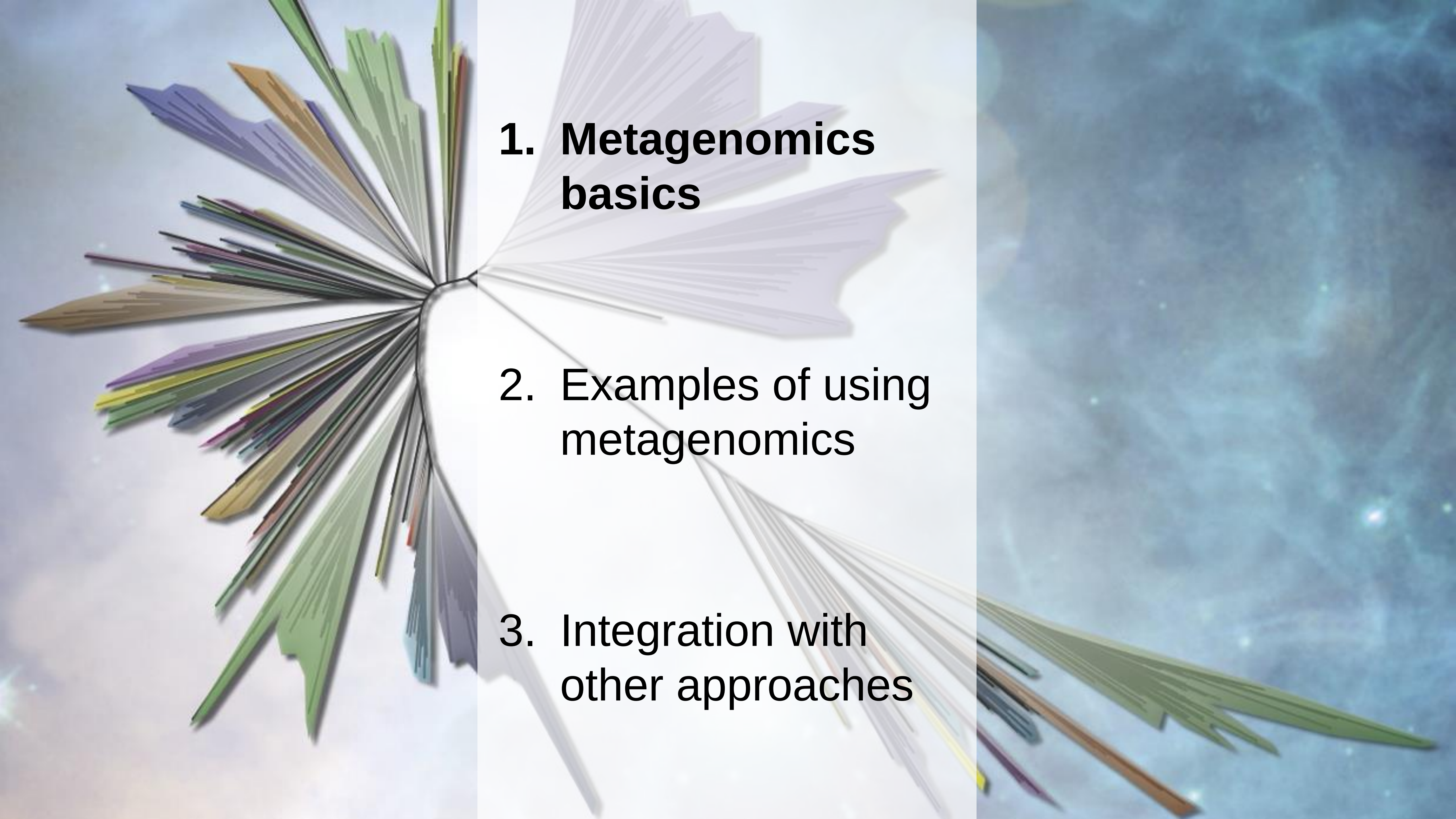
Metagenomics: a flexible tool for discovery

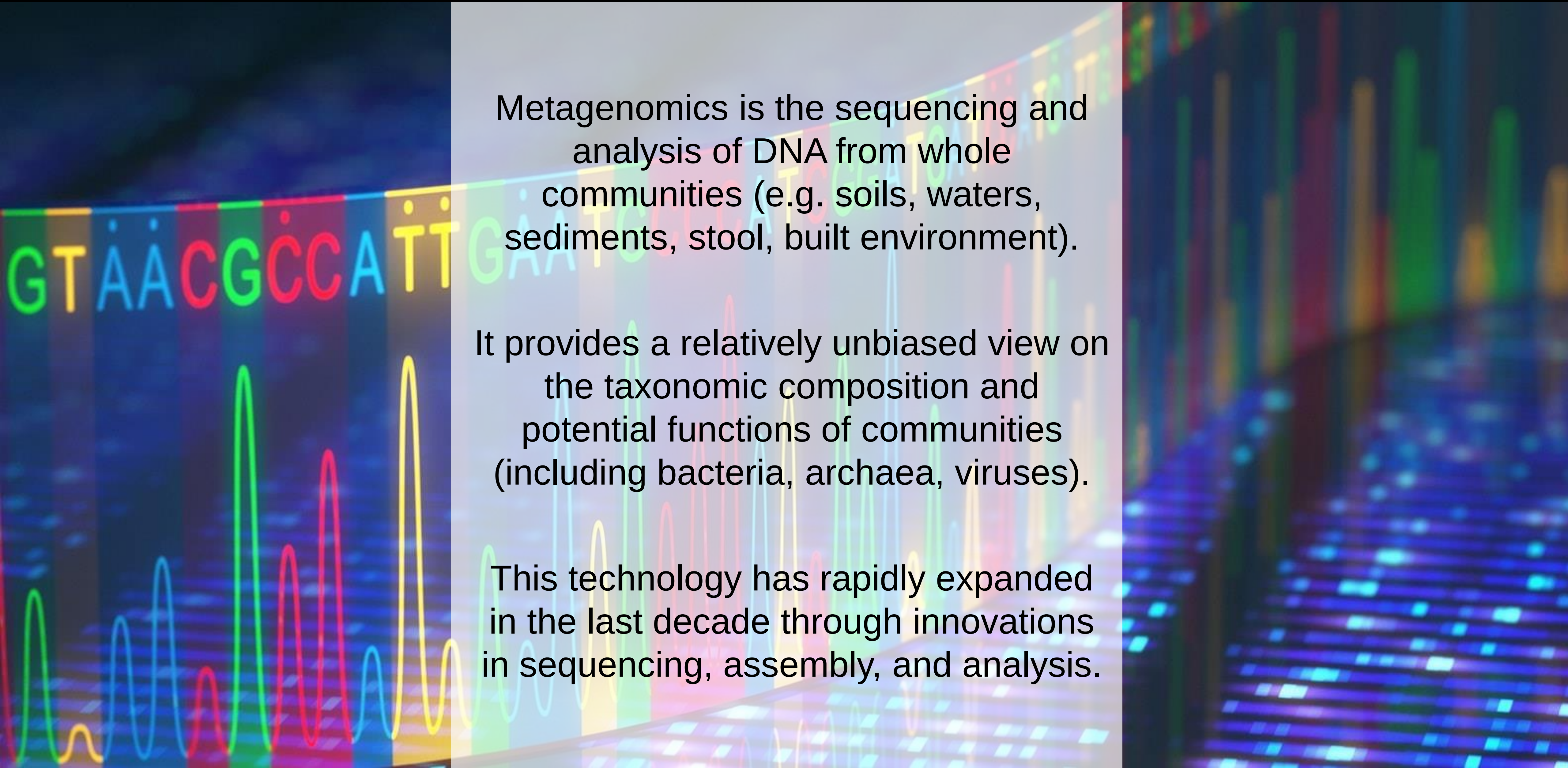
Assoc Prof Chris Greening
ARC DECRA Fellow & NHMRC EL2 Fellow
School of Biological Sciences
Monash University, Melbourne, Australia

MONASH
University



1. **Metagenomics basics**
2. Examples of using metagenomics
3. Integration with other approaches

What is metagenomics?

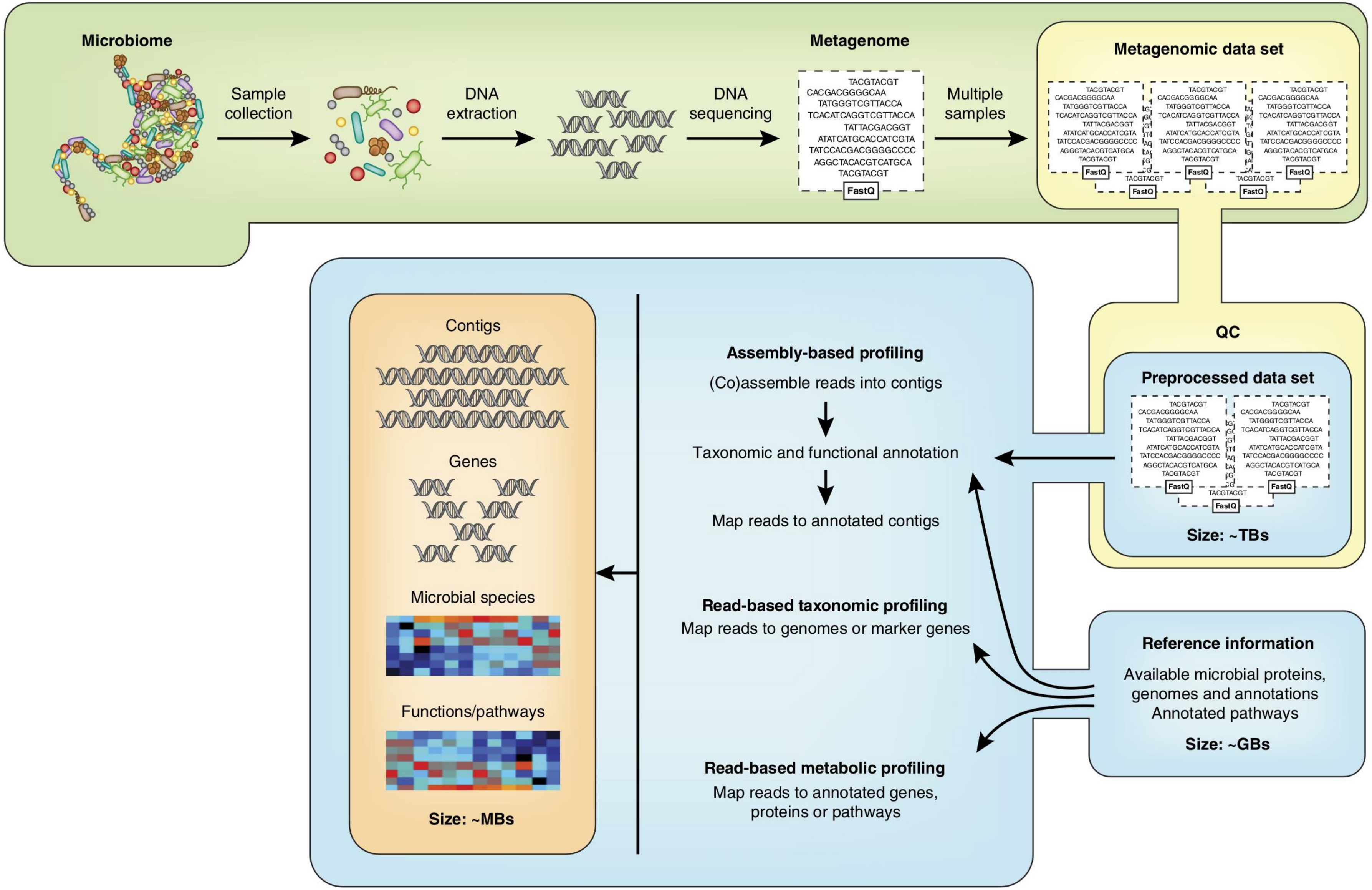


Metagenomics is the sequencing and analysis of DNA from whole communities (e.g. soils, waters, sediments, stool, built environment).

It provides a relatively unbiased view on the taxonomic composition and potential functions of communities (including bacteria, archaea, viruses).

This technology has rapidly expanded in the last decade through innovations in sequencing, assembly, and analysis.

A typical workflow for metagenome sequencing and analysis



Microbiome

Sample collection

DNA extraction

DNA sequencing

Metagenome

Multiple samples

Metagenomic data set

QC

Preprocessed data set

Size: ~TBs

Assembly-based profiling

(Co)assemble reads into contigs

Taxonomic and functional annotation

Map reads to annotated contigs

Read-based taxonomic profiling

Map reads to genomes or marker genes

Read-based metabolic profiling

Map reads to annotated genes, proteins or pathways

Reference information

Available microbial proteins, genomes and annotations

Annotated pathways

Size: ~GBs

Contigs

Genes

Microbial species

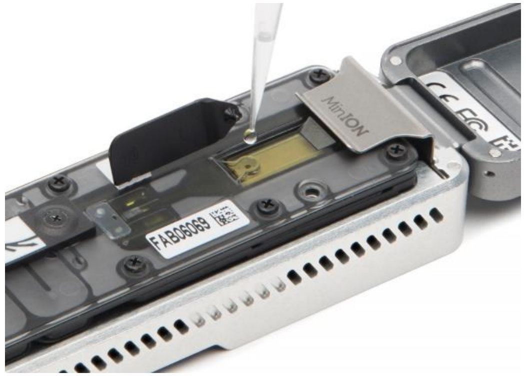
Functions/pathways

Size: ~MBs

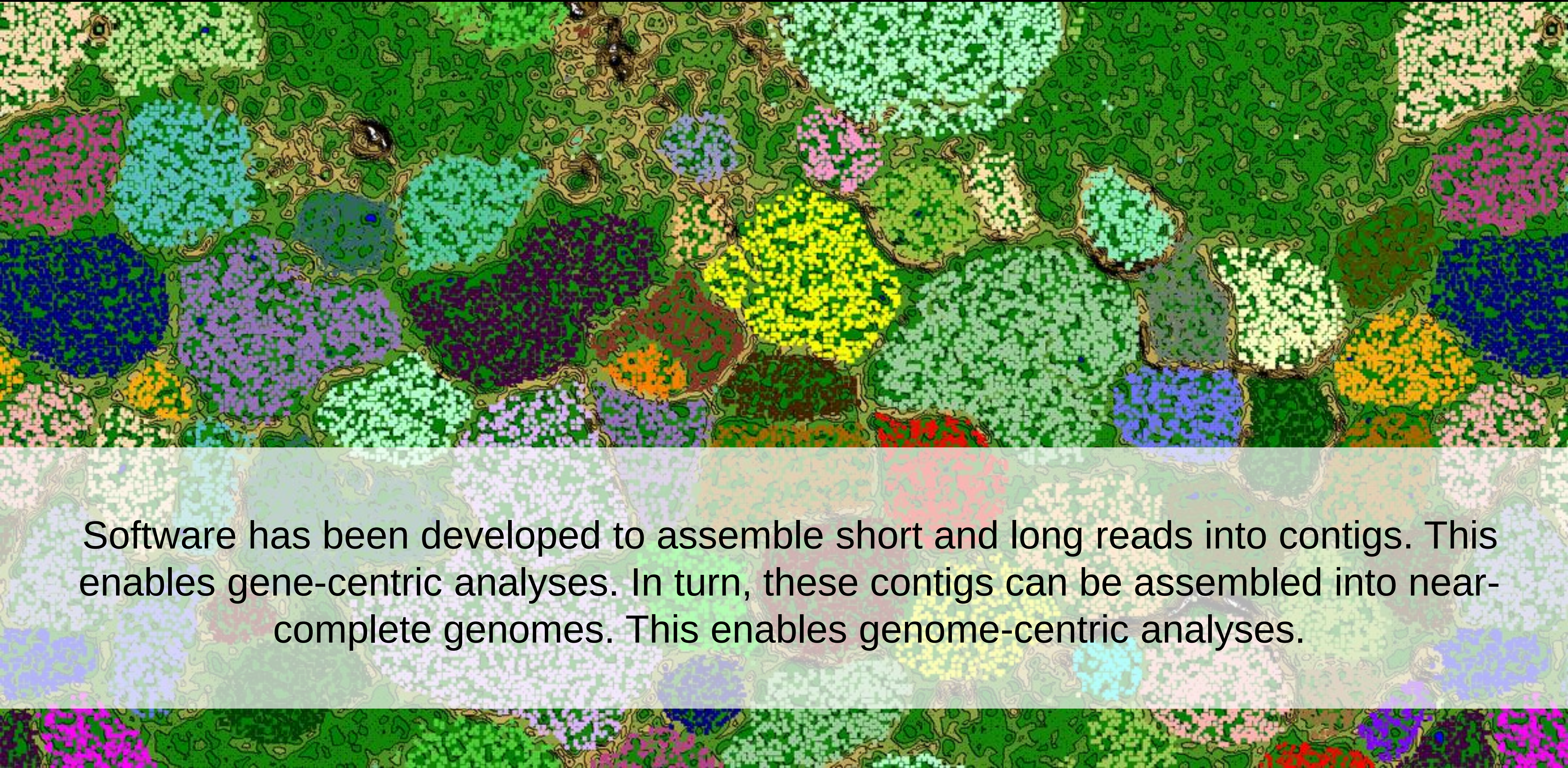
Quince et al (2017), Nat Biotechnol 35, 833-844

Two main types of sequencing available

	Short-read	Long-read
Platform	Illumina NextSeq, HiSeq, NovaSeq	Oxford Nanopore MinIlON, PacBio Sequel
Method	Amplified templates sequenced by synthesis	Single molecule flow or synthesis in real-time
Output	Short paired-end reads (2 × 150 bp)	Long reads of variable length (1 kbp to 1 mbp)
Error rate	~0.1%	~10%
Cost	$100/sample library prep, $60/Gbp sequence	~$250/Gbp library prep and sequencing

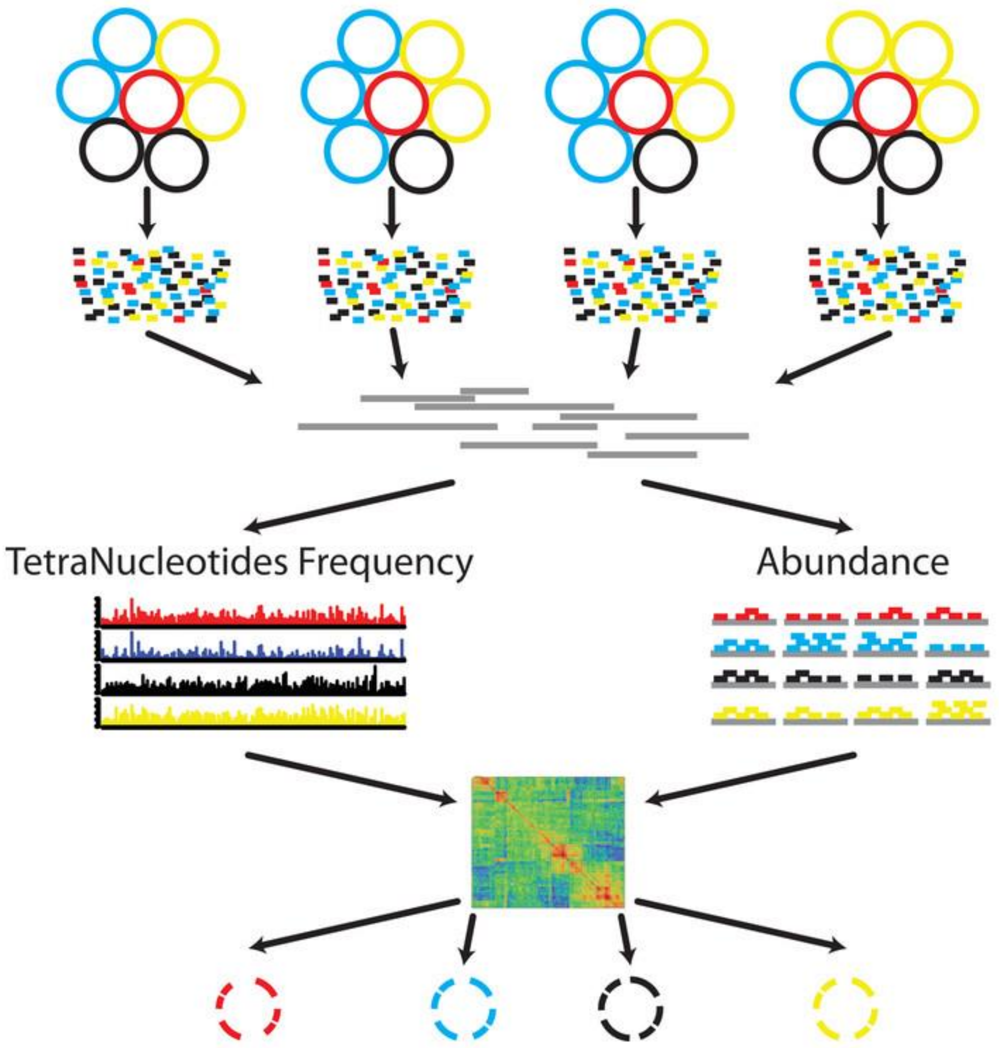


Gene-centric and genome-centric analyses



Software has been developed to assemble short and long reads into contigs. This enables gene-centric analyses. In turn, these contigs can be assembled into near-complete genomes. This enables genome-centric analyses.

Just one approach for assembly and binning



Sequencing and assembly

1. Samples from multiple sites or times
2. Metagenome libraries
3. Initial *de novo* individual or combined assembly

Binning with MetaBAT

4. Calculate TNF for each contig
5. Calculate abundance per library for each contig
6. Form and check genome bins iteratively

Kang et al (2015) PeerJ 3, e1165

From there it’s time to start answering the big questions

What is the composition and diversity of the community?

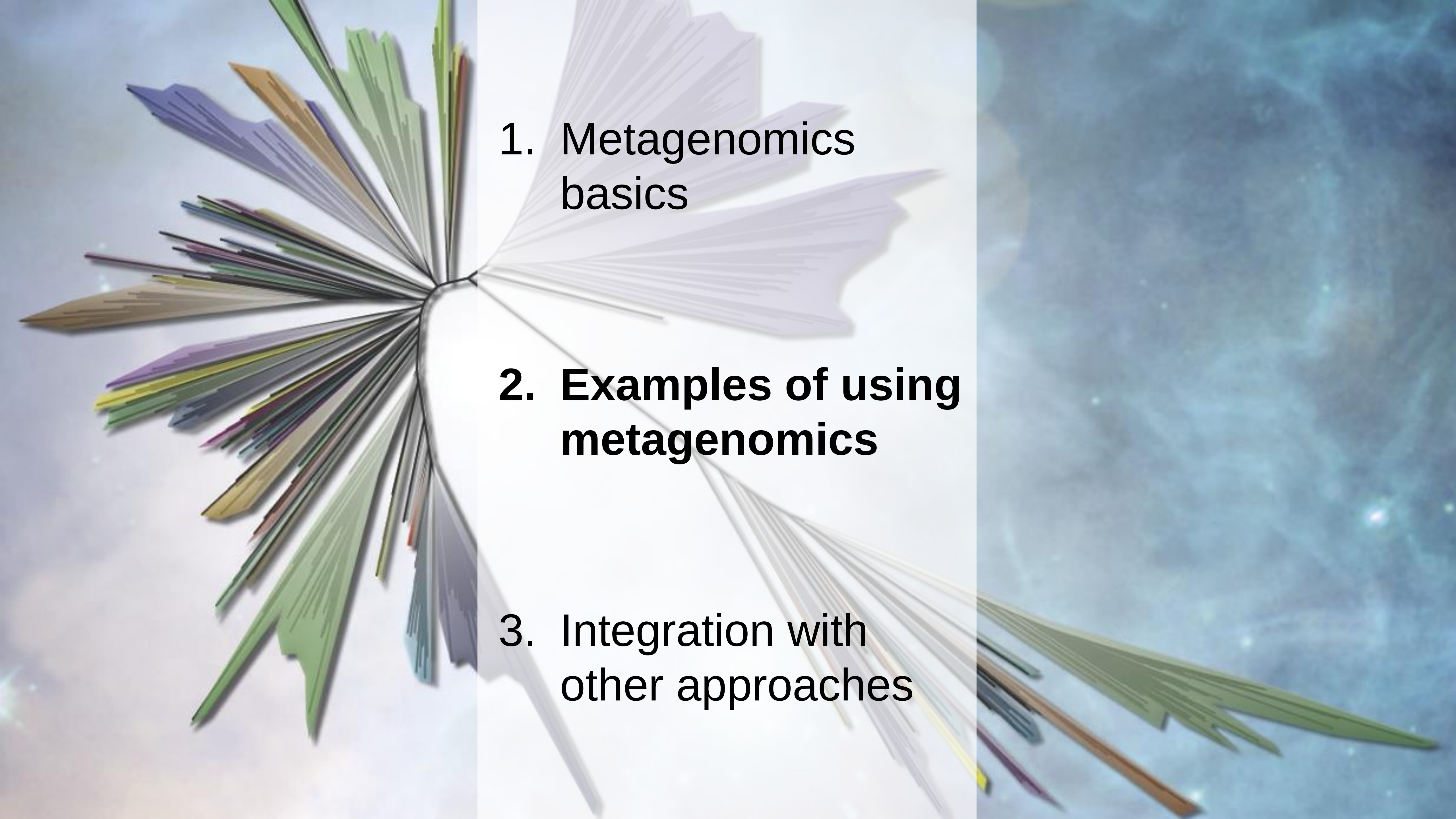
What are the functional capabilities of the community?

How might community members interact with the environment?

How do community members potentially interact with each other?

How does composition and function vary across time and space?

How can we use these findings to advance fundamental or scientific research?



1. Metagenomics basics
2. **Examples of using metagenomics**
3. Integration with other approaches

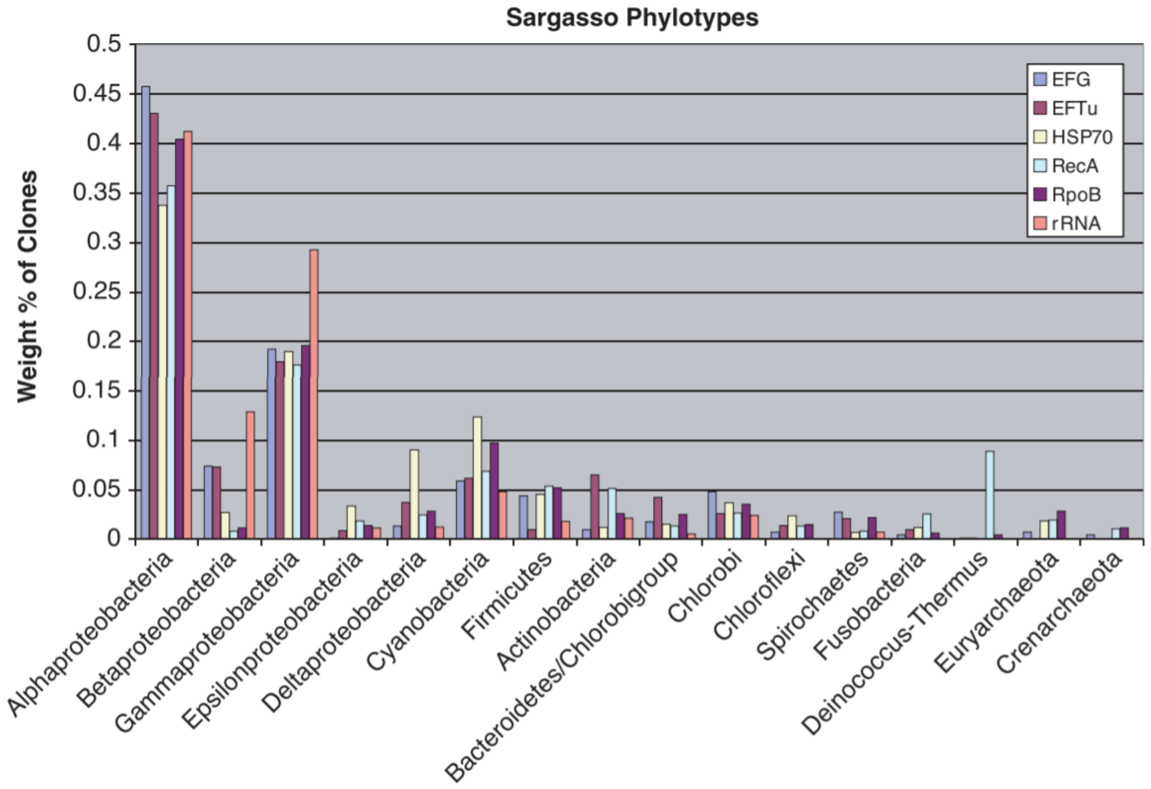
Landmark papers uncovered vast compositional and functional diversity

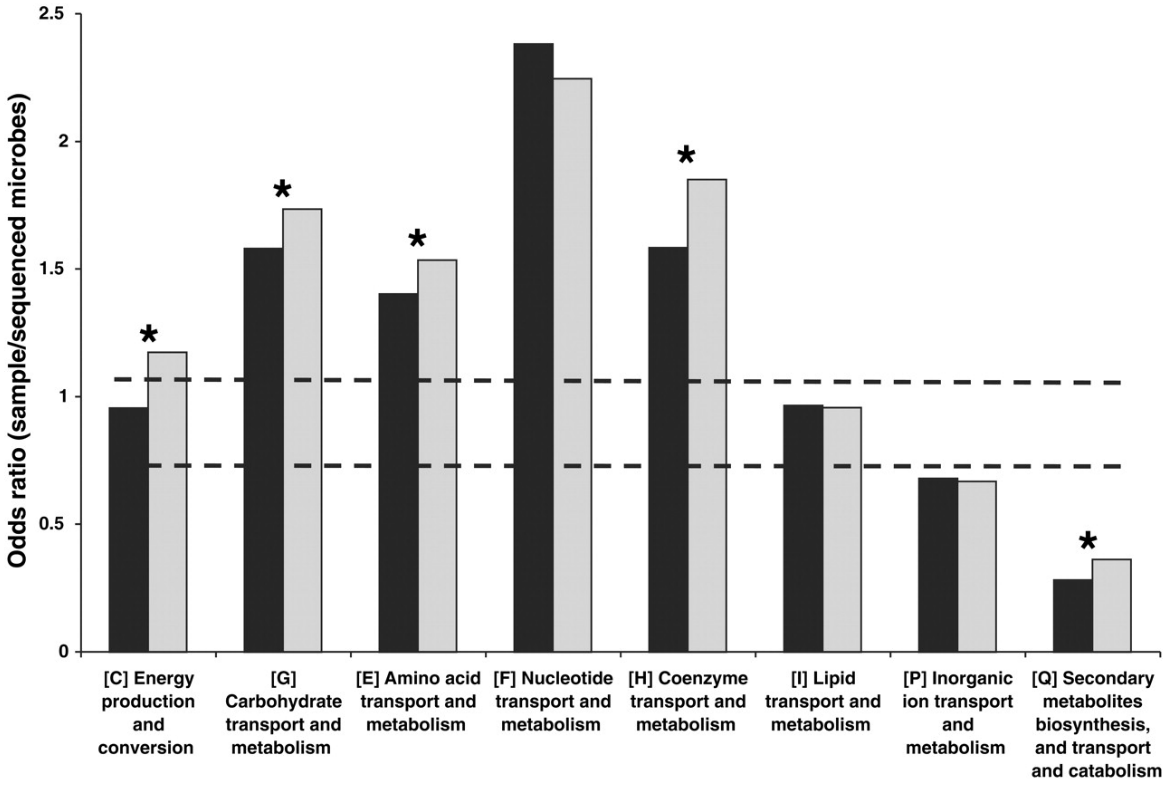
Venter et al., Science 2004

Metagenomic sequencing of the Sargasso sea detected 1800 phylotypes and 1.2M novel genes

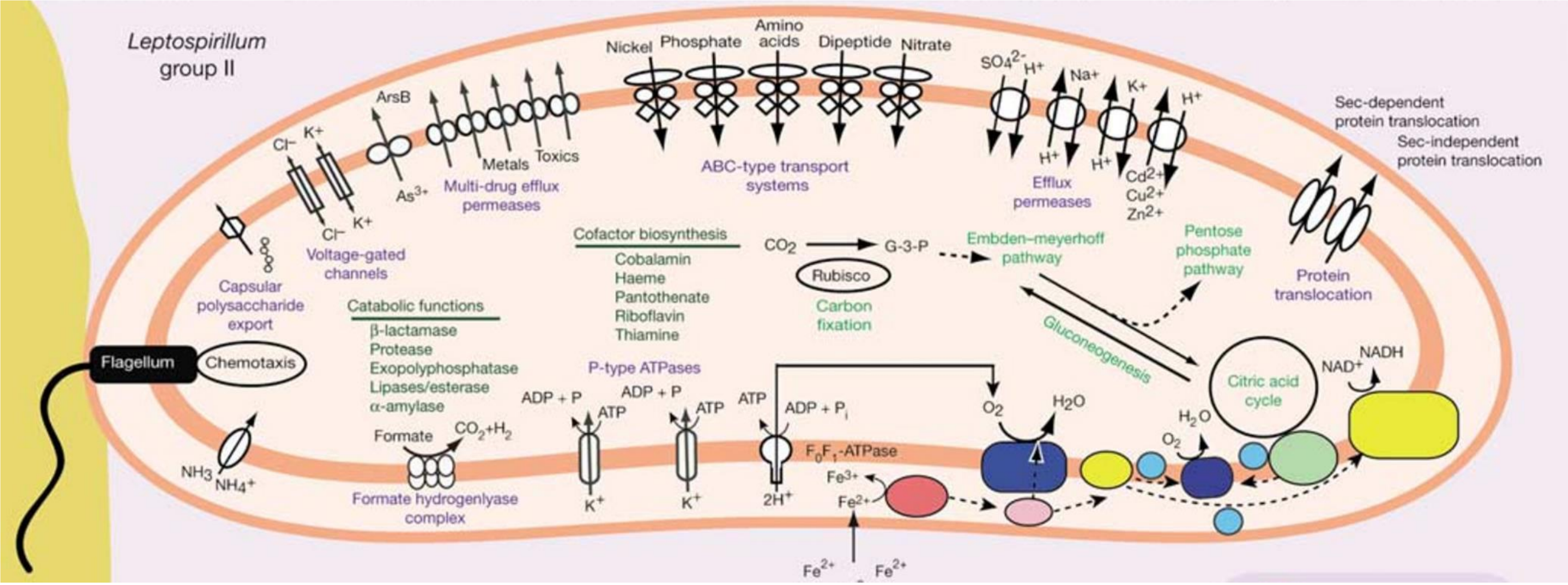
Gill et al., Science 2006

Metagenomic sequencing of the human gut revealed an enrichment of digestion genes



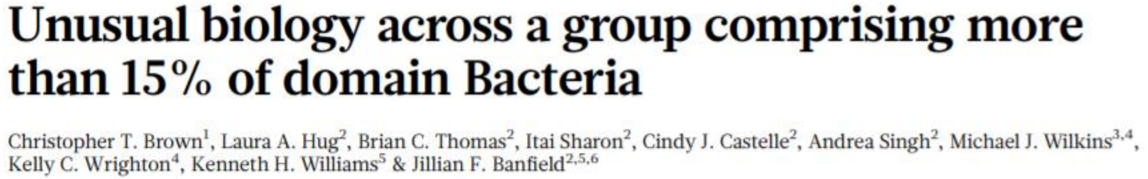


The first of many metagenome-assembled genomes



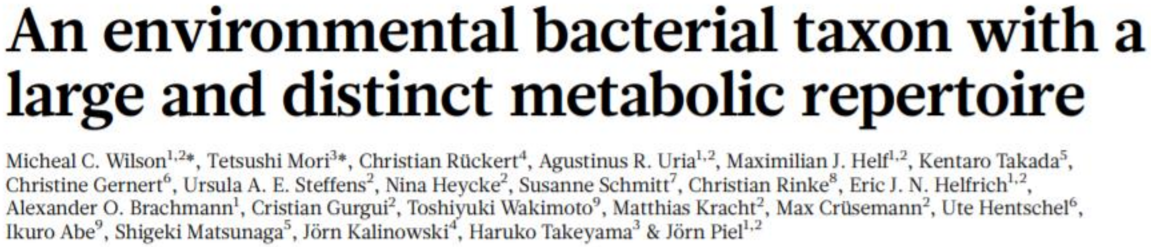
Tyson et al. (2004) Nature 428, 37-43..

Genome-resolved approaches have discovered numerous new phyla



Unusual biology across a group comprising more than 15% of domain Bacteria

Christopher T. Brown[1], Laura A. Hug[2], Brian C. Thomas[2], Itai Sharon[2], Cindy J. Castelle[2], Andrea Singh[2], Michael J. Wilkins[3,4], Kelly C. Wrighton[4], Kenneth H. Williams[5] & Jillian F. Banfield[2,5,6]

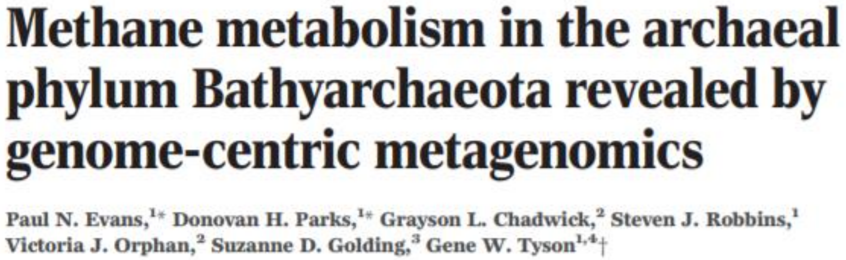


An environmental bacterial taxon with a large and distinct metabolic repertoire

Micheal C. Wilson[1,2]*, Tetsushi Mori[3]*, Christian Rückert[4], Agustinus R. Uria[1,2], Maximilian J. Helf[1,2], Kentaro Takada[5], Christine Gernert[6], Ursula A. E. Steffens[2], Nina Heycke[2], Susanne Schmitt[7], Christian Rinke[8], Eric J. N. Helfrich[1,2], Alexander O. Brachmann[1], Cristian Gurgui[2], Toshiyuki Wakimoto[9], Matthias Kracht[2], Max Crüsemann[2], Ute Hentschel[6], Ikuro Abe[9], Shigeki Matsunaga[5], Jörn Kalinowski[4], Haruko Takeyama[3] & Jörn Piel[1,2]

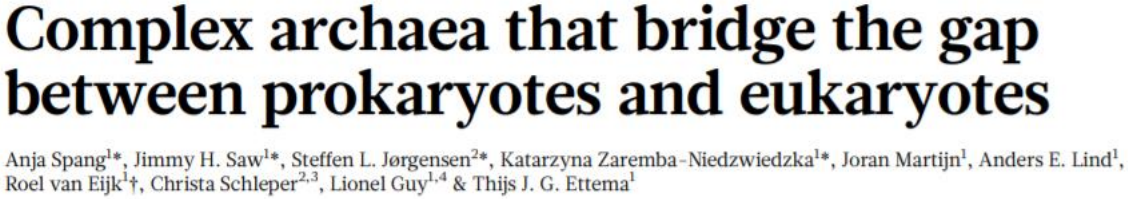
Atmospheric trace gases support primary production in Antarctic desert surface soil

Mukan Ji[1]*, Chris Greening[2]*, Inka Vanwonterghem[3], Carlo R. Carere[4], Sean K. Bay[2], Jason A. Steen[3], Kate Montgomery[1], Thomas Lines[2], John Beardall[2], Josie van Dorst[1], Ian Snape[5], Matthew B. Stott[4], Philip Hugenholtz[3] & Belinda C. Ferrari[1]



Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics

Paul N. Evans,[1]* Donovan H. Parks,[1]* Grayson L. Chadwick,[2] Steven J. Robbins,[1] Victoria J. Orphan,[2] Suzanne D. Golding,[3] Gene W. Tyson[1,4]†



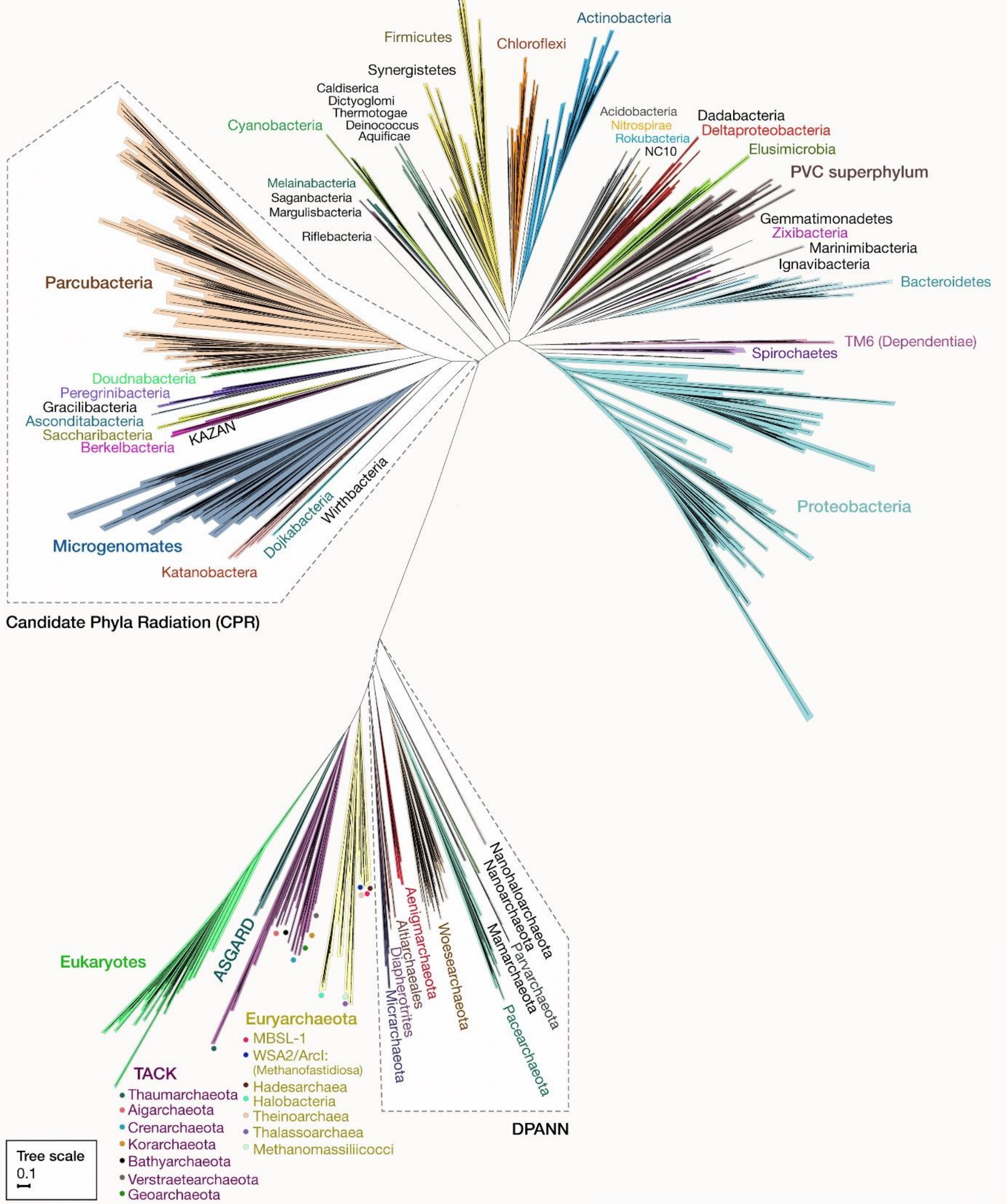
Complex archaea that bridge the gap between prokaryotes and eukaryotes

Anja Spang[1]*, Jimmy H. Saw[1]*, Steffen L. Jørgensen[2]*, Katarzyna Zaremba-Niedzwiedzka[1]*, Joran Martijn[1], Anders E. Lind[1], Roel van Eijk[1]†, Christa Schleper[2,3], Lionel Guy[1,4] & Thijs J. G. Ettema[1]

A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes

Bas E. Dutilh[1,2,3,4], Noriko Cassman[3,†], Katelyn McNair[2], Savannah E. Sanchez[3], Genivaldo G.Z. Silva[5], Lance Boling[3], Jeremy J. Barr[3], Daan R. Speth[6], Victor Seguritan[3], Ramy K. Aziz[2,7], Ben Felts[8], Elizabeth A. Dinsdale[3,5], John L. Mokili[3] & Robert A. Edwards[2,4,5,9]

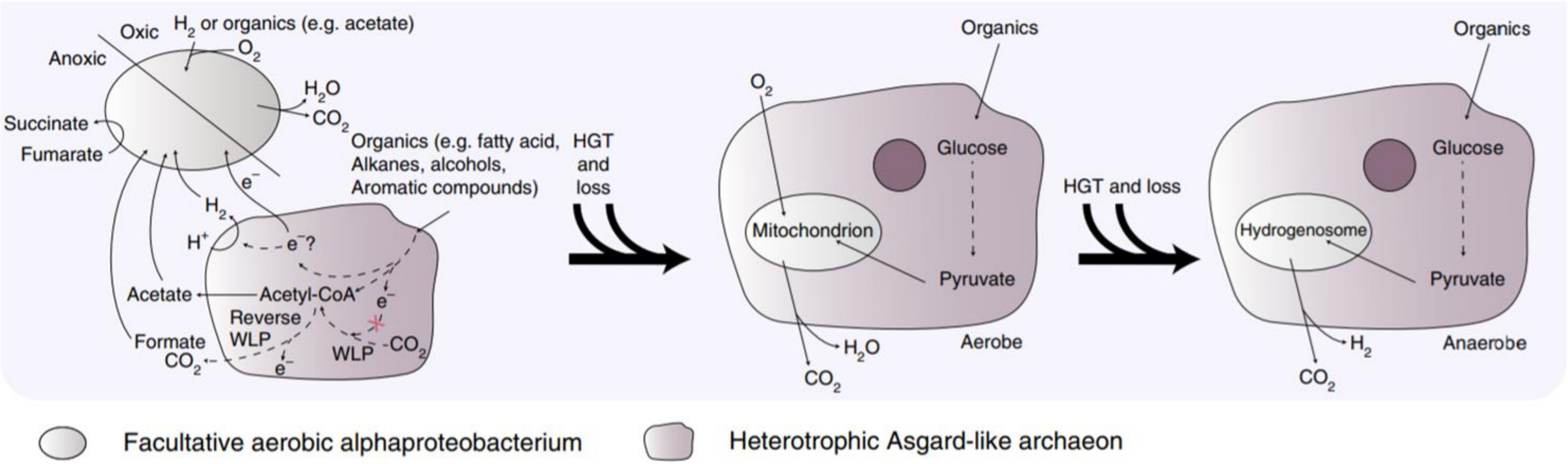
This has reshaped the tree of life…



Castelle and Banfield (2019) Cell 172, 1181-1197

…and led to new theories on abiogenesis and eukaryogenesis

A recent study reconstructed the metabolism of the Asgard archaea. We proposed that these archaea were fermenters and their alphaproteobacterial partners recycled the endproducts. The bacteria eventually became mitochondria.



Spang et al (2019), Nat Microbiol 4, 1138-1148

We’ve used metagenomics to understand how various ecosystems function

Deserts, forests, and grasslands

Greening et al, ISME 2016; Ji & Greening et al, Nature 2017; etc.

Human and animal tracts

Wolf et al, Gut Microbes 2016; Greening et al, ISME 2019, etc.

Subsurface environments

Dong et al, ISME 2018 & Nat Comms 2019; Spang et al, Nat Microbiol 2019

Marine waters and sediments

Kessleret al, Nat Microbiol 2019; Baltar et al., in preparation, etc.

The built environment

Taruc et al., Lancet Planetary Health 2019; Lappan et al., in prep

Global aerosphere

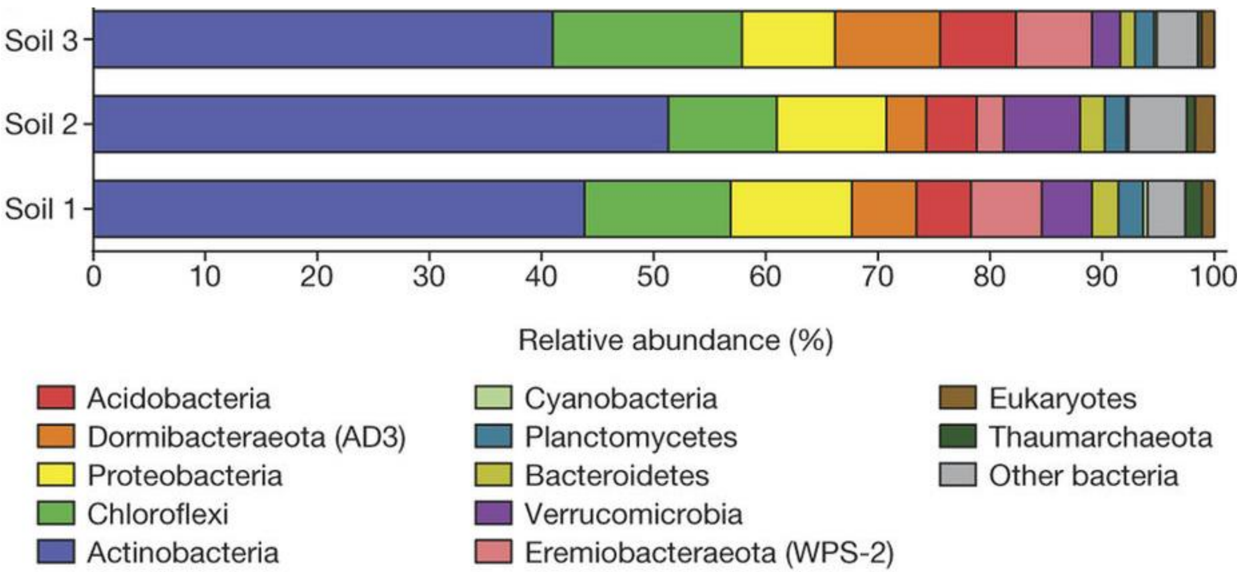
Archer & Lappan et al., in prep

Some of these studies have changed the way we think about life

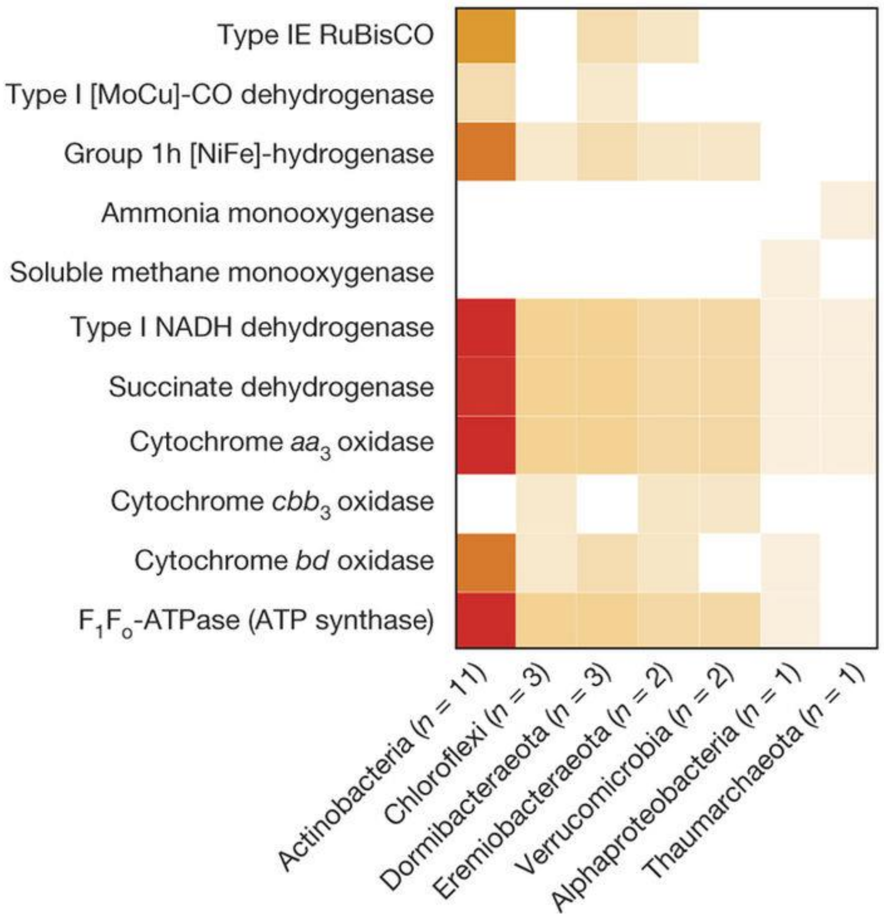
It’s thought that most soils are sustained by sunlight. However, we provided strong evidence that atmospheric energy sources support the diverse, abundant communities in terrestrial Antarctica, namely hydrogen and carbon monoxide.

Genome-resolved metagenomics that was later validated

These soils are phylogenetically diverse but contain few phototrophs



Genome reconstructions show Antarctic soils dominated by gas scavengers



Ji & Greening et al. (2017), Nature 552, 400-403

A boom in using metagenomics to understand health and disease

Metagenomics is being for diverse purposes: to profile the gastrointestinal microbiome, diagnose infectious disease, or track spread of antimicrobial resistance.

rise

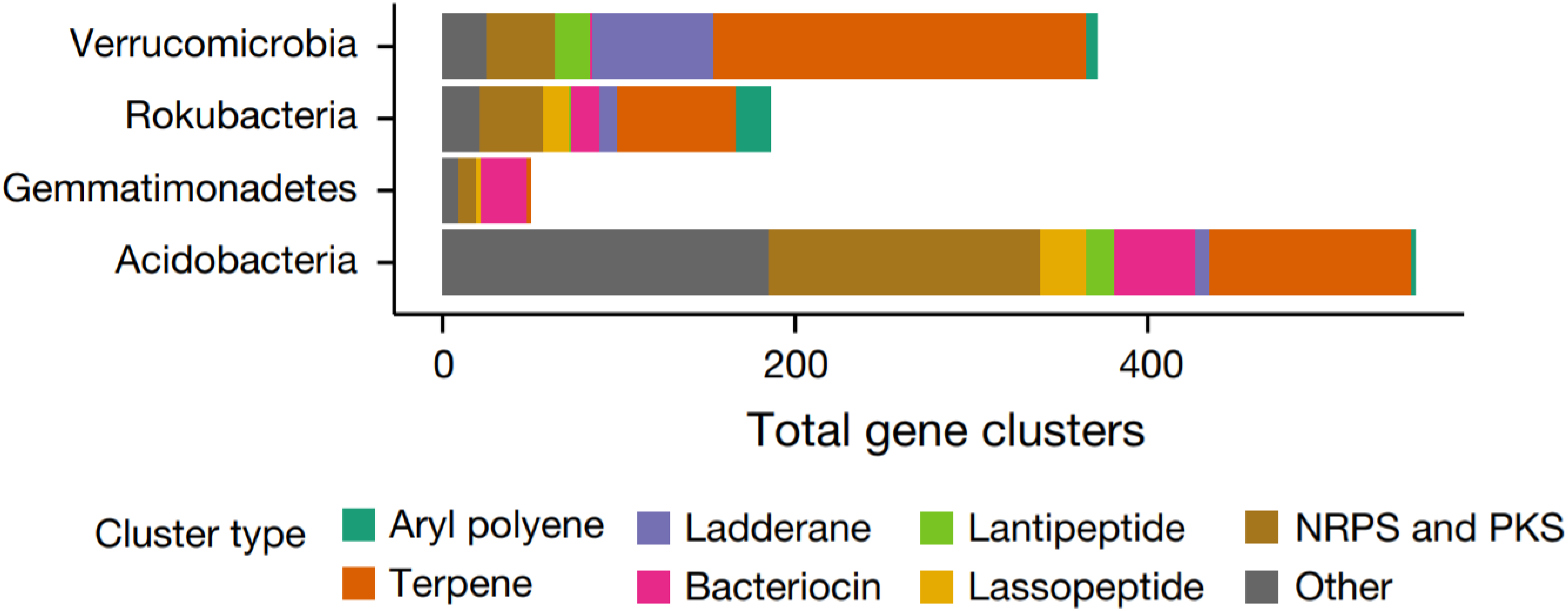
REVITALISING INFORMAL
SETTLEMENTS AND
THEIR ENVIRONMENTS

We are using metagenomics to profile antimicrobial resistance genes. We are looking at environmental samples, animal scats, and human stool before and after a public health intervention.

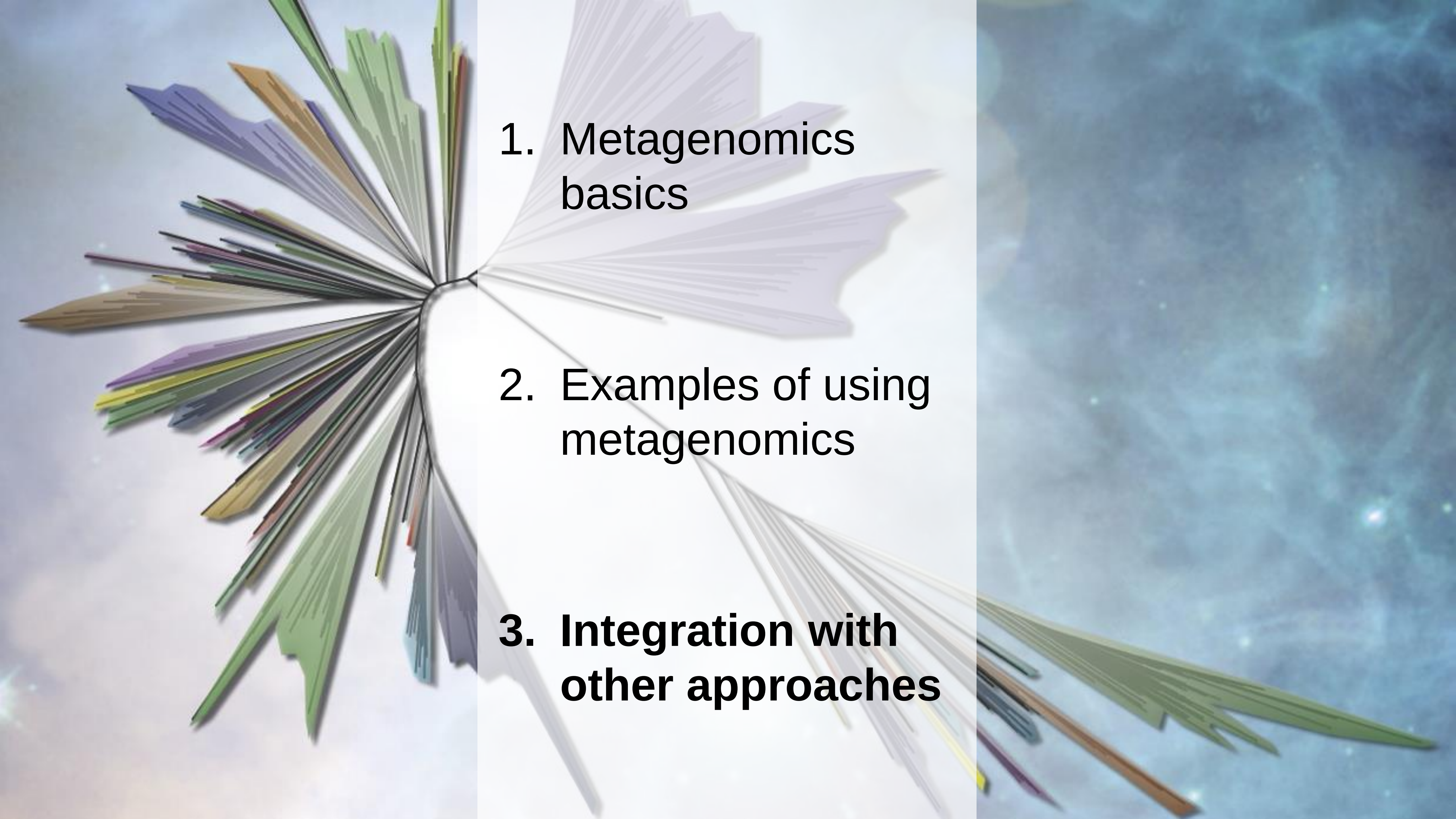
Taruc et al. (2019), Lancet Planetary Health 3, S8

And provides a novel means for drug discovery

For example, metagenomics has been used to discovery gene clusters for novel antibiotics across soil phyla such as the Acidobacteria and Verrucomicrobia.



Crits-Christoph et al (2018), Nature 558, 440-444



1. Metagenomics basics
2. Examples of using metagenomics
3. **Integration with other approaches**

Metagenomes are a resource to generate testable hypotheses

> “Metagenomics can only provide inferences about how these organisms live, rather than any sort of proofs. Moreover, these approaches have limited capacity to discover or characterize genes encoding novel functions.”

Greening et al. (2019), mSystems 4, e00107-19

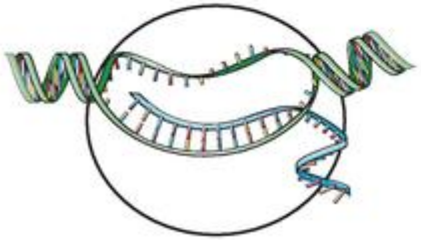
Going more "meta" can help to develop insights

Metagenomics

- DNA sequencing and assembly.
- What is there?
- What can they do?

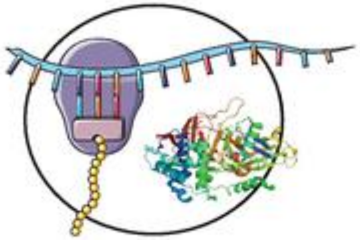
Metaproteomics

- Protein separation and mass spectrometry.
- What is translated?
- Are complexes formed?



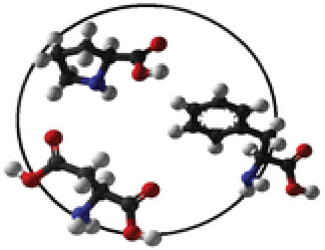
Metatranscriptomics

- RNA reverse transcription and sequencing.
- What is active?
- How do they respond?



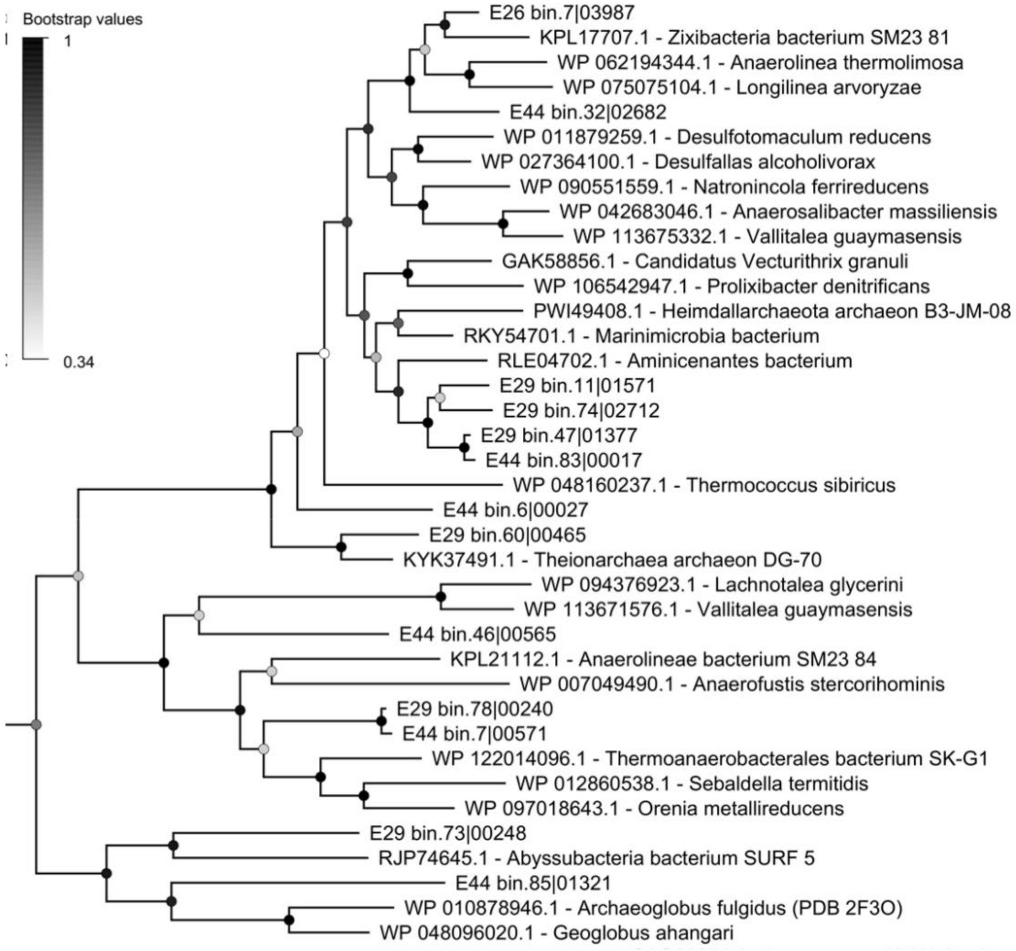
Metabolomics

- Metabolite separation and mass spectrometry.
- What pathways are active?
- How is environment modified?

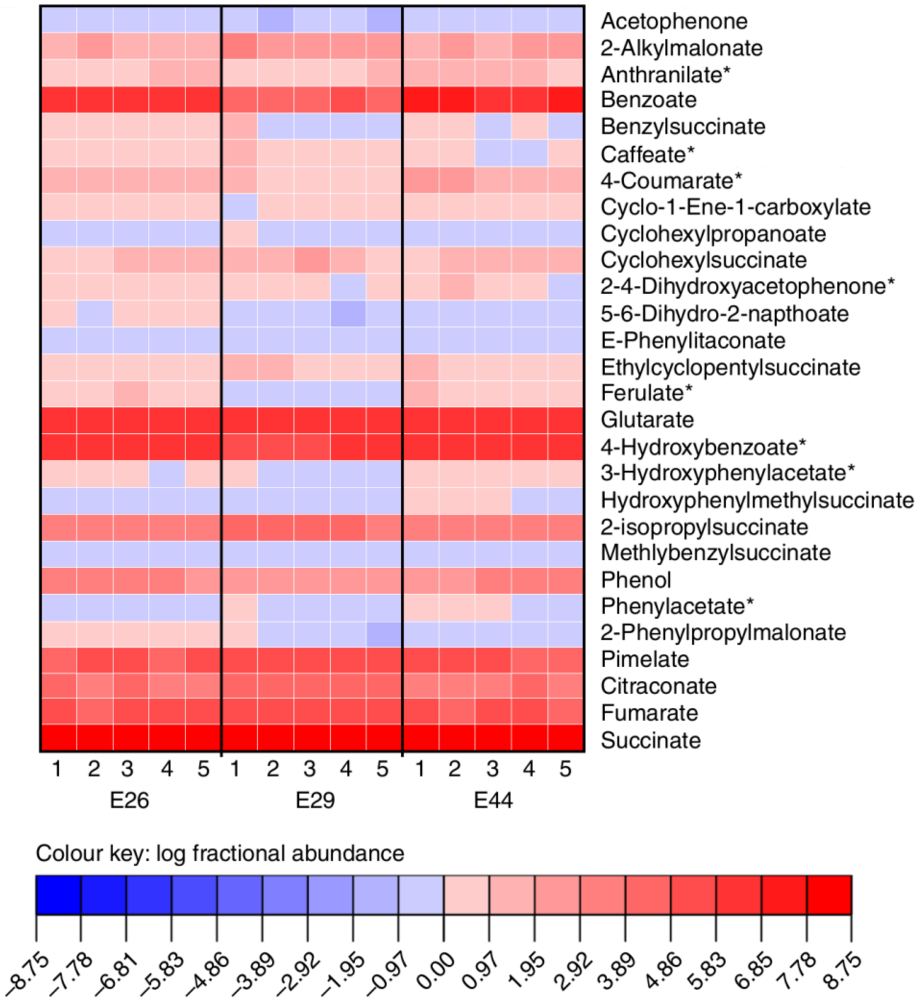


Metabolomics confirms hydrocarbon activation pathway in petroleum seeps

Many MAGs encode novel enzymes that may activate hydrocarbons by fumarate addition



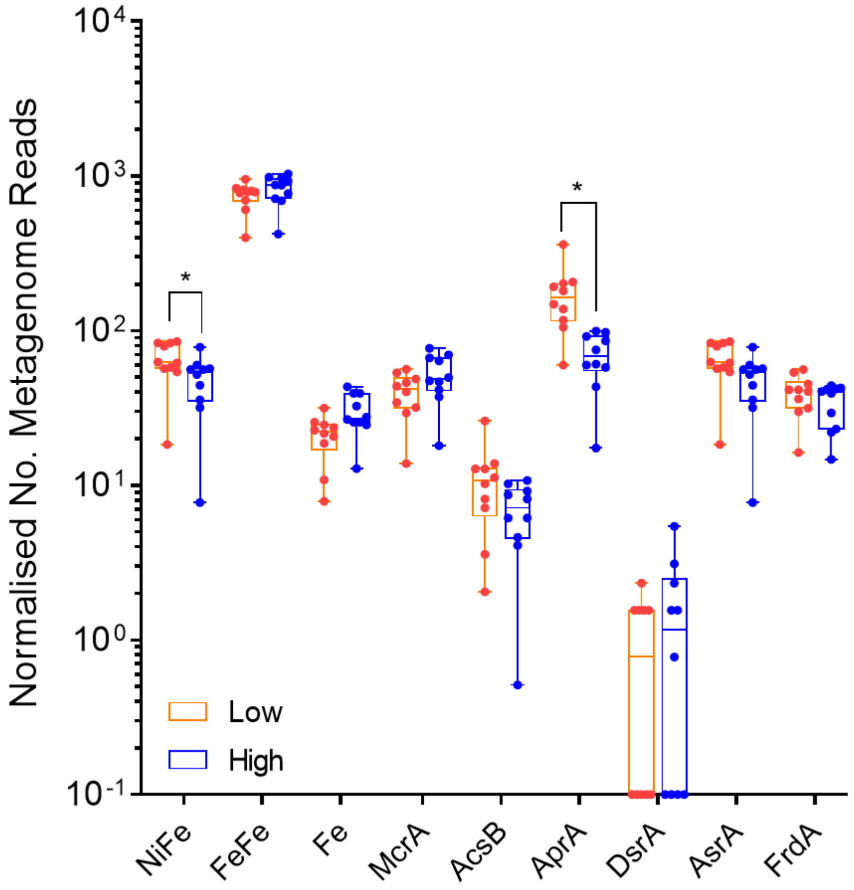
Metabolomics confirmed that succinate-hydrocarbon conjugates are highly abundant



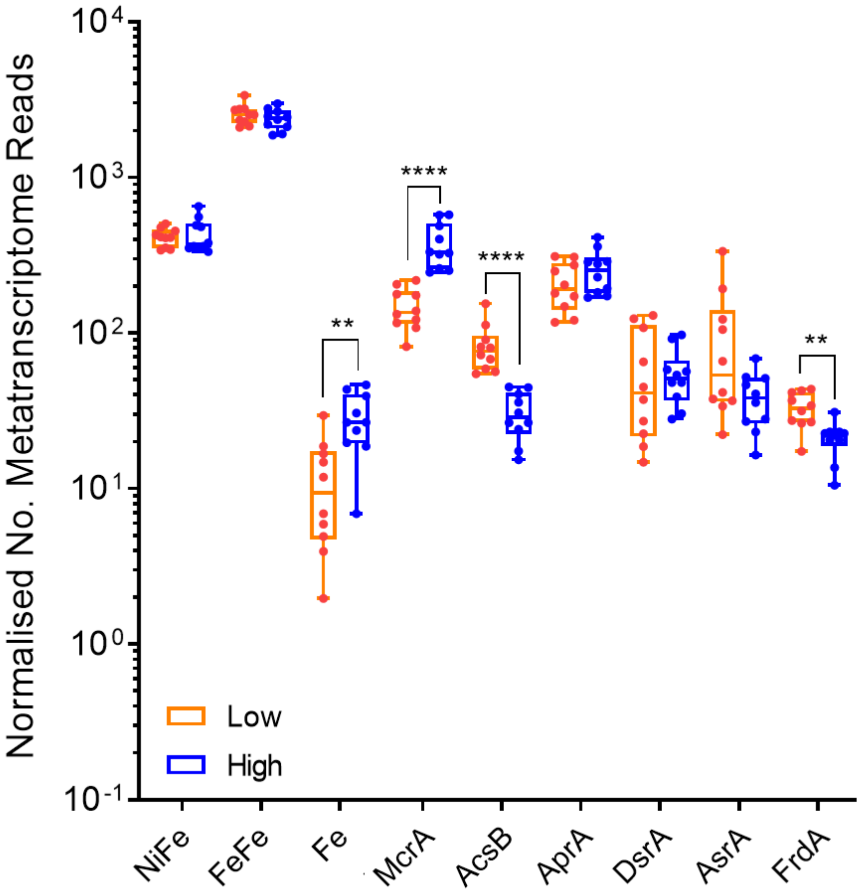
Dong et al (2019), Nat Commun 10, 1816

Metatranscriptomics explains methane yield differences in farmed sheep

Few differences in community composition or gene abundance between low and high methane yield sheep



Methanogenesis downregulated in low yield sheep in favour of acetogenesis and fumarate reduction



Greening et al. (2019), ISME J 13, 2617-2632

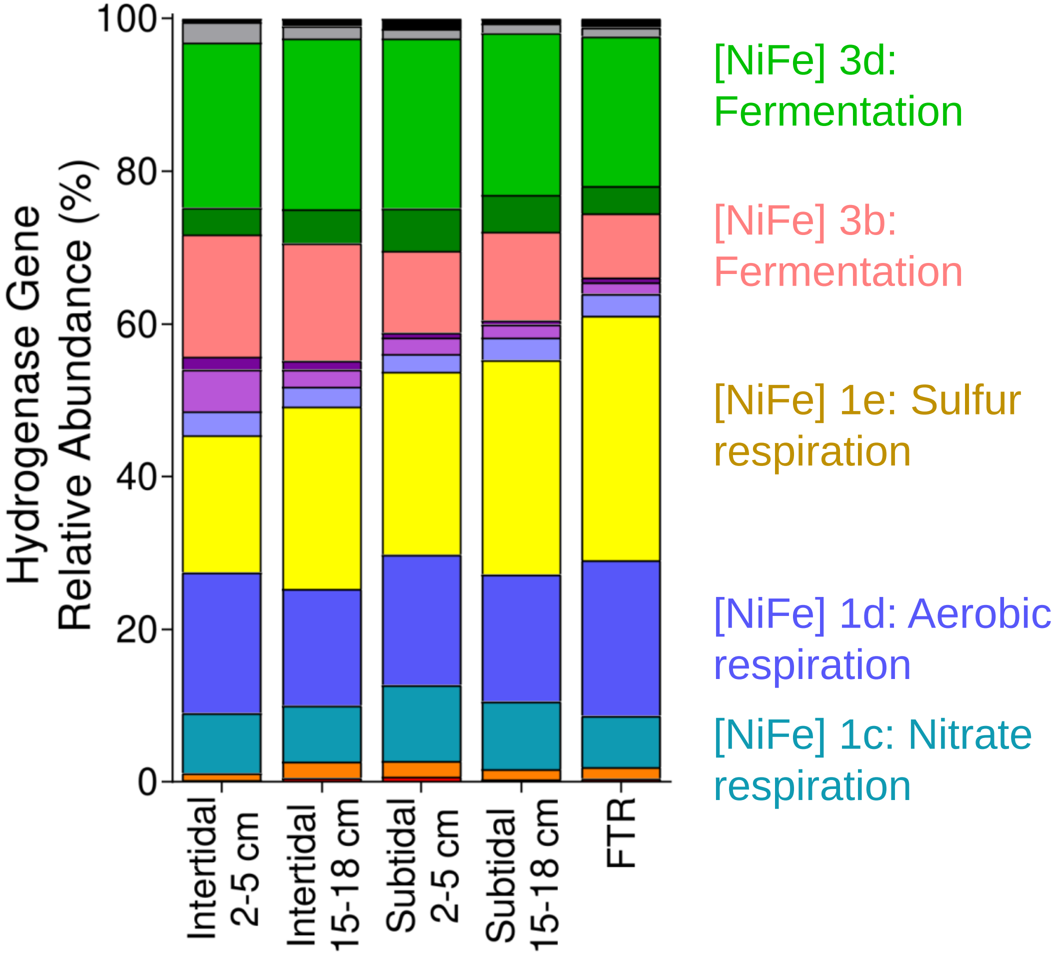
But more targeted approaches are also needed to validate findings

> “We must integrate meta-omics technologies with more targeted approaches. Only through such combinations will it be possible to gain both a broad and a deep understanding of how the uncultured majority live.”

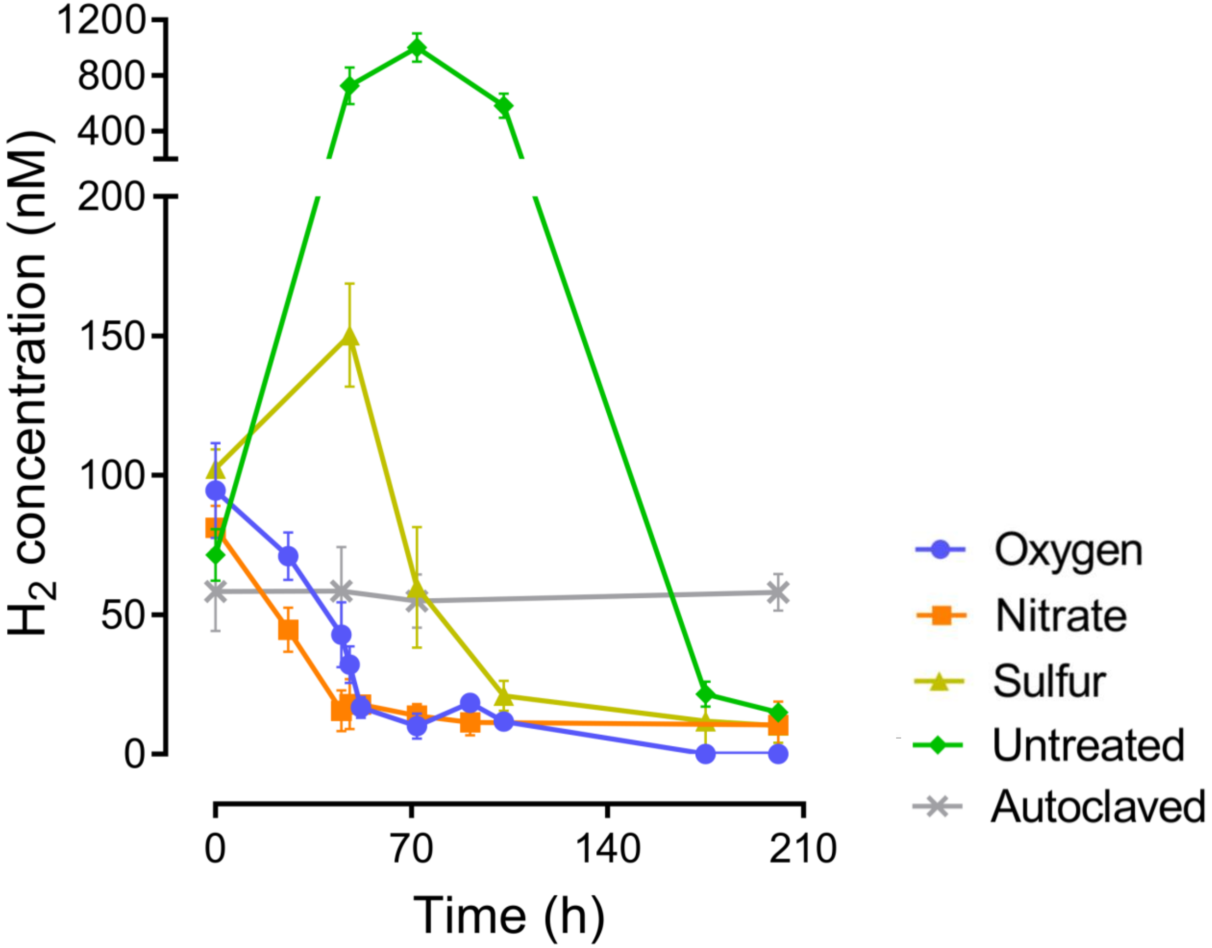
Greening et al. (2019), mSystems 4, e00107-19

Linking microbial communities to biogeochemical function in beach sands

Metagenomic profiling suggests diverse capabilities for hydrogen metabolism



Biogeochemical perturbation studies perfectly validates metagenomic predictions



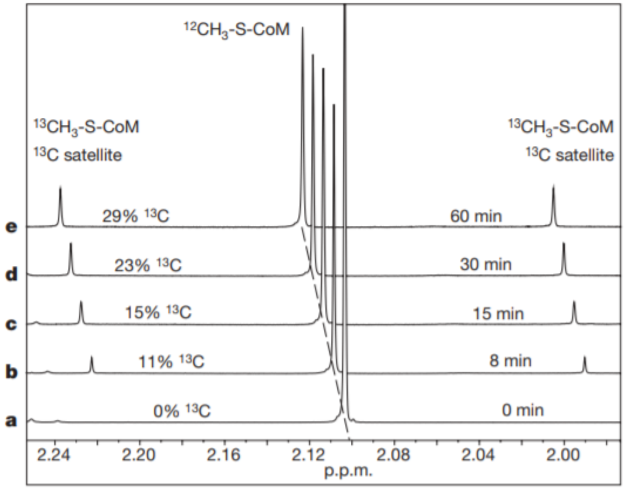
Kessler & Chen et al. (2019), Nature Microbiol 4, 1014-1023

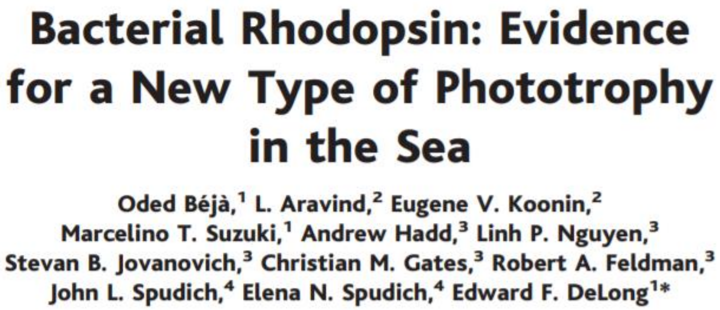
Classical genetics and biochemistry can validate novel gene functions

LETTERS

The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane

Silvan Scheller[1], Meike Goenrich[2], Reinhard Boecher[2], Rudolf K. Thauer[2] & Bernhard Jaun[1]





Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea

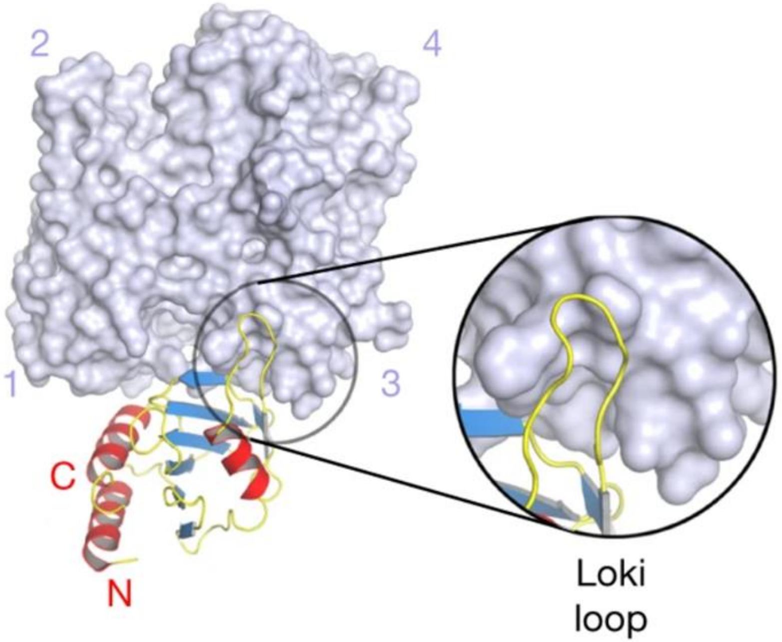
Oded Béjà,[1] L. Aravind,[2] Eugene V. Koonin,[2]
Marcelino T. Suzuki,[1] Andrew Hadd,[3] Linh P. Nguyen,[3]
Stevan B. Jovanovich,[3] Christian M. Gates,[3] Robert A. Feldman,[3]
John L. Spudich,[4] Elena N. Spudich,[4] Edward F. DeLong[1]*

LETTER

https://doi.org/10.1038/s41586-018-0548-6

Genomes of Asgard archaea encode profilins that regulate actin

Caner Akıl[1,2] & Robert C. Robinson[1,2,3]*

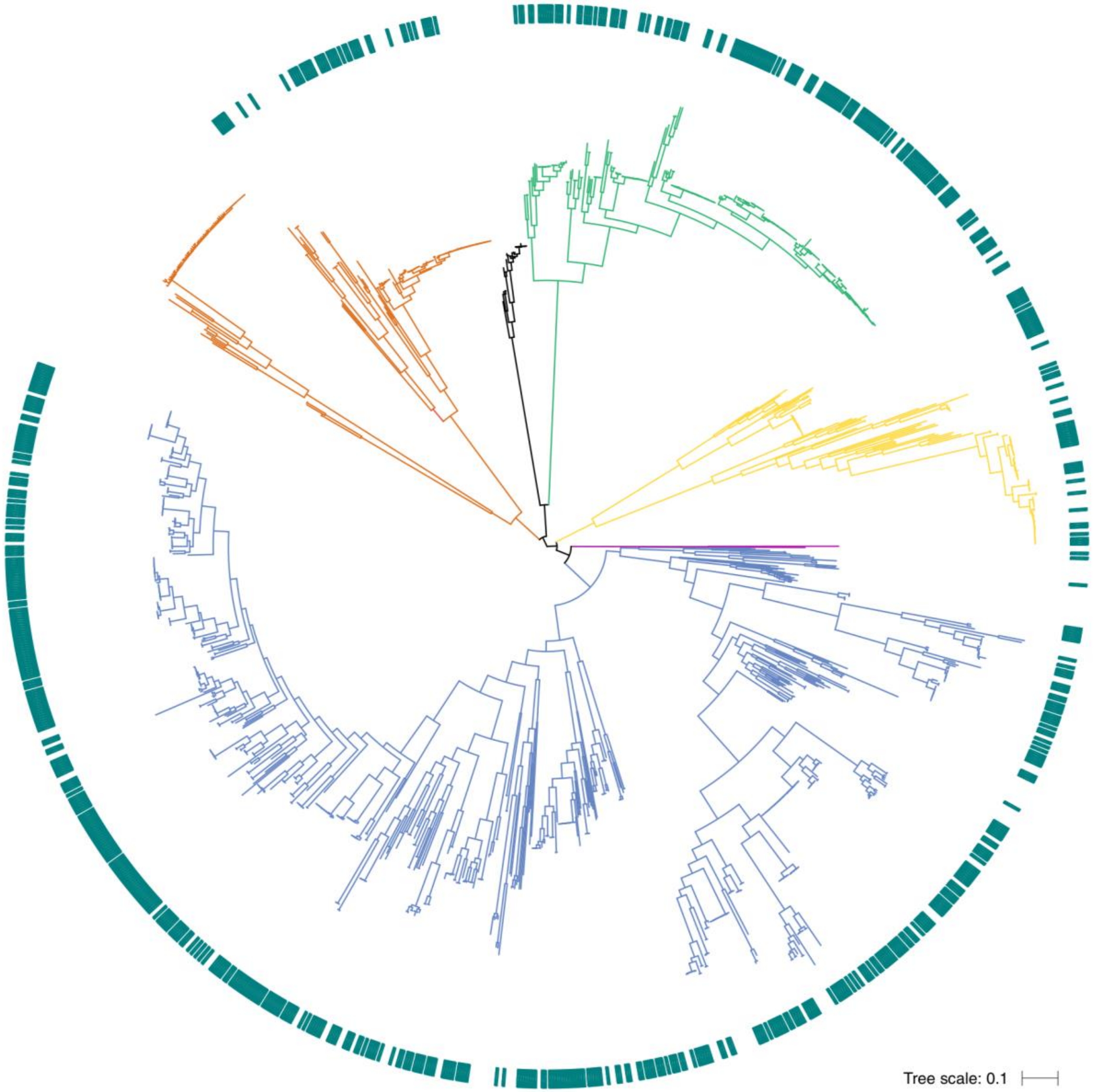


Going back to culture to bridge these divides

There has been a recent resurgence in culturing. Both targeted and untargeted approaches have been developed.

For example, researchers have performed high-throughput cultivation and sequencing of thousands of human gut bacterial isolates.

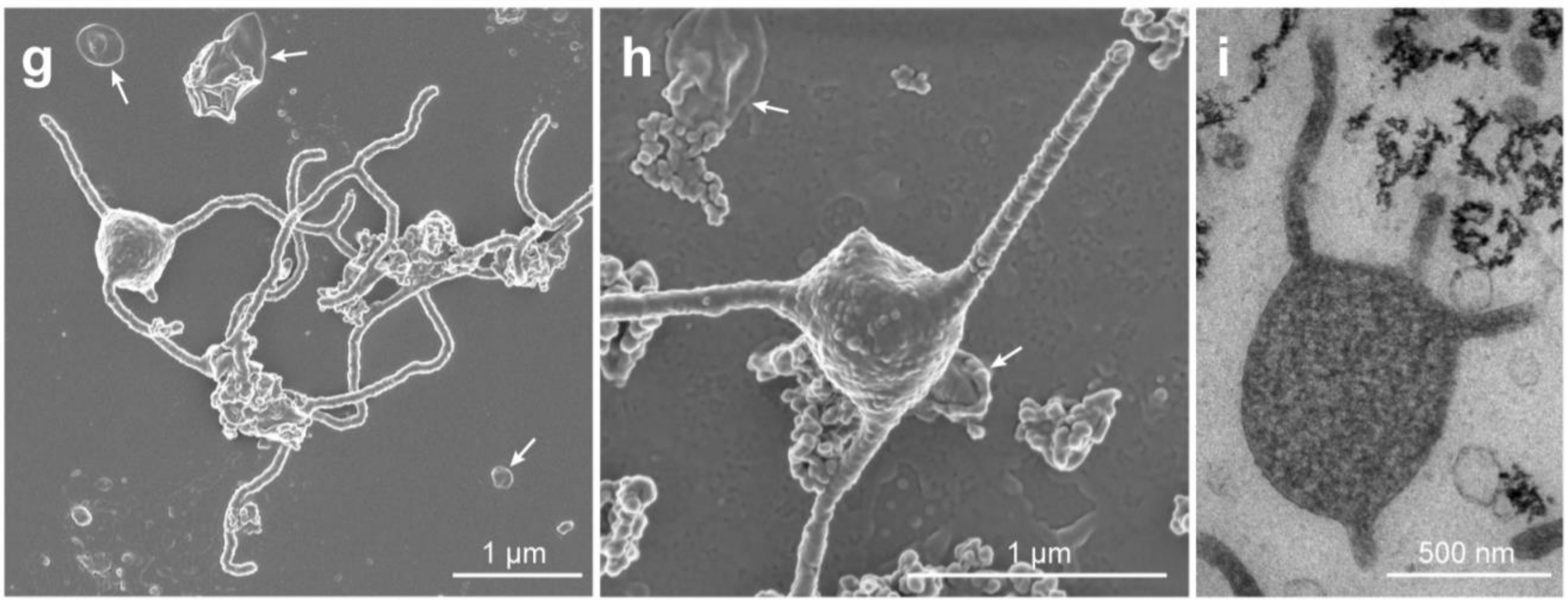
Researchers have now cultured organisms discovered by metagenomics, e.g. Lokiarchaeota, Atribacteria, Latesibacteria.



Forster et al. (2019), Nat Biotechnol 37, 186-192

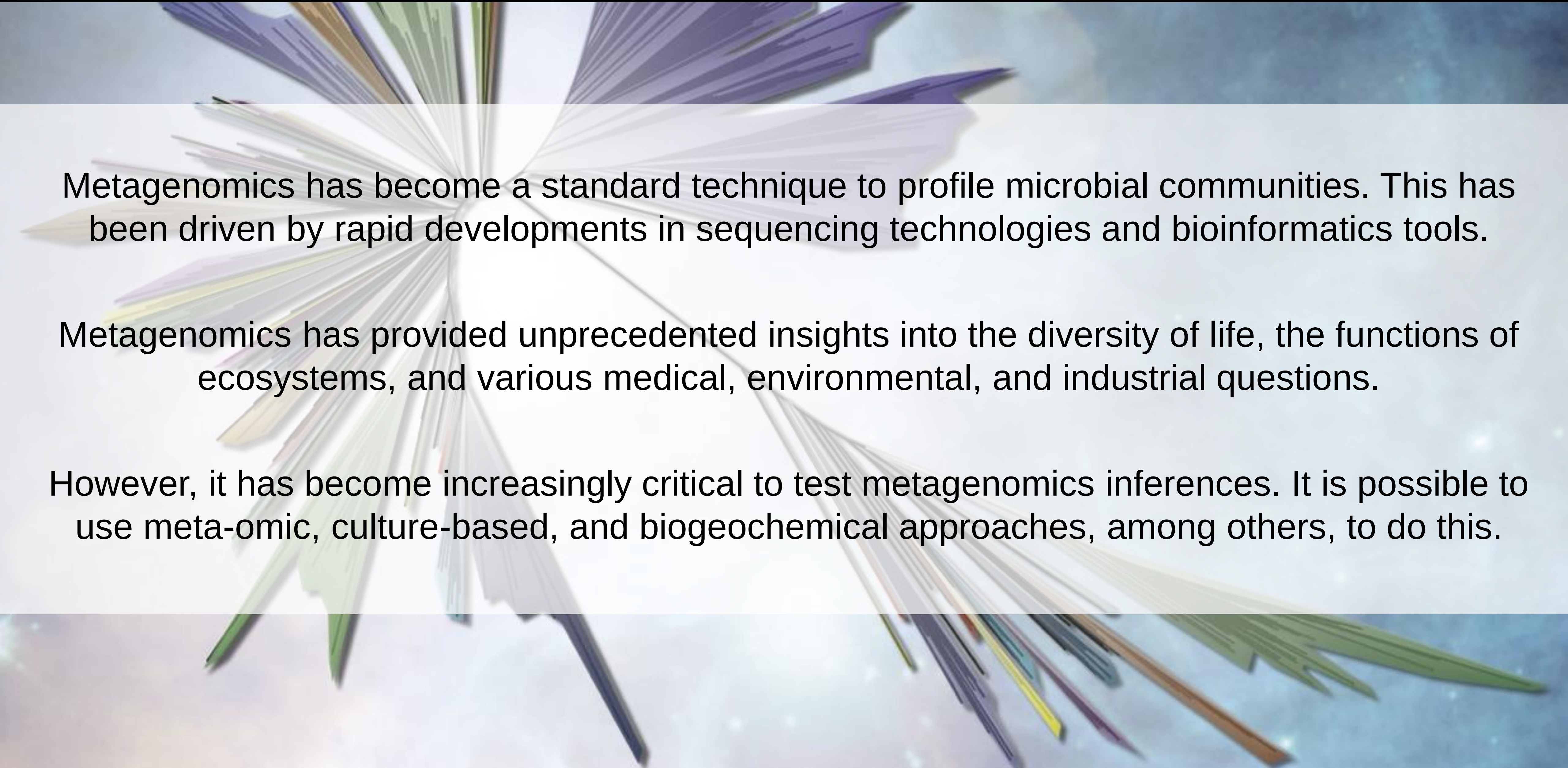
Cultivation is often higher-risk but higher-reward

A spectacular example of this is the successful seven-year isolation of a Lokiarchaeum. This confirmed metagenomics-based inferences about its phylogeny, morphology, and metabolism, while providing many new discoveries.



Imachi & Nobu et al. (2019), bioRxiv, 10.1101/726976

Take-home messages



Metagenomics has become a standard technique to profile microbial communities. This has been driven by rapid developments in sequencing technologies and bioinformatics tools.

Metagenomics has provided unprecedented insights into the diversity of life, the functions of ecosystems, and various medical, environmental, and industrial questions.

However, it has become increasingly critical to test metagenomics inferences. It is possible to use meta-omic, culture-based, and biogeochemical approaches, among others, to do this.

Any questions? Thanks for listening!

greeninglab.com chris.greening@monash.edu twitter.com/greeninglab

Australian Government
Australian Research Council

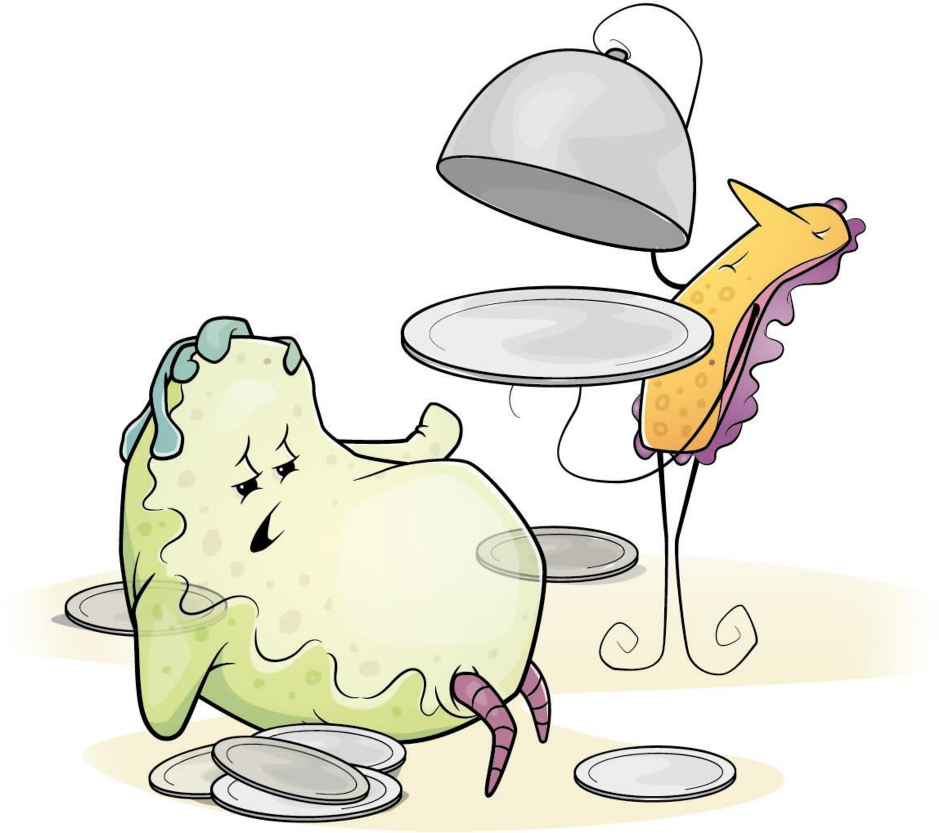
Australian Government
Department of the Environment
Australian Antarctic Division

Australian Government
National Health and Medical Research Council

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

wellcome

FNSNF



No more air for me, thanks