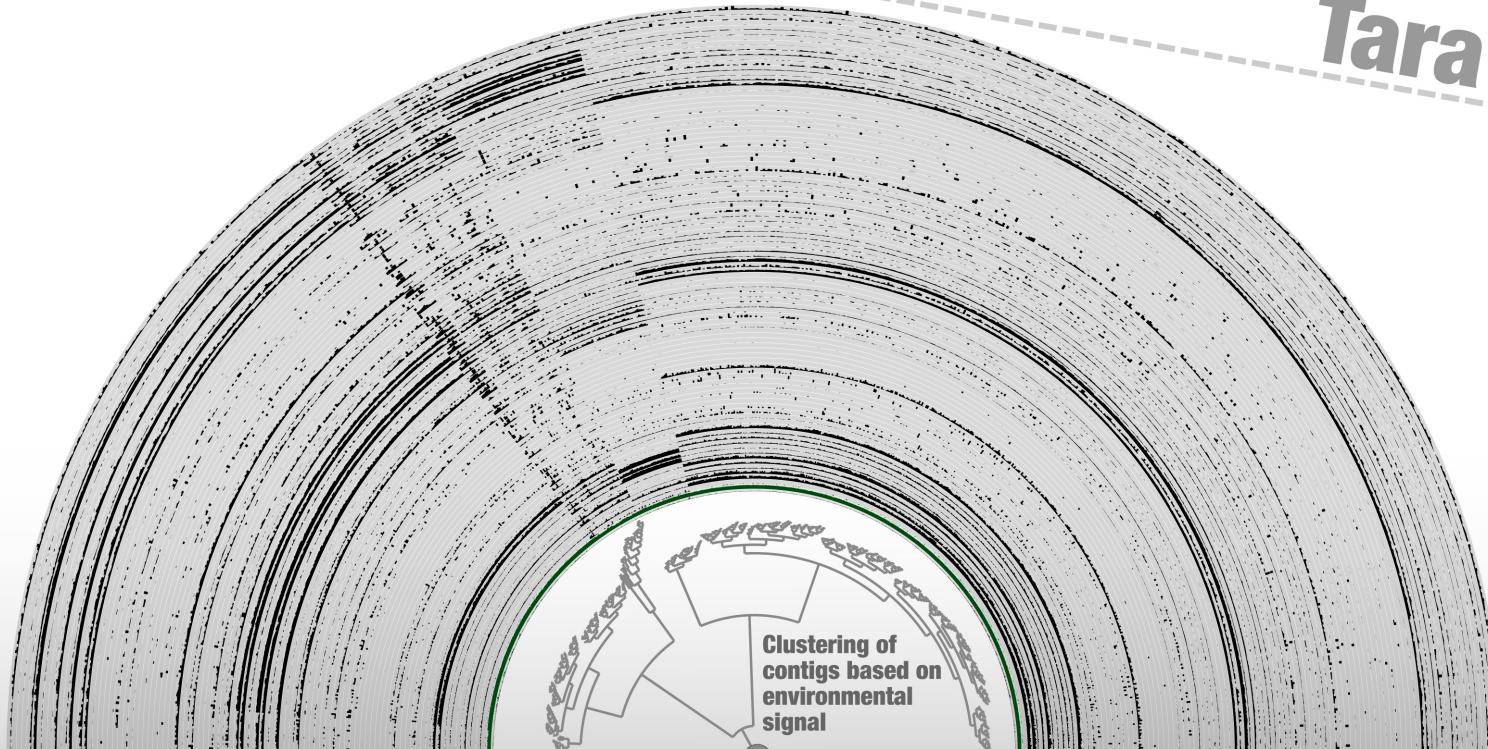
C

Genome-resolved metagenomics

A Metagenomic assemblies



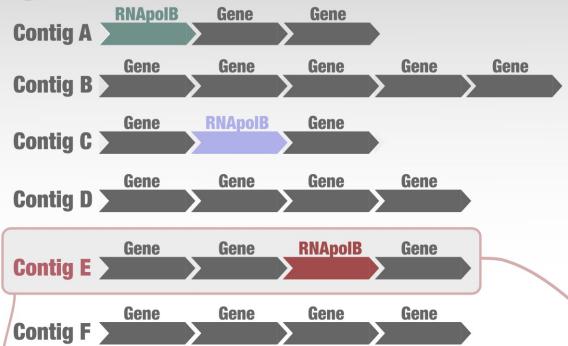
Tara Oceans
Anvi'o



Genome-resolved metagenomics

C

A Metagenomic assemblies

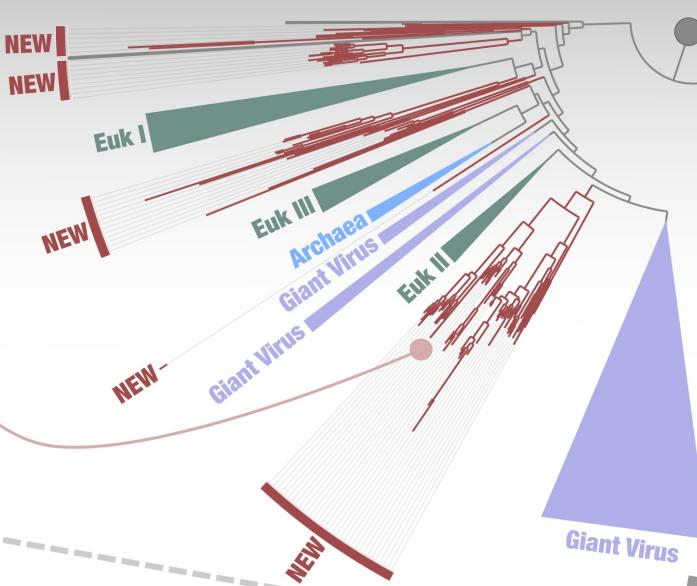


Population genome

50 genomic fragments
~370,000 nt in length

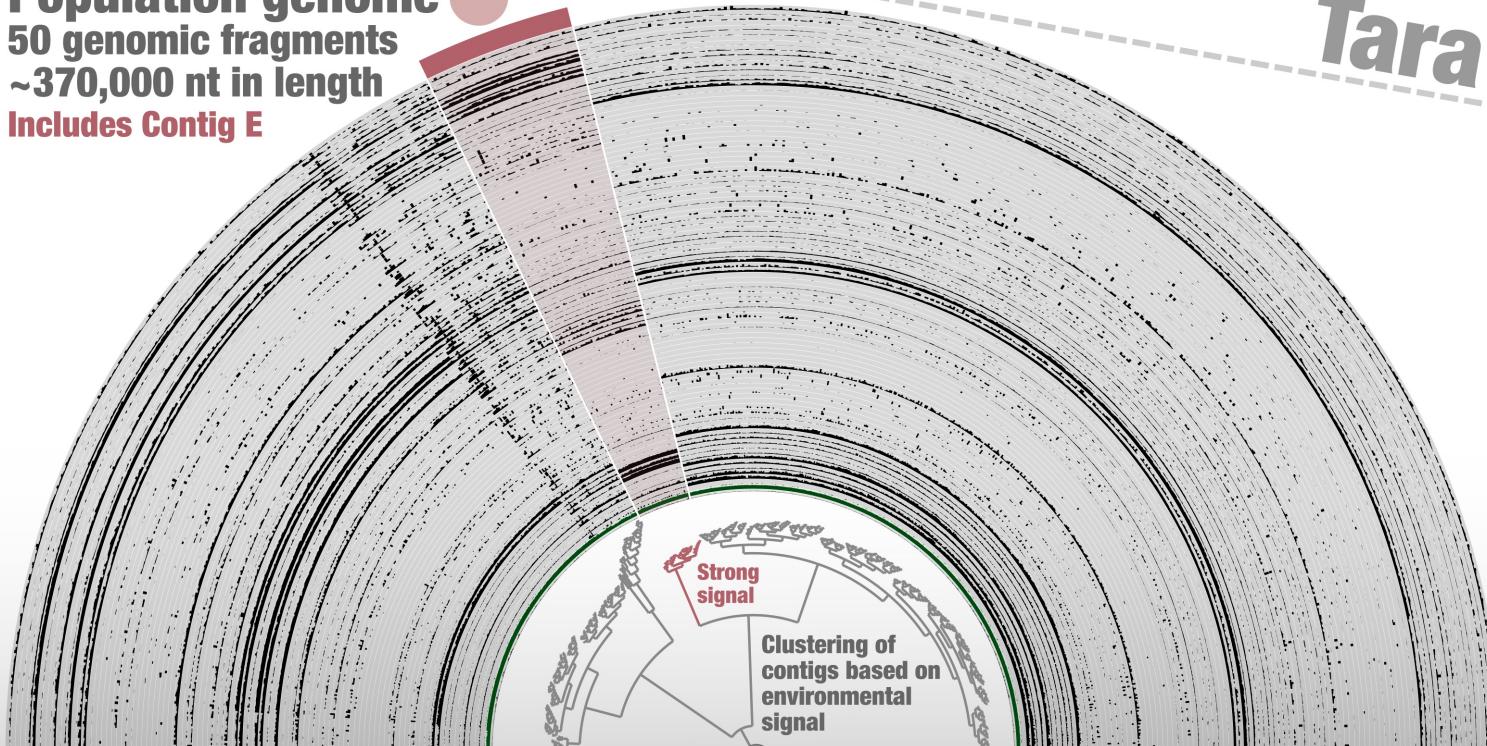
Includes Contig E

B Phylogeny of the RNApolB



Bacteria
Chloroplasts
Mitochondria

Tara Oceans
Anvi'o



Phylogeny-guided genome-resolved metagenomics

C

Article | [Open Access](#) | Published: 19 April 2023

Mirusviruses link herpesviruses to giant viruses

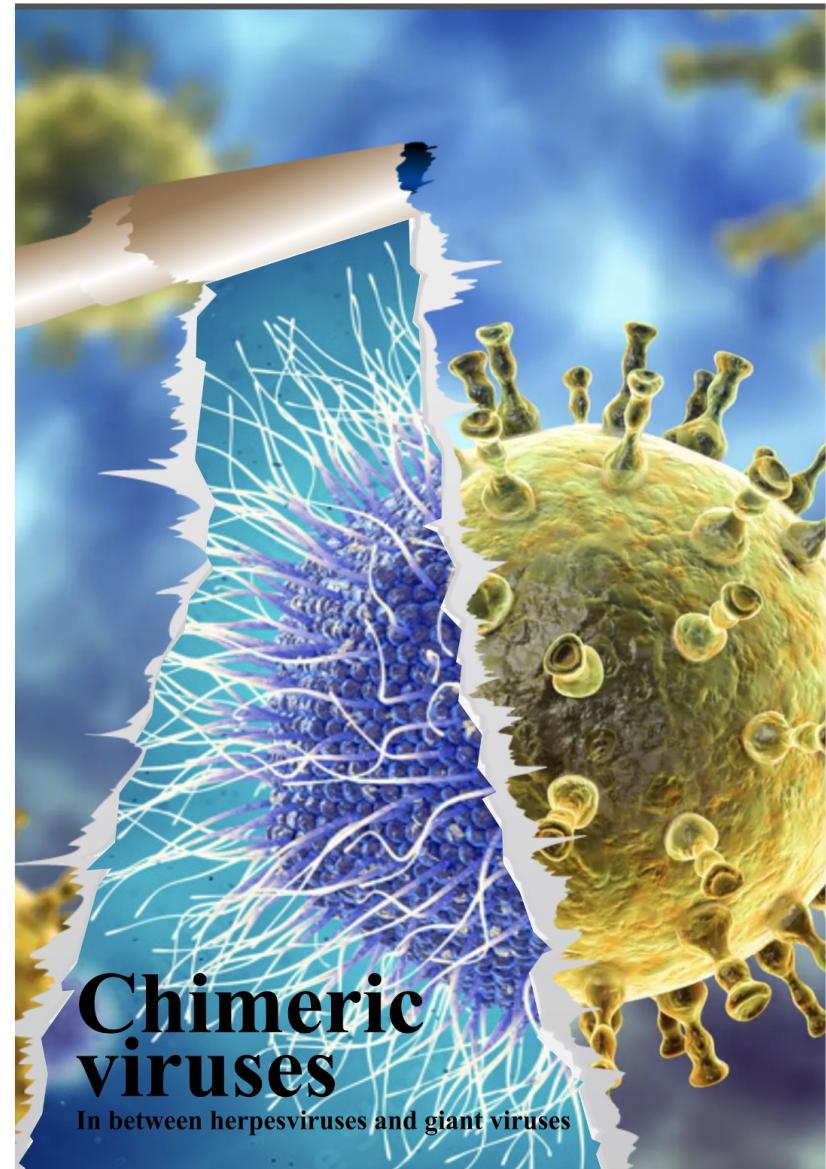
Morgan Gaia, Lingjie Meng, Eric Pelletier, Patrick Forterre, Chiara Vanni, Antonio Fernandez-Guerra, Olivier Jaillon, Patrick Wincker, Hiroyuki Ogata, Mart Krupovic & Tom O. Delmont 

[Nature](#) 616, 783–789 (2023) | [Cite this article](#)

19k Accesses | 6 Citations | 535 Altmetric | [Metrics](#)

Abstract

DNA viruses have a major influence on the ecology and evolution of cellular organisms^{1,2,3,4} but their overall diversity and evolutionary trajectories remain elusive⁵. Here we carried out a phylogeny-guided genome-resolved metagenomic survey of the sunlit oceans and discovered plankton-infecting relatives of herpesviruses that form a putative new phylum dubbed *Mirusviricota*. The virion morphogenesis module of this large monophyletic clade is typical of viruses from the realm *Duplodnaviria*⁶, with multiple components strongly indicating a common ancestry with animal-infecting *Herpesvirales*. Yet, a substantial fraction of mirusvirus genes, including hallmark transcription machinery genes missing in herpesviruses, are closely related homologues of giant eukaryotic DNA viruses from another viral realm, *Varidnaviria*. These remarkable chimaeric attributes connecting *Mirusviricota* to herpesviruses and giant eukaryotic viruses are supported by more than 100 environmental mirusvirus genomes, including a near-complete contiguous genome of 432 kilobases. Moreover, mirusviruses are among the most abundant and active eukaryotic viruses characterized in the sunlit oceans, encoding a diverse array of functions used during the infection of microbial eukaryotes from pole to pole. The prevalence, functional activity, diversification and atypical chimaeric attributes of mirusviruses point to a lasting role of *Mirusviricota* in the ecology of marine ecosystems and in the evolution of eukaryotic DNA viruses.



Nearly ten years of environmental genomics

Methodological developments

Platform dedicated
to genome-resolved
metagenomics



Environmental genomics of plankton in the sunlit oceans

2015

Tara Oceans
metagenomes
are made public

2016

2017

2018

Bacteria | Small
Archaea | size fraction

2019

2020

2021

2022

Bacteria & Archaea | Large
Eukaryotes (up to >1Gbp) | size fractions

2023

2024

2025



Nearly ten years of environmental genomics

Methodological developments



Platform dedicated
to genome-resolved
metagenomics

Evolution
as a compass
Phylogeny-guided
genome-resolved
metagenomics

Environmental genomics of plankton in the sunlit oceans

2015

Tara Oceans
metagenomes
are made public

2016

2017

Bacteria | Small
Archaea | size fraction

2019

2020

2021

2022

Bacteria & Archaea | Large
Eukaryotes (up to >1Gbp) | size fractions

2023

2024

2025



Nearly ten years of environmental genomics

Methodological developments



Platform dedicated
to genome-resolved
metagenomics

Evolution
as a compass
Phylogeny-guided
genome-resolved
metagenomics

Environmental genomics of plankton in the sunlit oceans

2015

Tara Oceans
metagenomes
are made public

2016

2017

Bacteria | Small
Archaea | size fraction

2019

2020

2021

2022

Bacteria & Archaea | Large
Eukaryotes (up to >1Gbp) | size fractions

2023

Giant viruses
Discovery of the mirusviruses

2024

Chloroplasts
Nucleomorphs
More classes of mirusviruses

2025



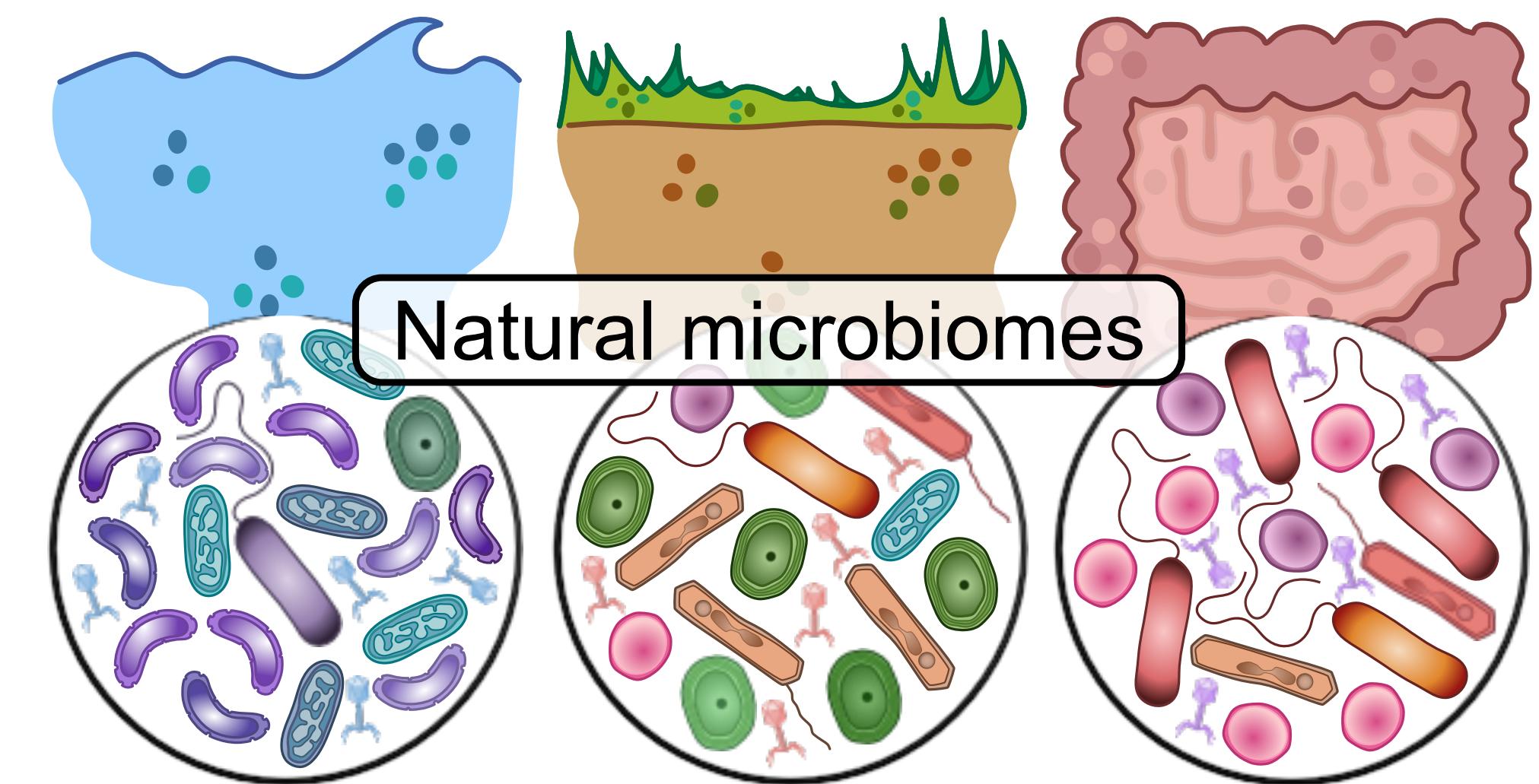
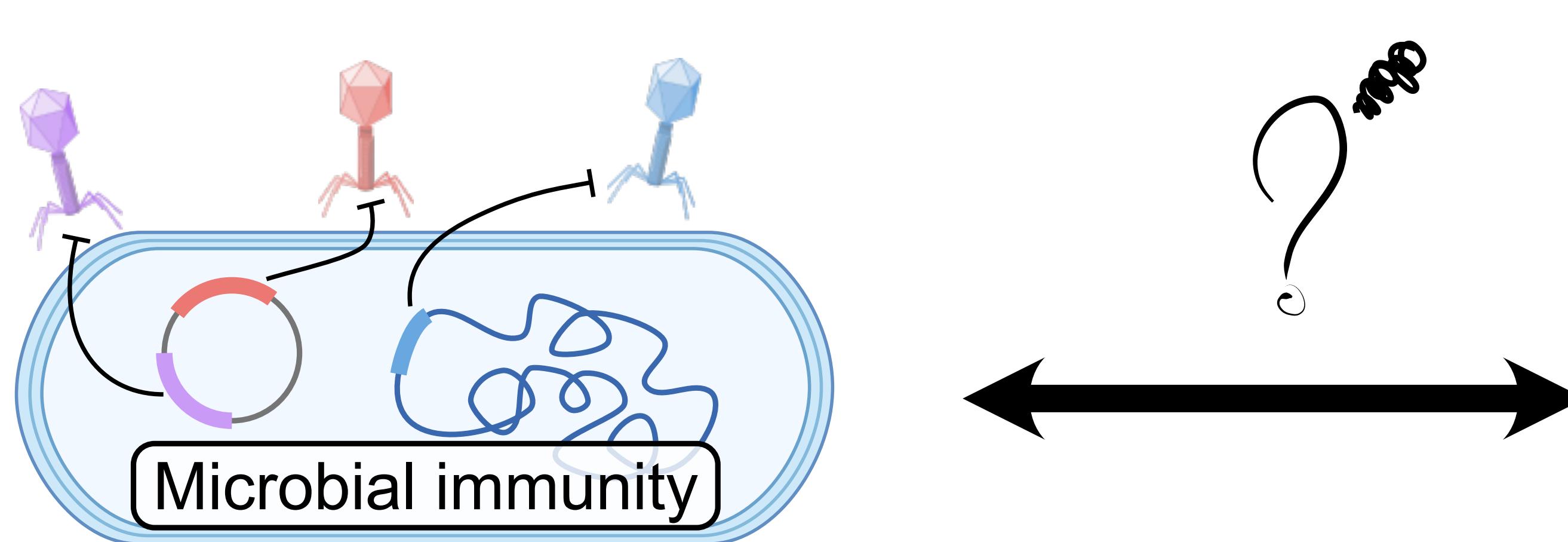
Master & PhD positions!



EPFL

Lab opening at EPFL
(Lausanne, Switzerland),
Spring 2024!

lucas.paoli@pasteur.fr
lucas.paoli@gmail.com



Plankton genomics goes like the wind



Thank you for
your attention

