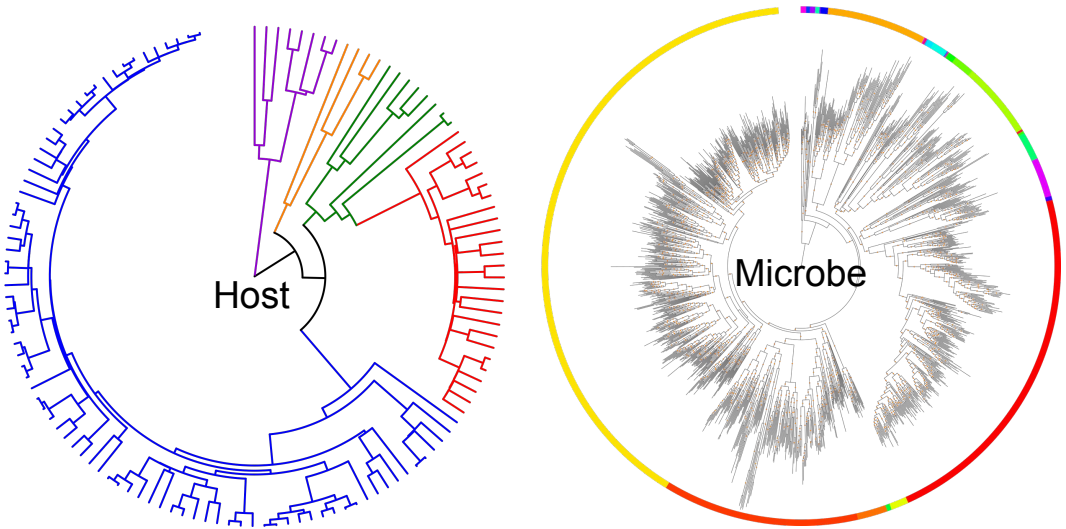
Bioinformatics in environmental and host-associated microbiome research

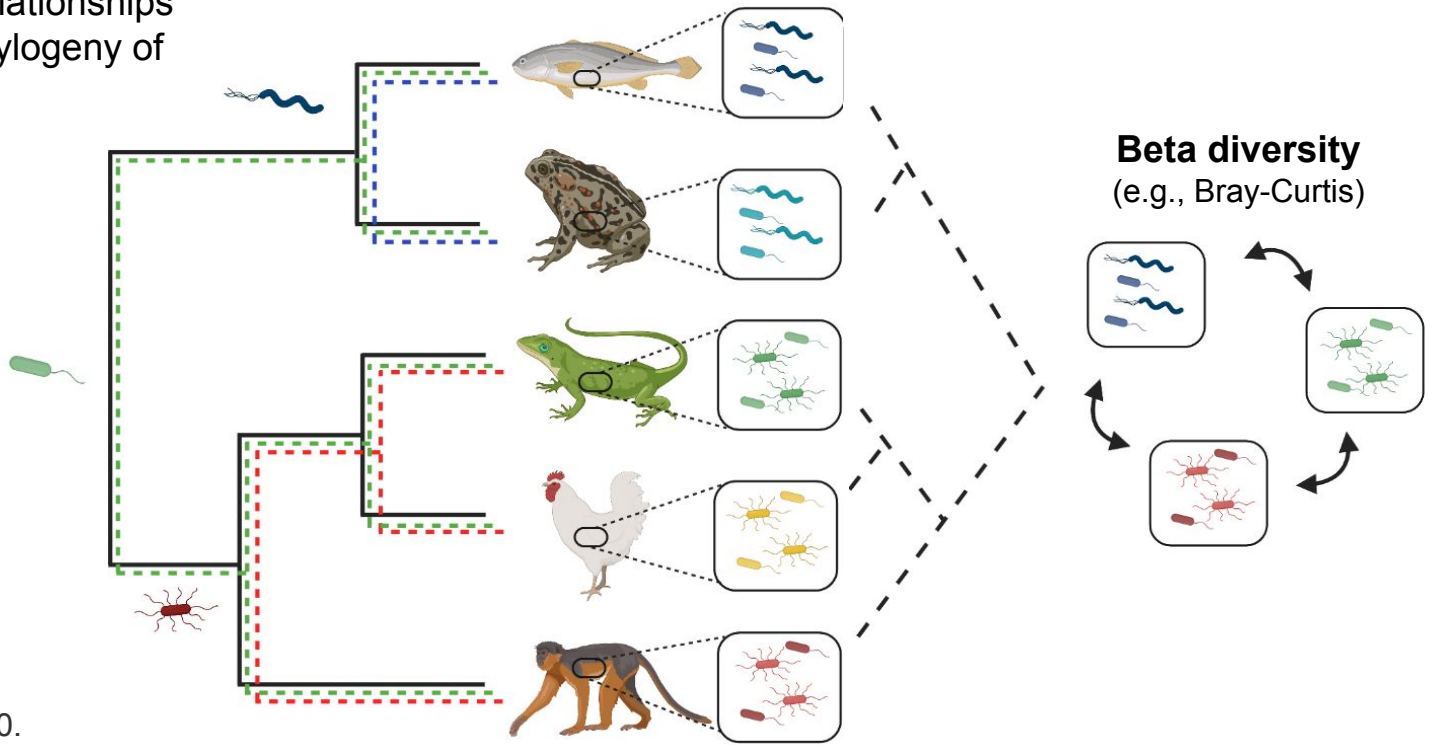


Nick Youngblut
Group Leader
Department of Microbiome Science
Max Planck Institute for Developmental Biology

Ley Lab

Microbial community assembly in the animal gut

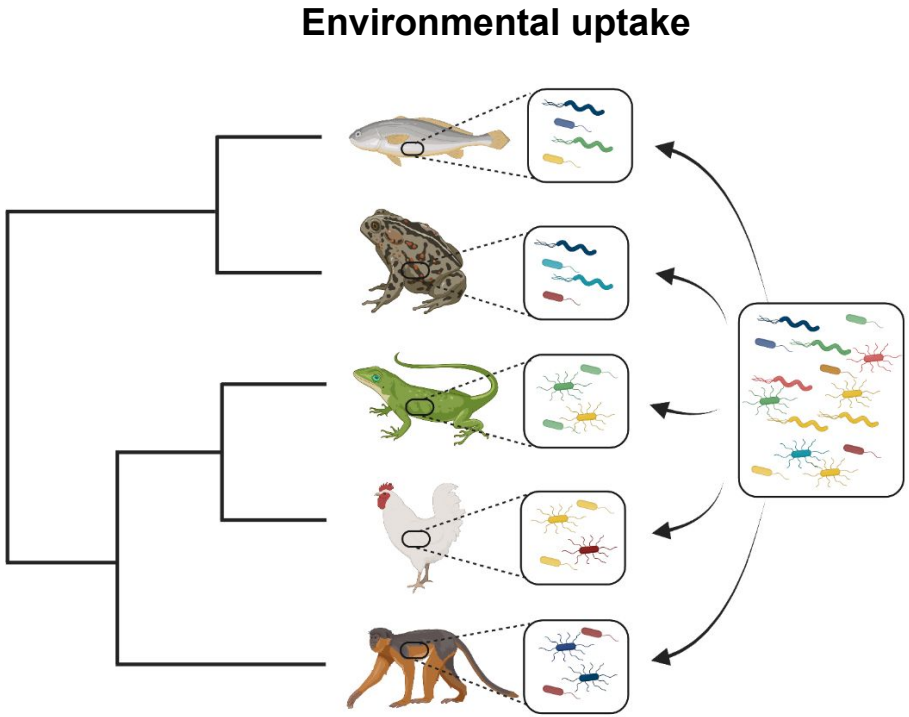
Phylosymbiosis:
microbial community relationships that recapitulate the phylogeny of their host

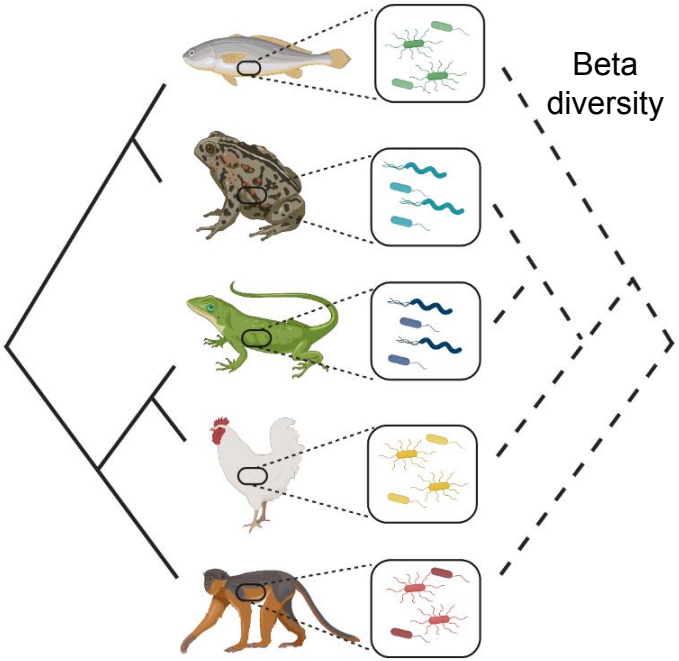


Lim and Bordenstein 2020.
Royal Society B

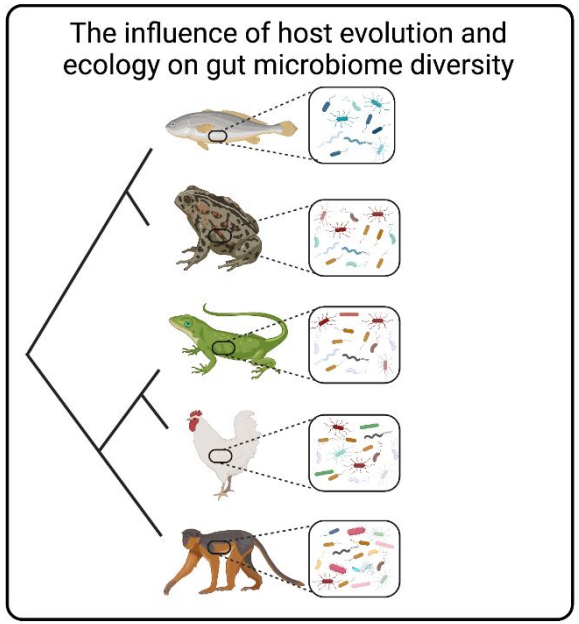
Created with BioRender

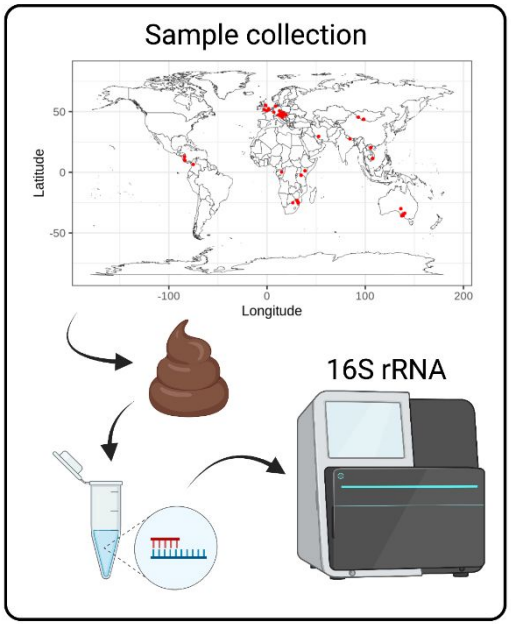
Microbial community assembly in the animal gut

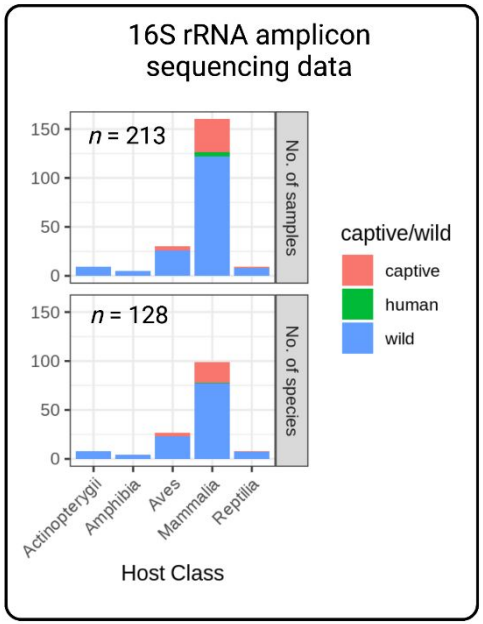




The vertebrate gut microbiome: influence of host evolution & ecology







Georg Reischer

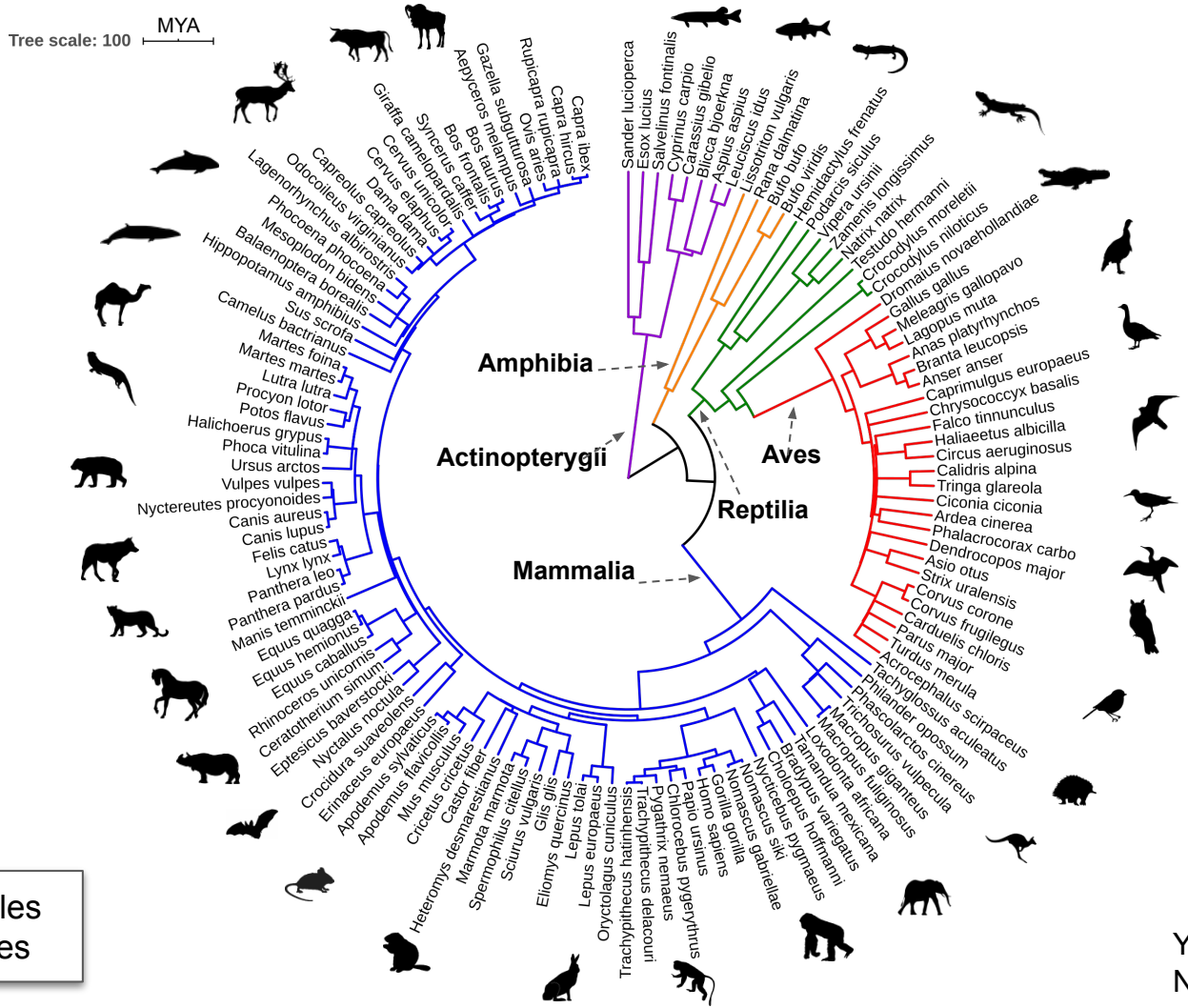
Andreas Farnleitner

Gabriella Stalder

Chris Walzer

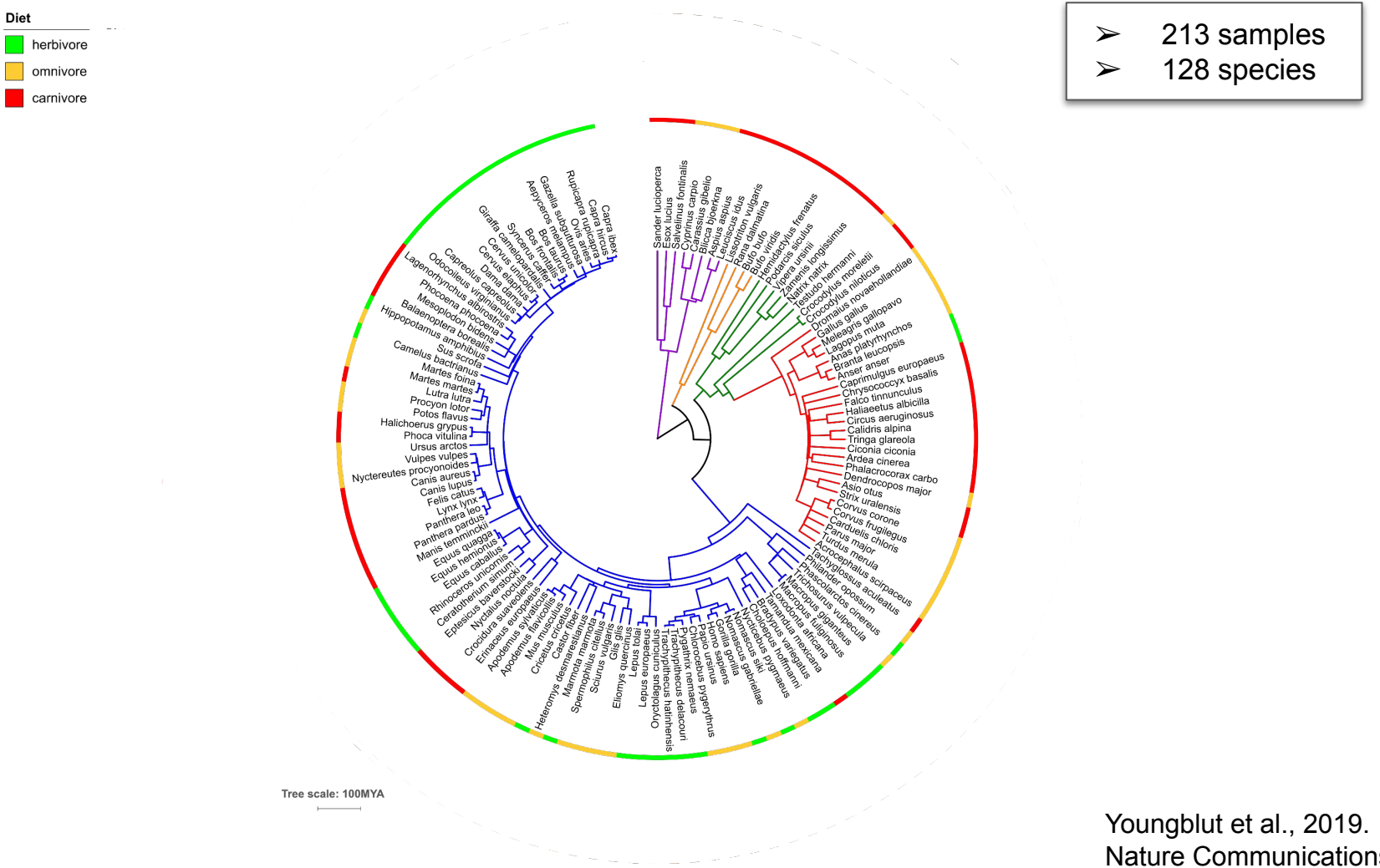
Youngblut et al., 2019. Nature Communications

Created with BioRender



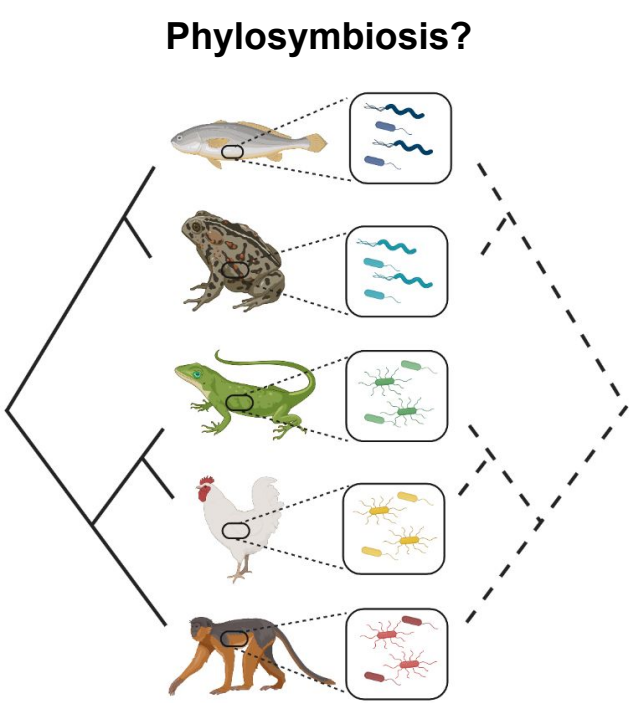
- 213 samples
- 128 species

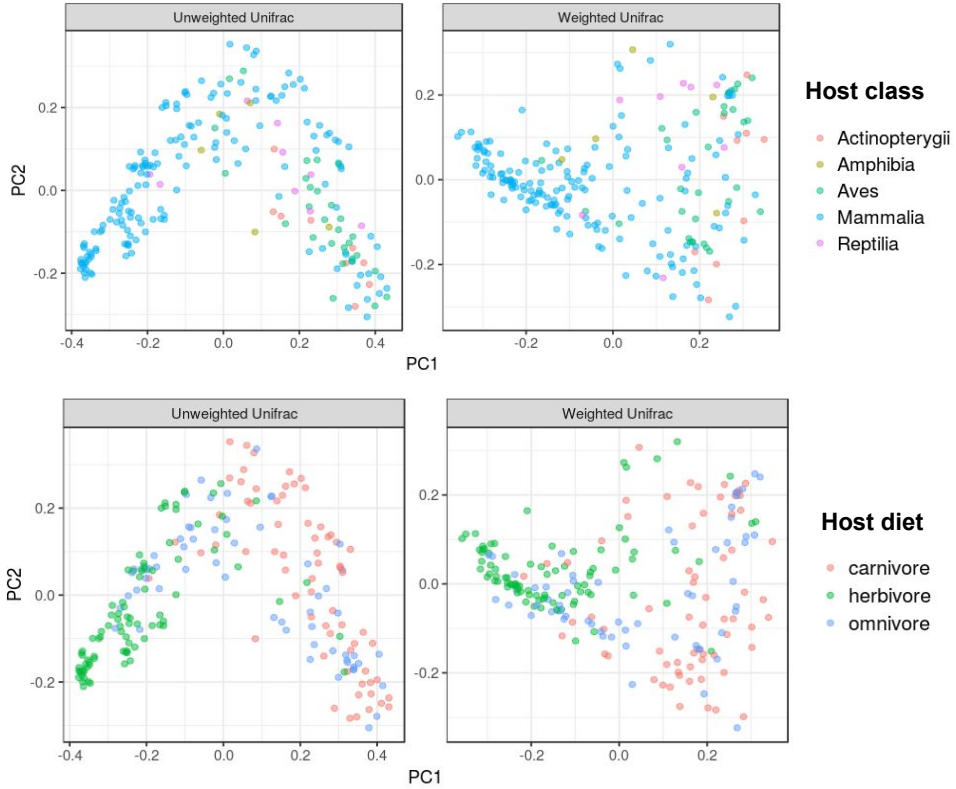
Youngblut et al., 2019.
Nature Communications



Youngblut et al., 2019.
Nature Communications

Pattern of phylosymbiosis?





Created with BioRender

Youngblut et al., 2019. Nature Communications

Phylosymbiosis versus other factors

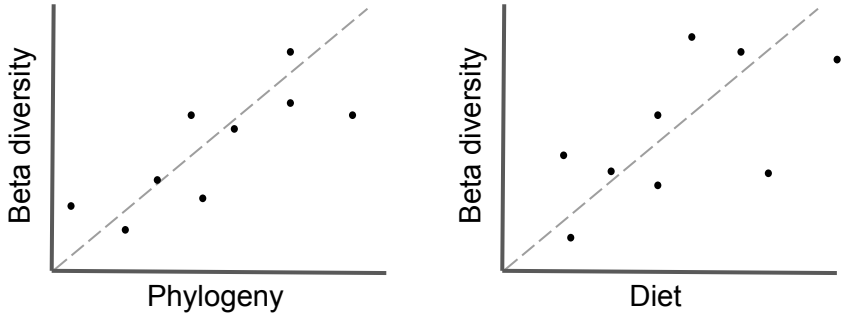
Multiple regression on matrices (MRM)

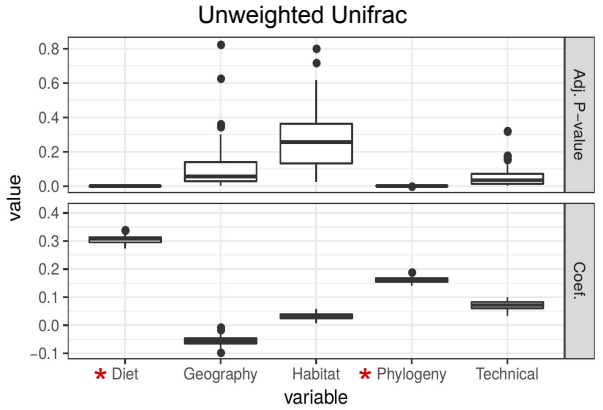
Phylogeny
(patristic distance)

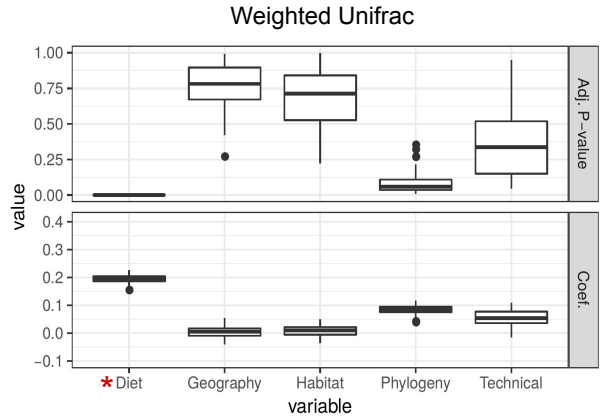
	S1	S2	S3
S1	0	...	...
S2	0.1	0	...
S3	0.3	0.2	0

Diet
(Gower's distance)

	S1	S2	S3
S1	0	...	...
S2	0.3	0	...
S3	0.6	0.1	0



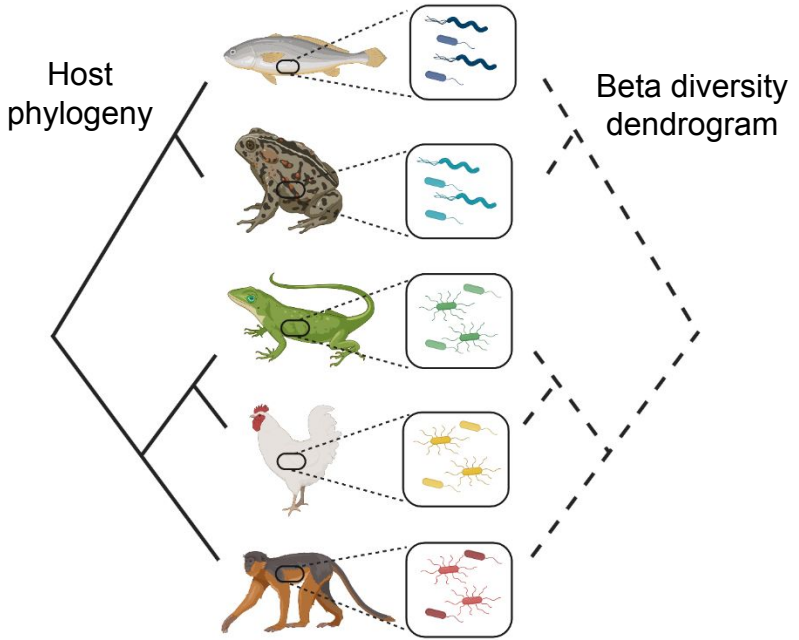




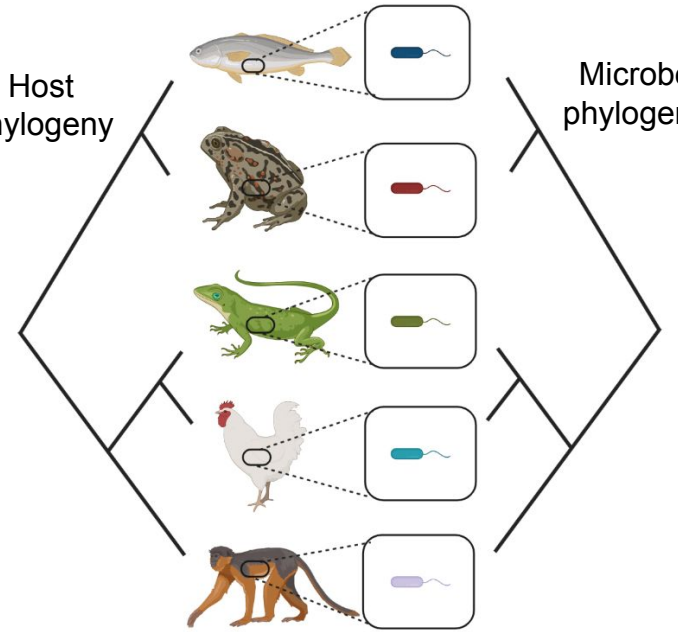
Youngblut et al., 2019. Nature Communications

Phylosymbiosis versus cophylogeny

Phylobiosis: community similarity



Cophylogeny: congruent evolutionary histories

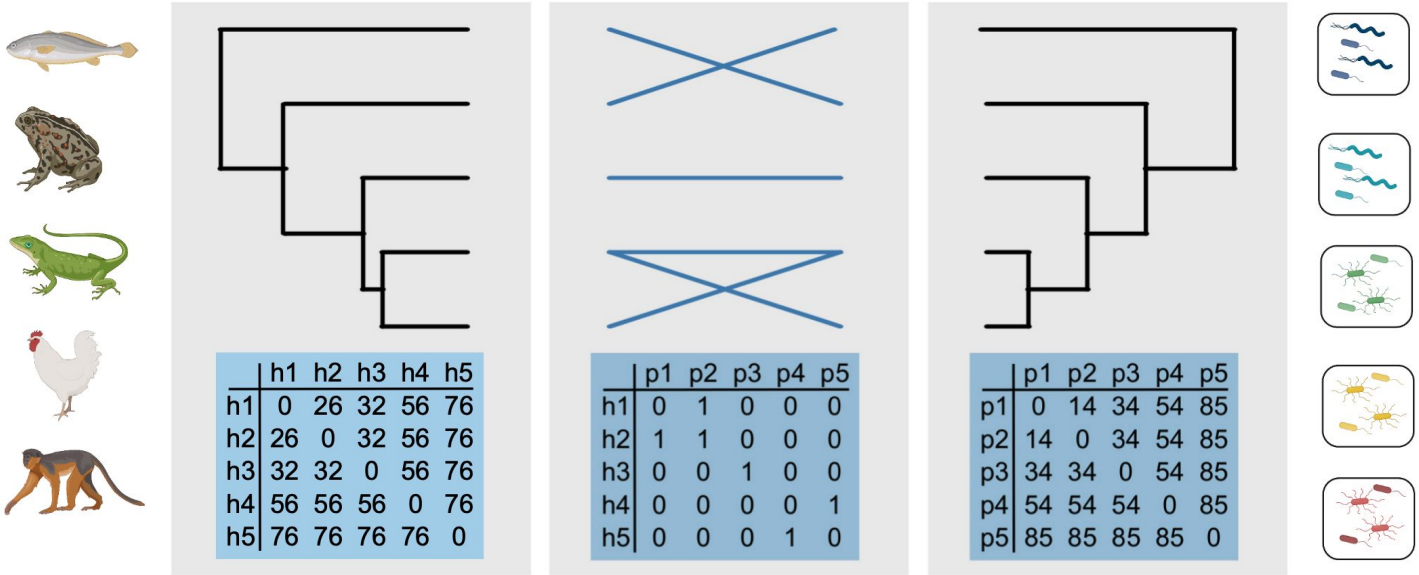


Created with
BioRender

Phylosymbiosis versus cophylogeny

Cophylogeny for the entire microbial community

Procrustean Approach to Cophylogeny (PACo)



	h1	h2	h3	h4	h5
h1	0	26	32	56	76
h2	26	0	32	56	76
h3	32	32	0	56	76
h4	56	56	56	0	76
h5	76	76	76	76	0

	p1	p2	p3	p4	p5
h1	0	1	0	0	0
h2	1	1	0	0	0
h3	0	0	1	0	0
h4	0	0	0	0	1
h5	0	0	0	1	0

	p1	p2	p3	p4	p5
p1	0	14	34	54	85
p2	14	0	34	54	85
p3	34	34	0	54	85
p4	54	54	54	0	85
p5	85	85	85	85	0

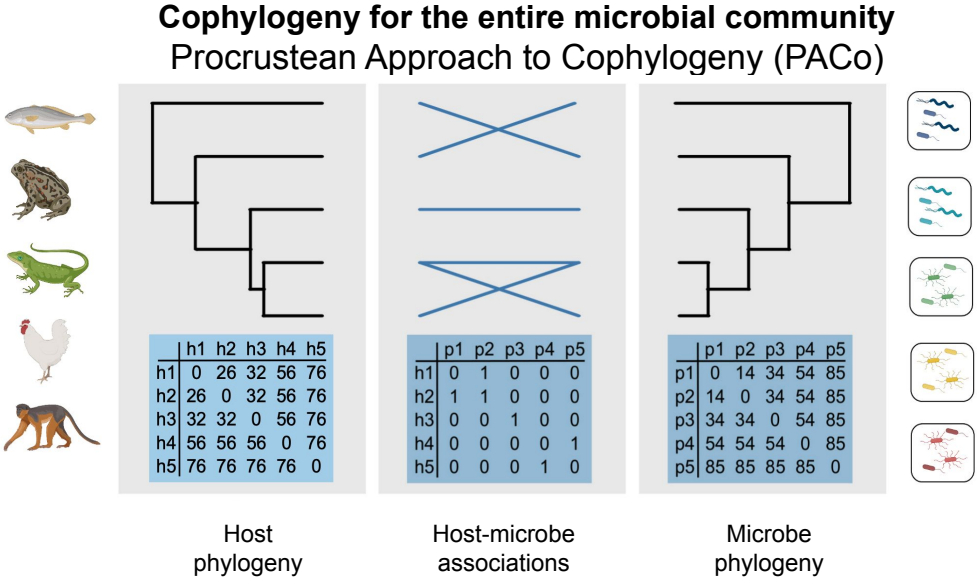
Host phylogeny

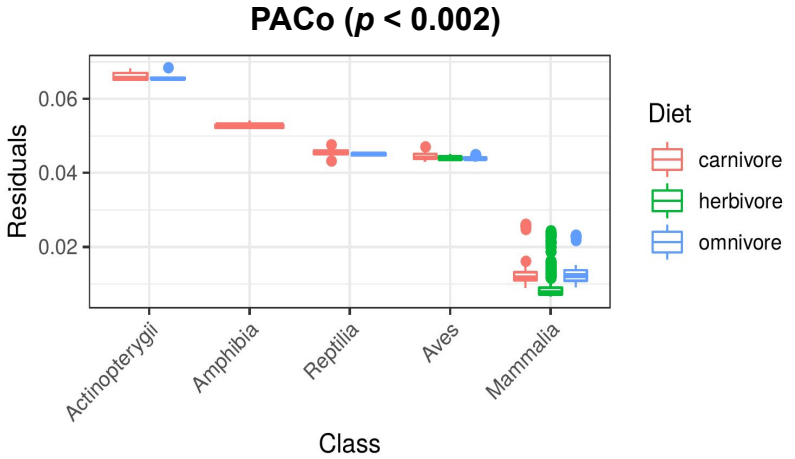
Host-microbe associations

Microbe phylogeny

Hutchinson et al., 2017. Methods in Ecology and Evolution

Pattern of cophylogeny?



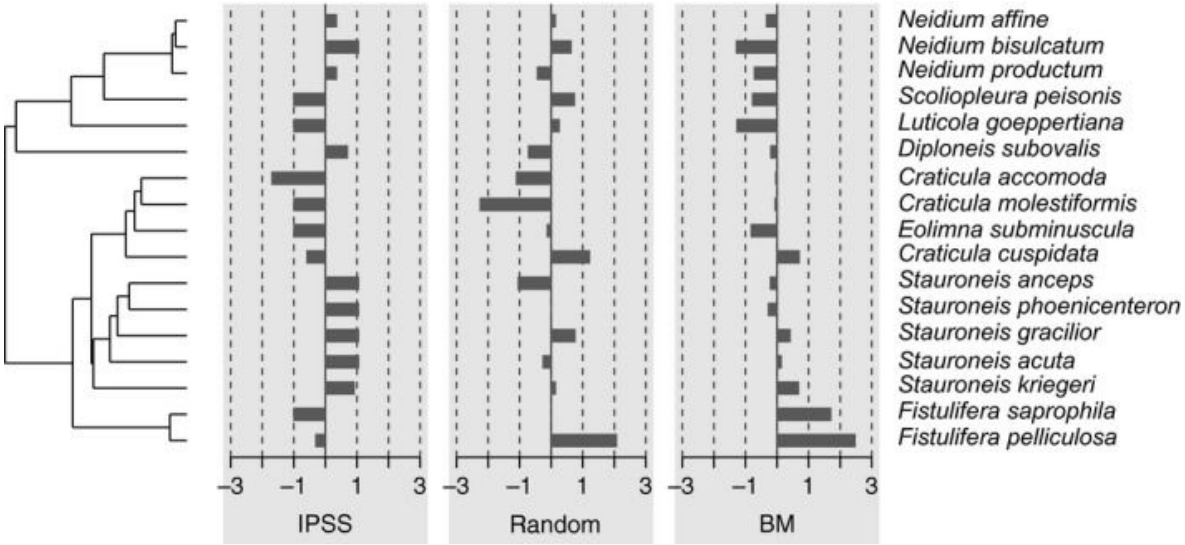


Hutchinson et al., 2017. Methods in Ecology and Evolution

Youngblut et al., 2019. Nature Communications

Specific microbe-host clade associations?

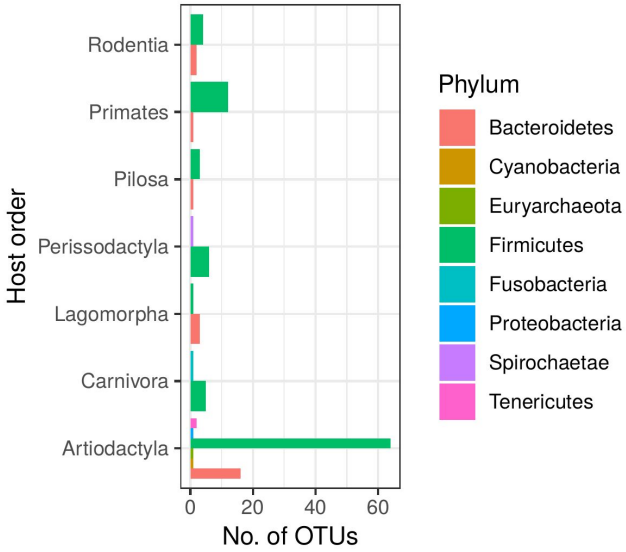
Phylogenetic signal: autocorrelation of ≥1 trait across the phylogeny

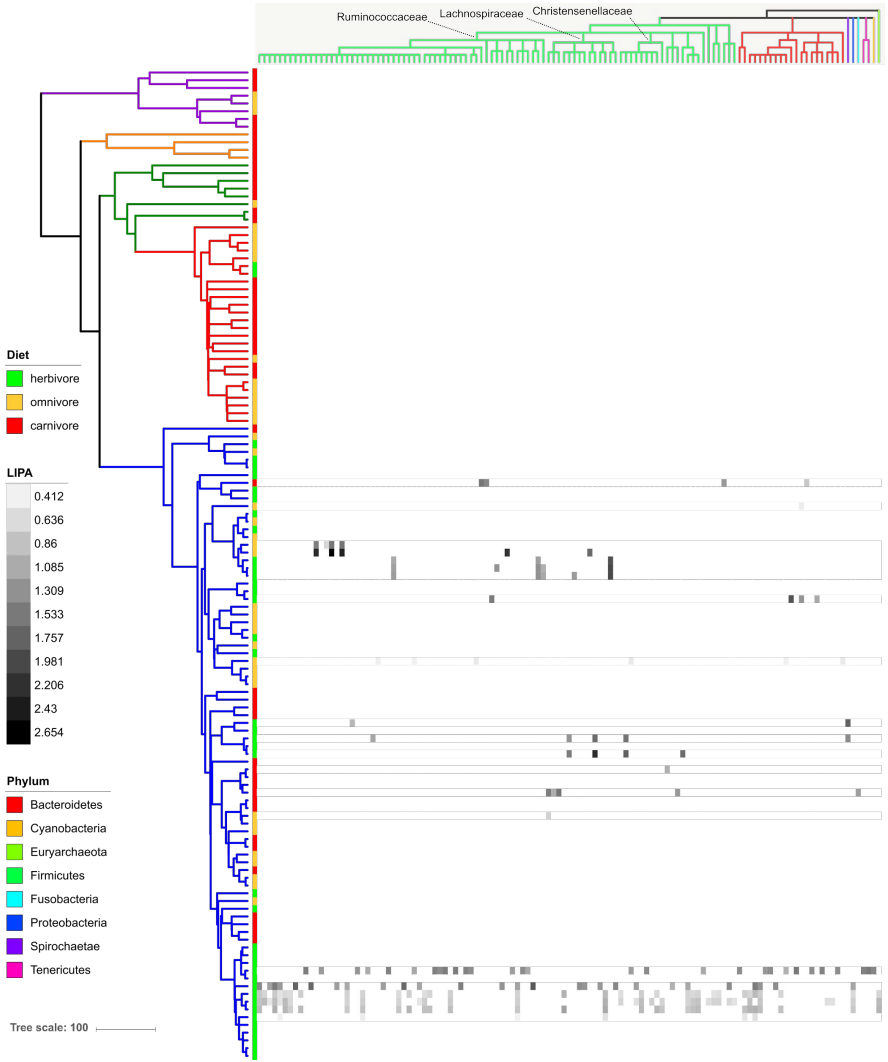


Keck et al., 2016.
Ecology and Evolution

Specific host-microbe associations

LIPA: Local Indicator of Phylogenetic Association





Youngblut et al., 2019. Nature Communications

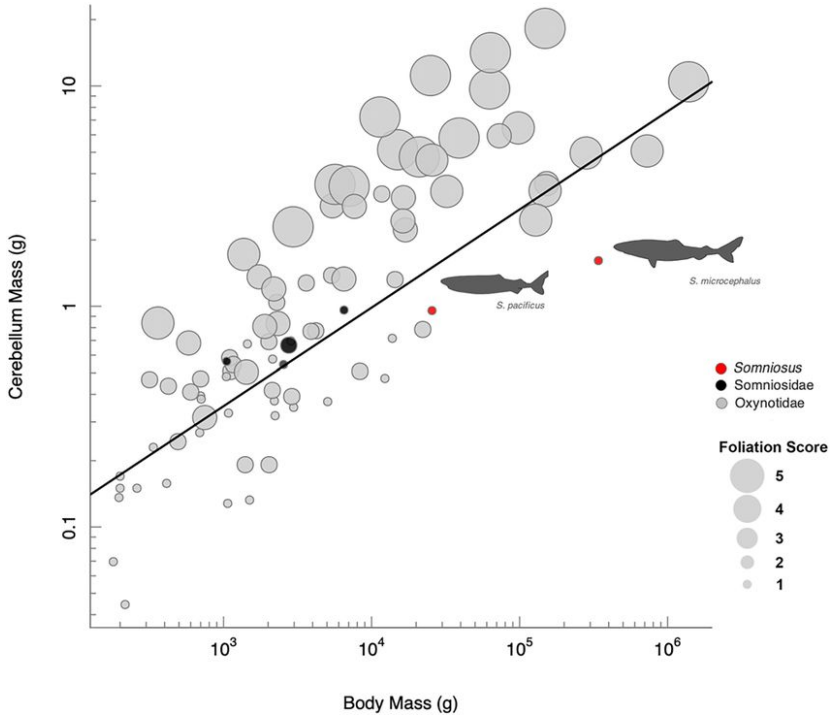
Associations with host diet

Trait-trait associations, while controlling for diet

- Phylogenetic Generalised Least Squares (PGLS)

- $Trait_1$ => host diet
- $Trait_2$ => microbial diversity

Example of PGLS on body morphology traits

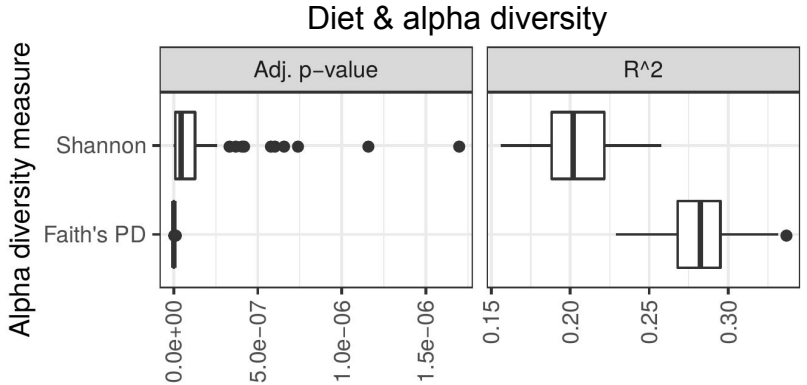


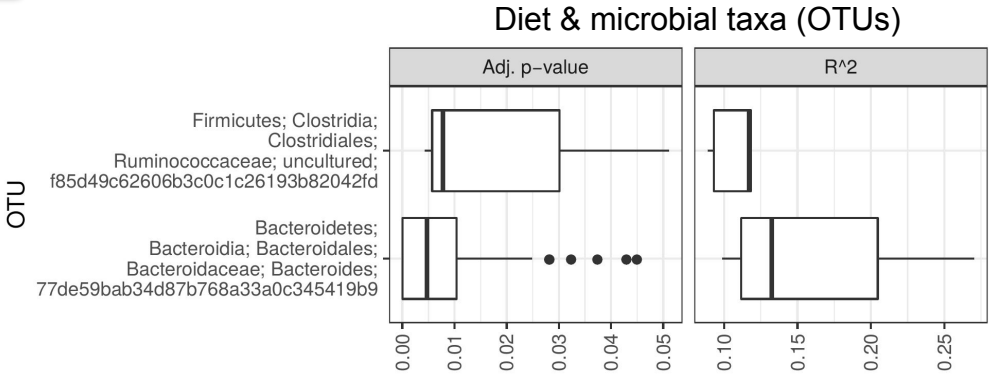
Yopak et al., 2019. Scientific Reports

Associations with host diet

Association with diet while controlling for host phylogeny?

➢ Phylogenetic Generalised Least Squares (PGLS)

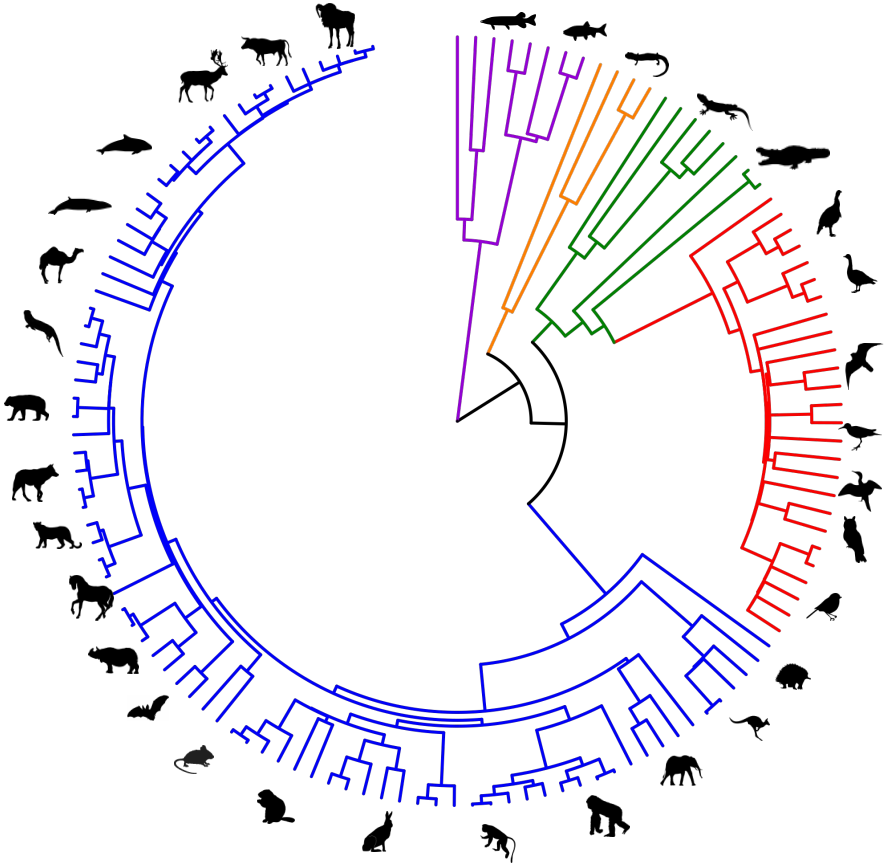




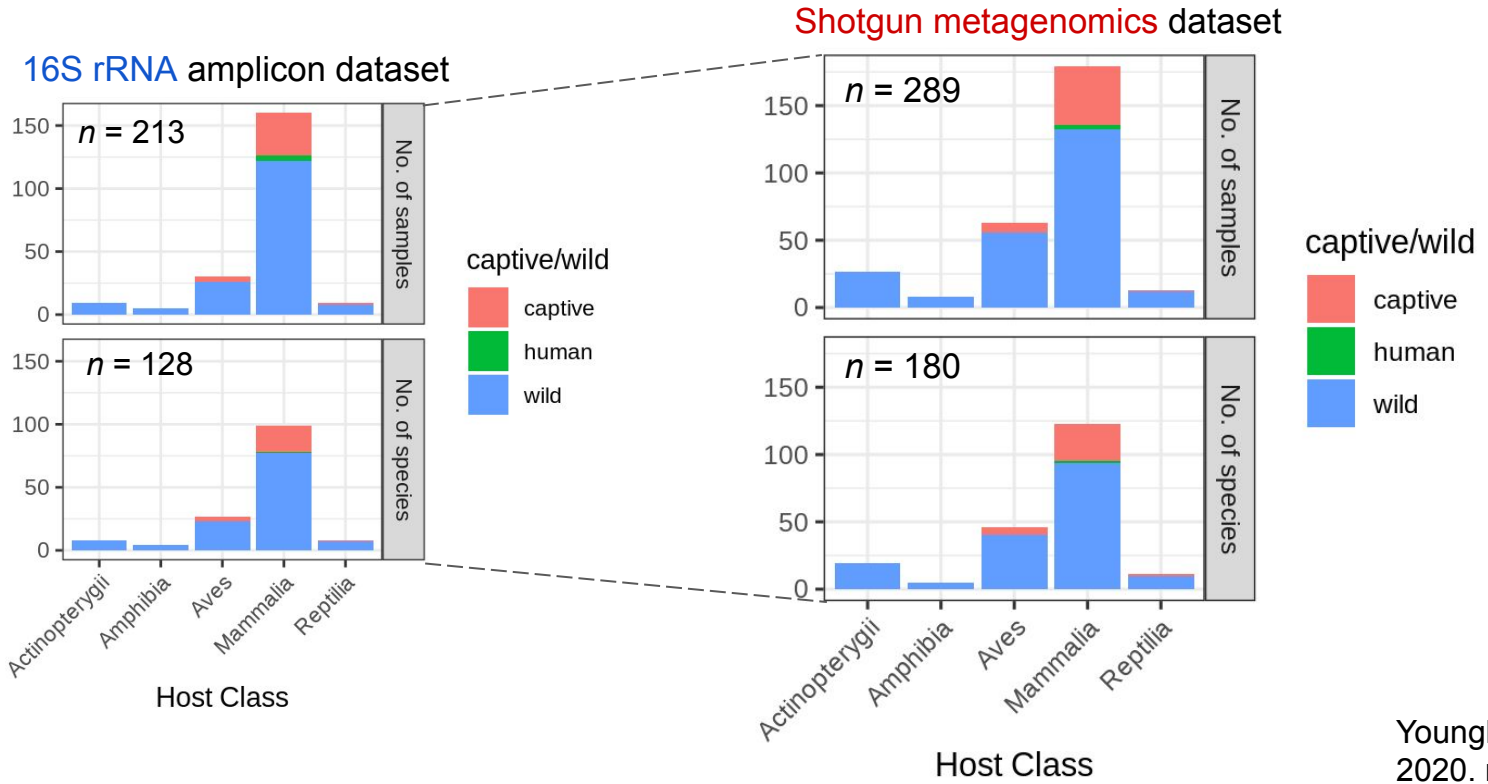
Youngblut et al., 2019. Nature Communications

Summary

- Phylosymbiosis pattern detected
  - Multiple regression on matrices (MRM)
  - Accounting for other factors
- Significant cophylogeny pattern
  - PACo
  - Strongest for mammals
- Diet stronger factor than host phylogeny
  - Selecting for total diversity versus particular microbes (PGLS)

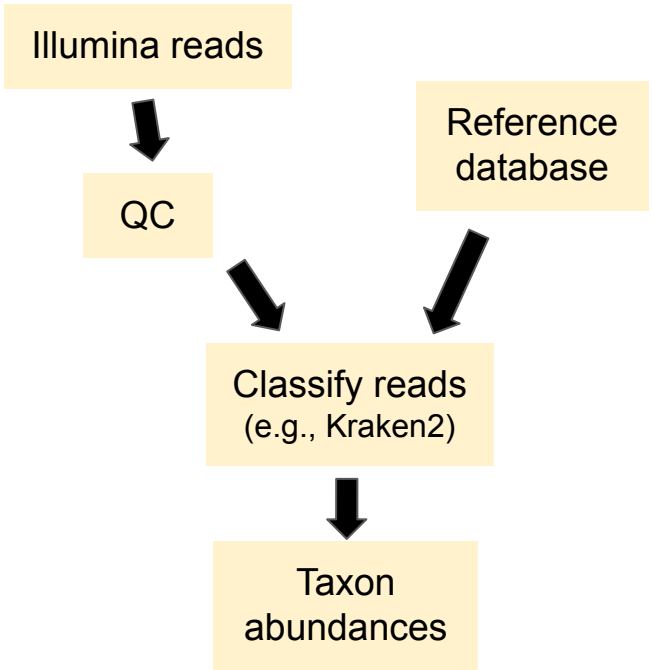


Beyond 16S rRNA: utilizing metagenomics



Youngblut et al., 2020. mSystems

Metagenome profiling: improving the reference database

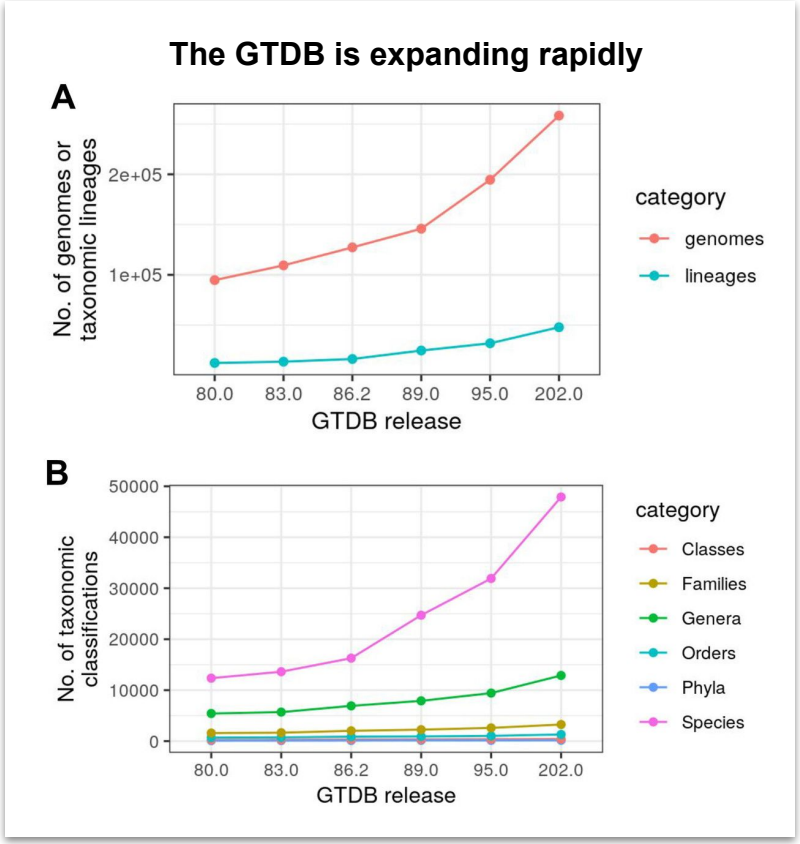


Genome Taxonomy Database (GTDB)

Phil Hugenholtz

Metagenome profiling: improving the reference database

Problem: rapidly expanding size of the GTDB database.
How to scale?



Solution

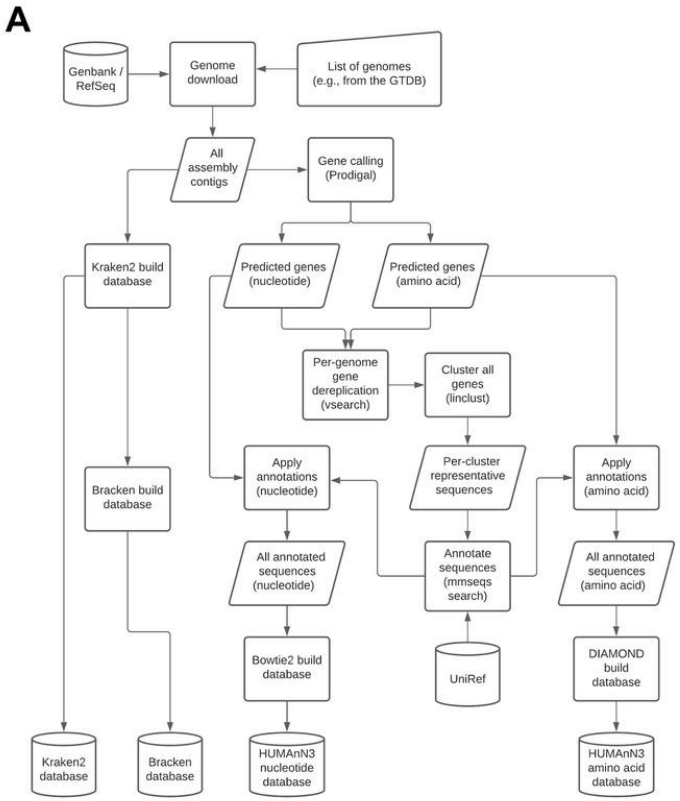
Struo2: efficient metagenome profiling database construction for ever-expanding microbial genome datasets

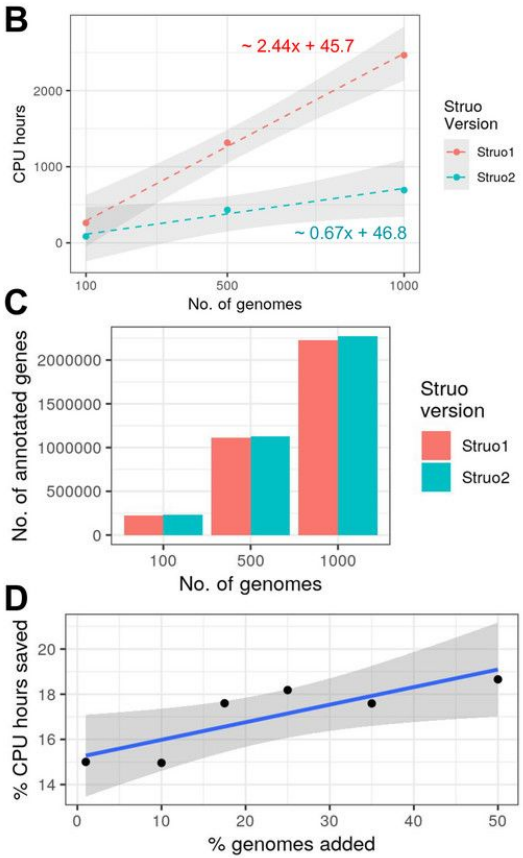
Youngblut and Ley 2021. PeerJ

Metagenome profiling: improving the reference database

Struo2: efficient metagenome profiling database construction for ever-expanding microbial genome datasets

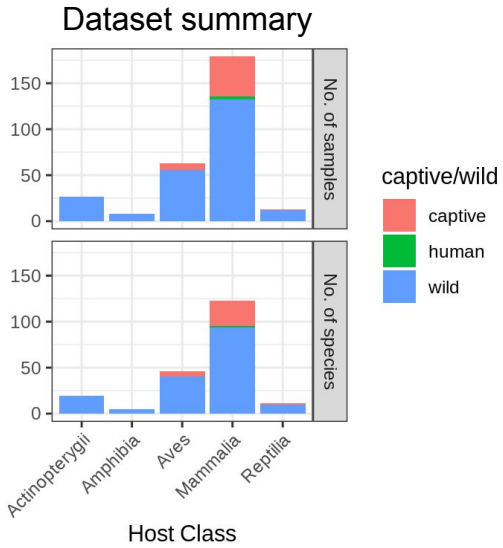
MMseqs2

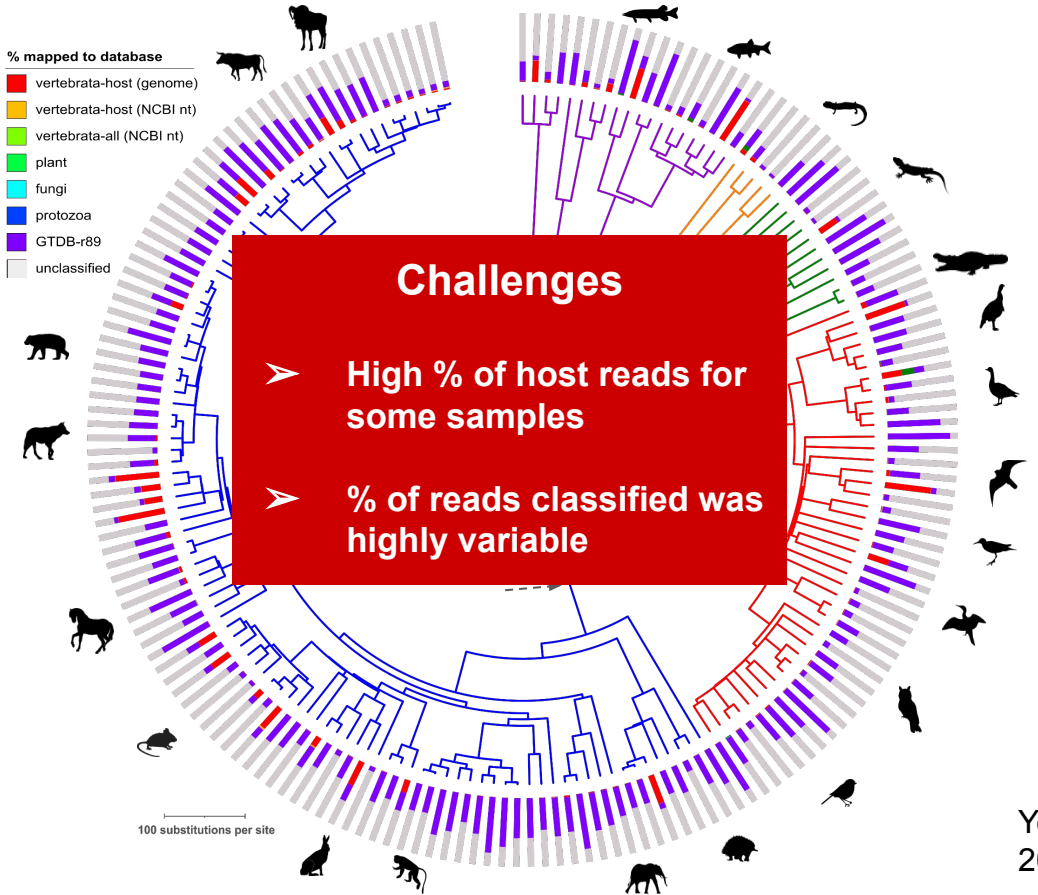




Youngblut and Ley 2021. PeerJ

Beyond 16S rRNA: utilizing metagenomics





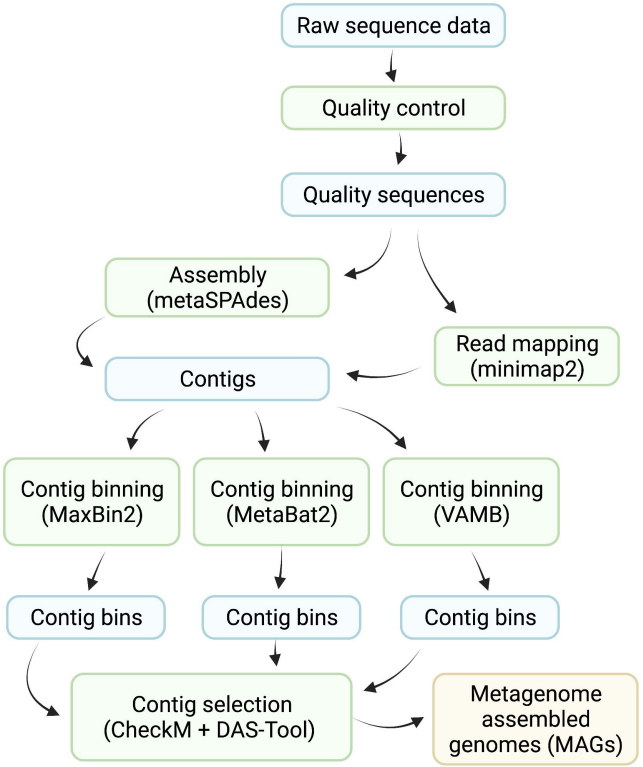
Challenges

- High % of host reads for some samples
- % of reads classified was highly variable

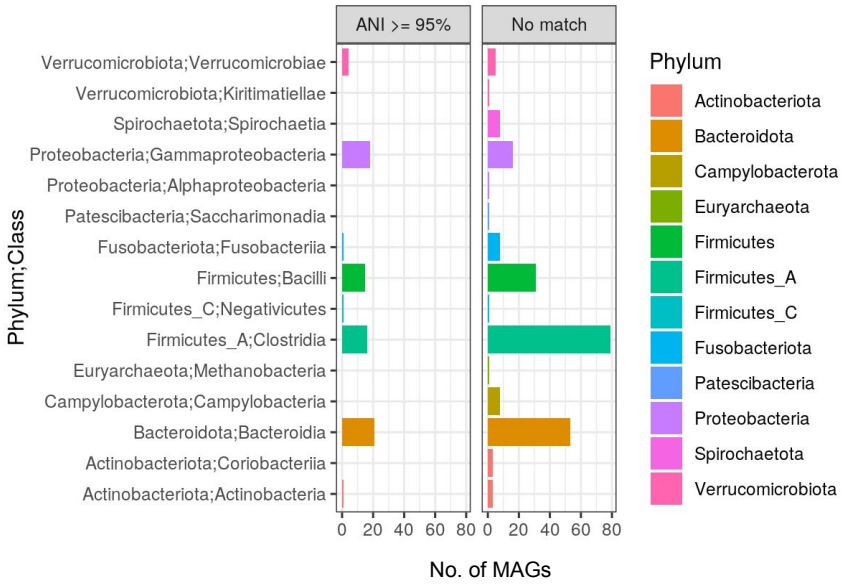
Youngblut et al., 2020. mSystems

Expanding known microbial genomic diversity via metagenome assembly

Metagenome assembly pipeline (simplified)

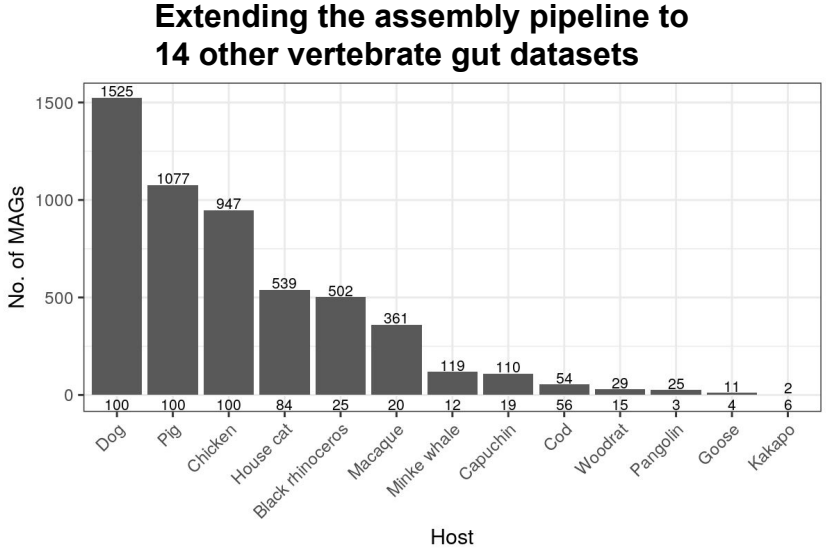


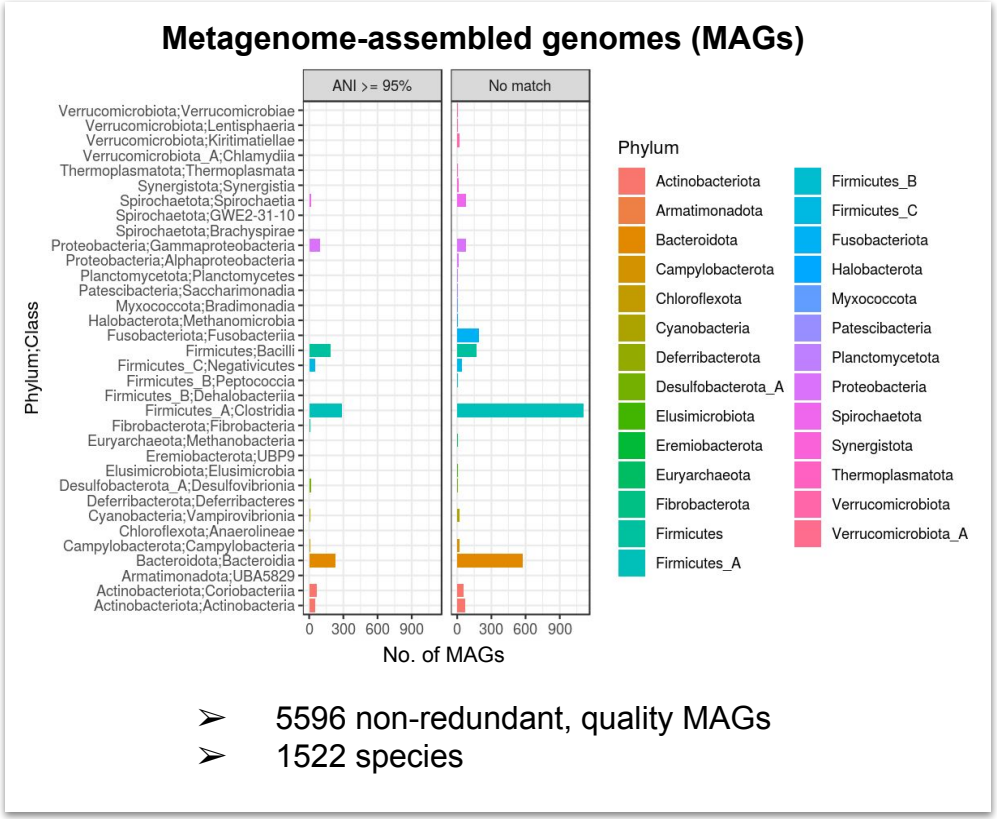
Metagenome-assembled genomes (MAGs)



- ➢ 296 non-redundant, quality MAGs
- ➢ 248 species

Expanding known microbial genomic diversity via metagenome assembly





- 5596 non-redundant, quality MAGs
- 1522 species

Youngblut et al., 2020. mSystems

Using deep learning to improve metagenome assemblies

Bioinformatics, 36(10), 2020, 3011–3017
doi: 10.1093/bioinformatics/btaa124
Advance Access Publication Date: 25 February 2020
Original Paper

OXFORD

Genome analysis

DeepMAsED: evaluating the quality of metagenomic assemblies

Olga Mineeva[1,2,†], Mateo Rojas-Carulla[1,†], Ruth E. Ley[3], Bernhard Schölkopf[1] and Nicholas D. Youngblut[3,*]

Goal: identify errors in metagenome assemblies via deep learning


Olga Mineeva

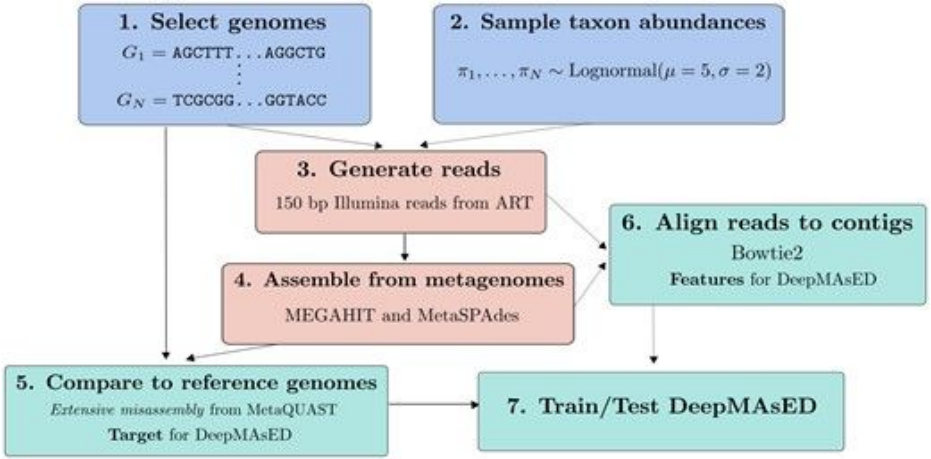

Mateo Rojas-Carulla


Bernhard Schölkopf

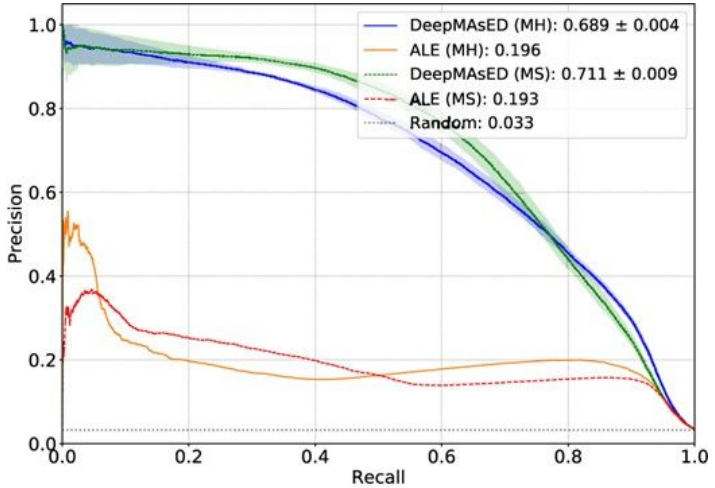

Gunnar Rätsch

Using deep learning to improve metagenome assemblies

Task: map reads back to contigs & predict misassembled contigs

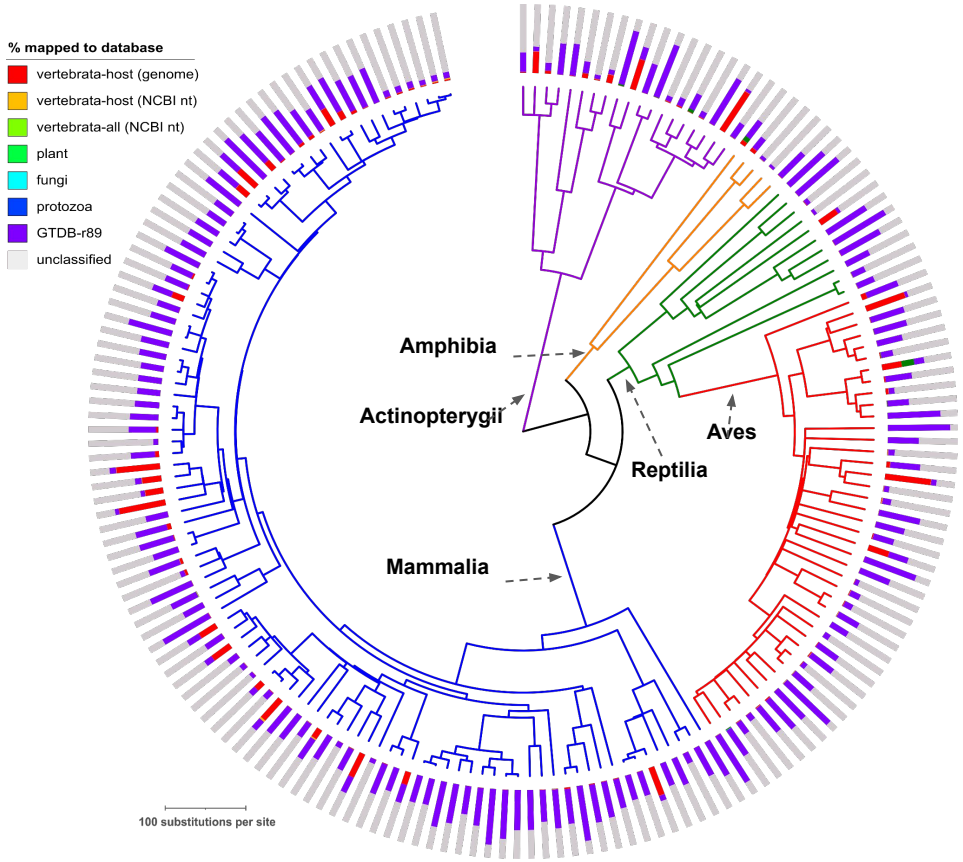


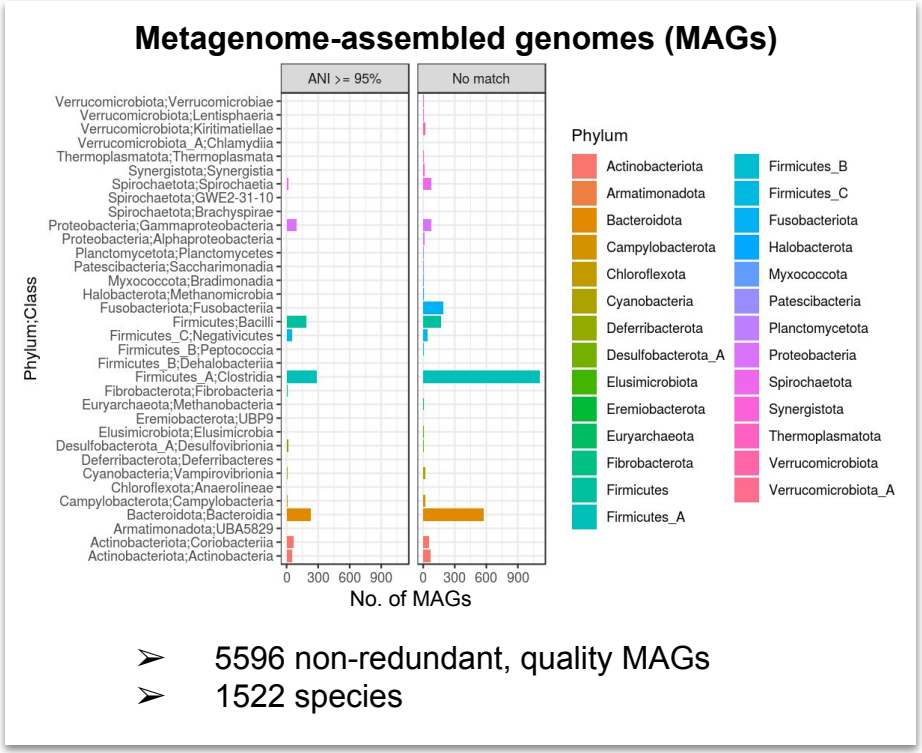
DeepMAsED is substantially better than the state-of-the-art



Mineeva et al., 2020. Bioinformatics

MAGs significantly increase the % of reads classified



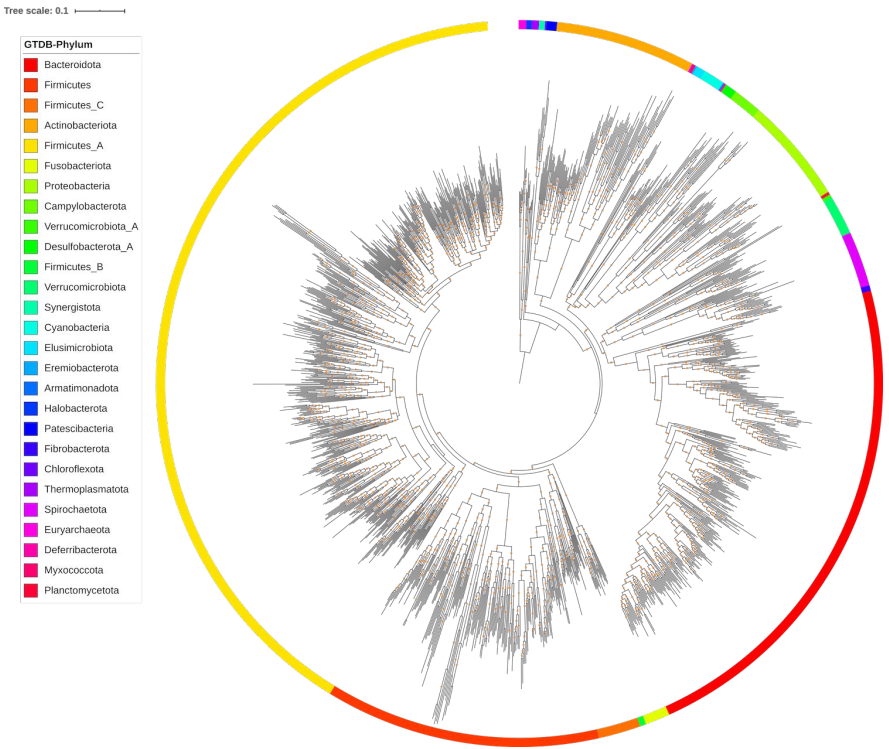


- 5596 non-redundant, quality MAGs
- 1522 species

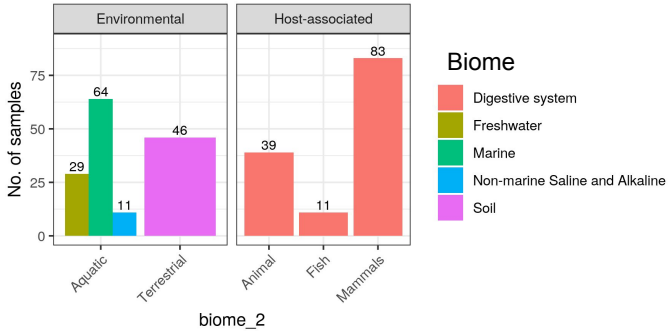
Youngblut et al., 2020. mSystems

Most MAGs are host-associated

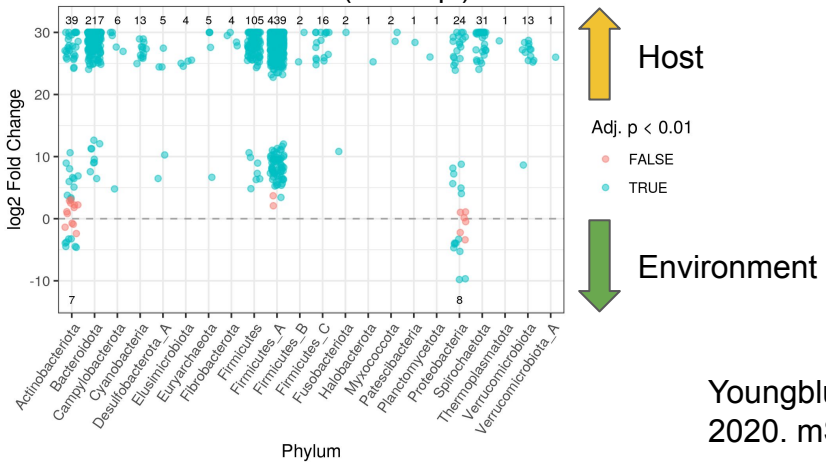
1522 species-level MAGs
(5596 at the strain-level)



"host-environment" metagenome dataset

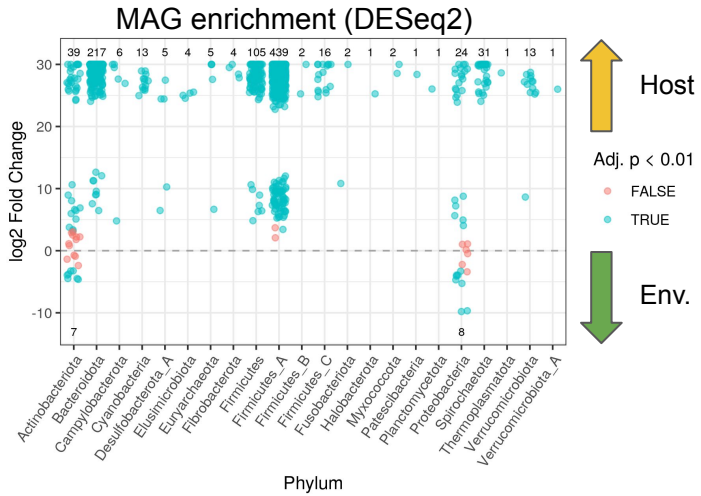


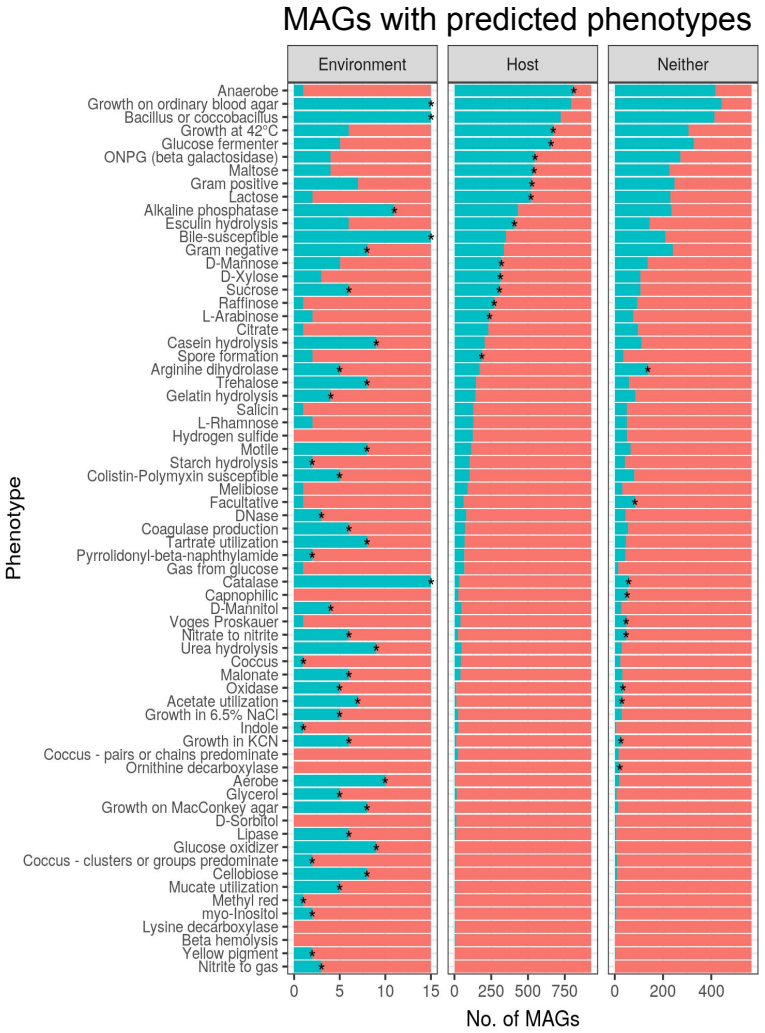
MAG enrichment (DESeq2)



Youngblut et al., 2020. mSystems

Host & env. associated traits

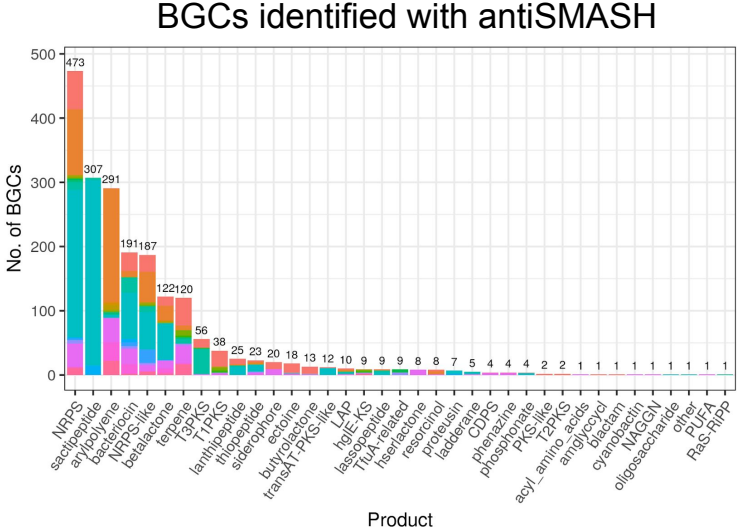


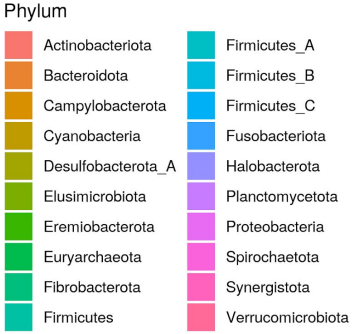


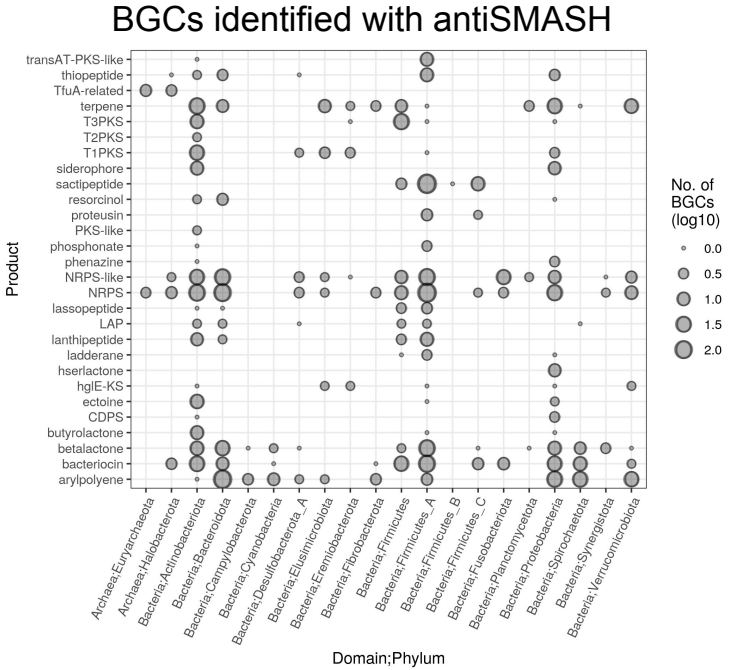
Machine learning: predicting traits based on Pfam counts

Youngblut et al., 2020. mSystems

A large diversity of biosynthetic gene clusters (BGCs)



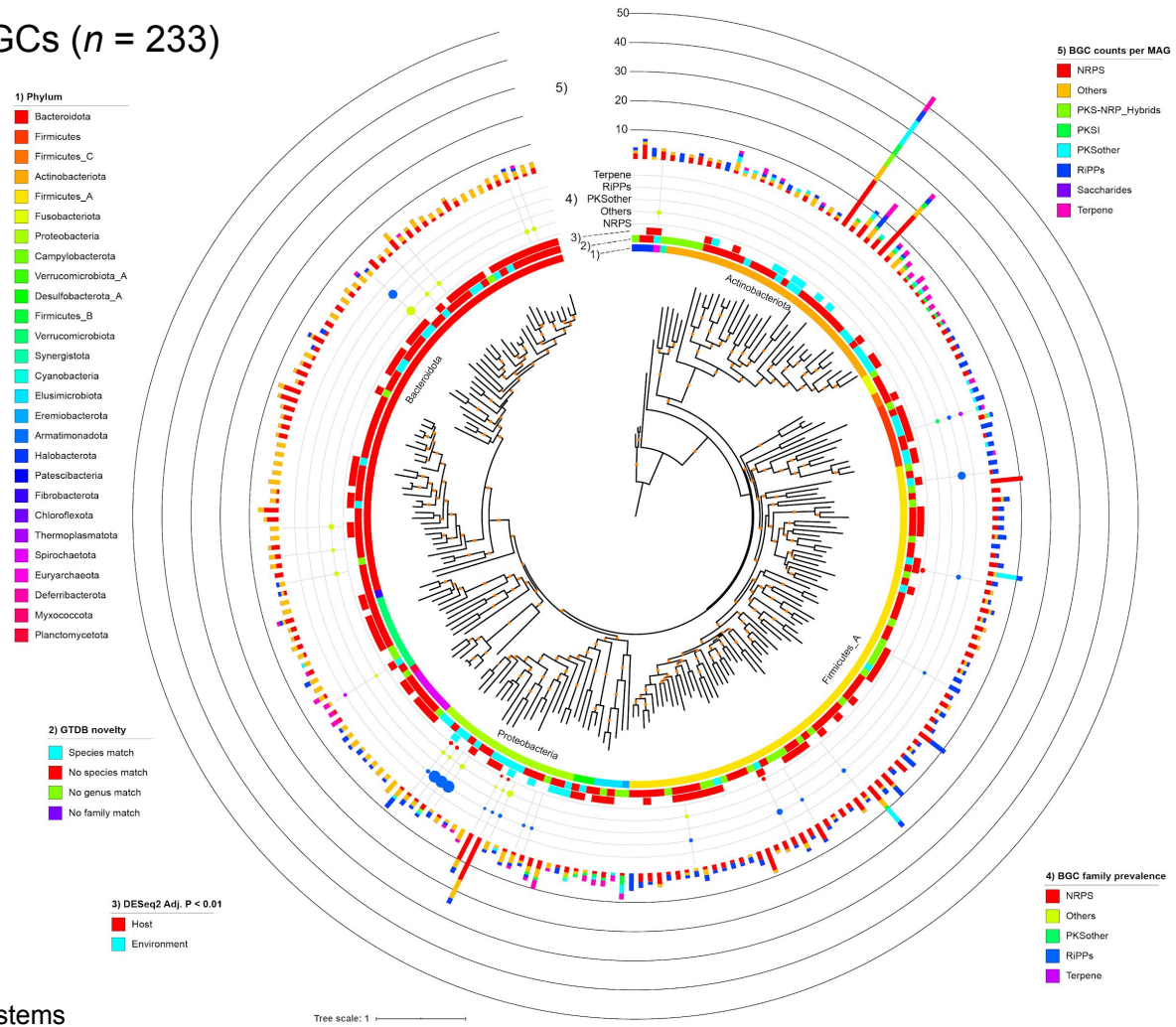




- 1986 BGCs
- High novelty based on MIBiG

Youngblut et al., 2020. mSystems

Species with ≥3 BGCs (*n* = 233)



Youngblut et al., 2020. mSystems

A large amount of gene-level diversity

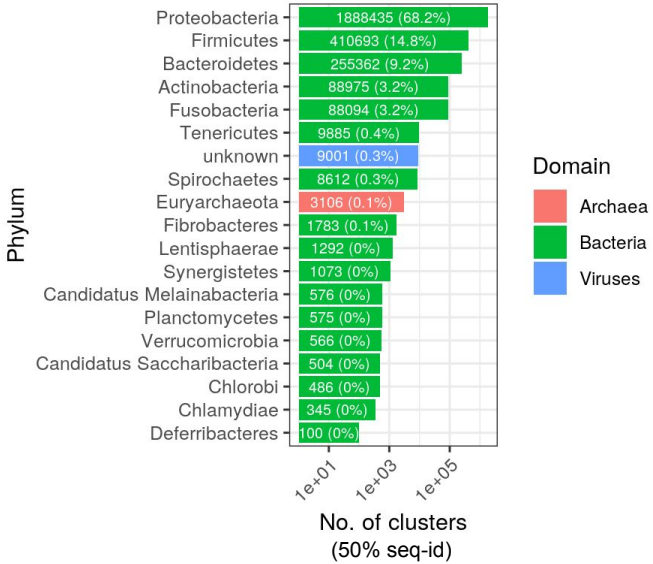
Shotgun metagenomics
(Illumina HiSeq)

Gene-based assembly: PLASS

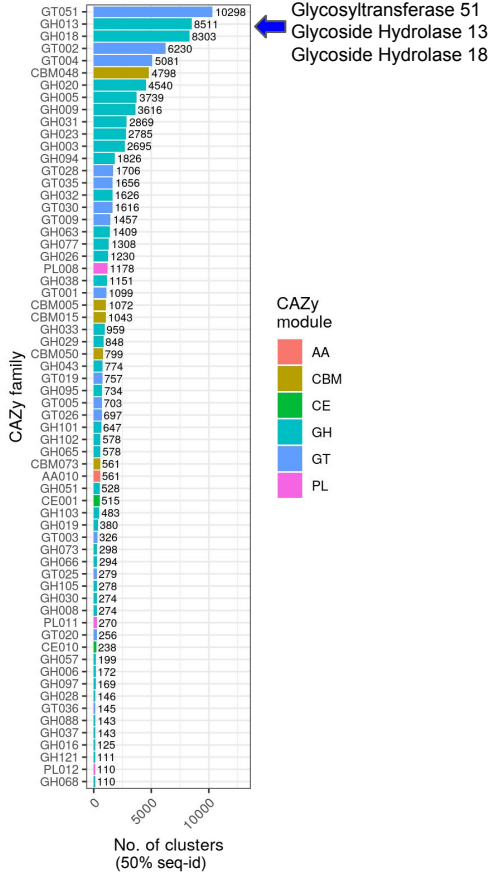
Cluster: linclust

Steinegger et al., 2019

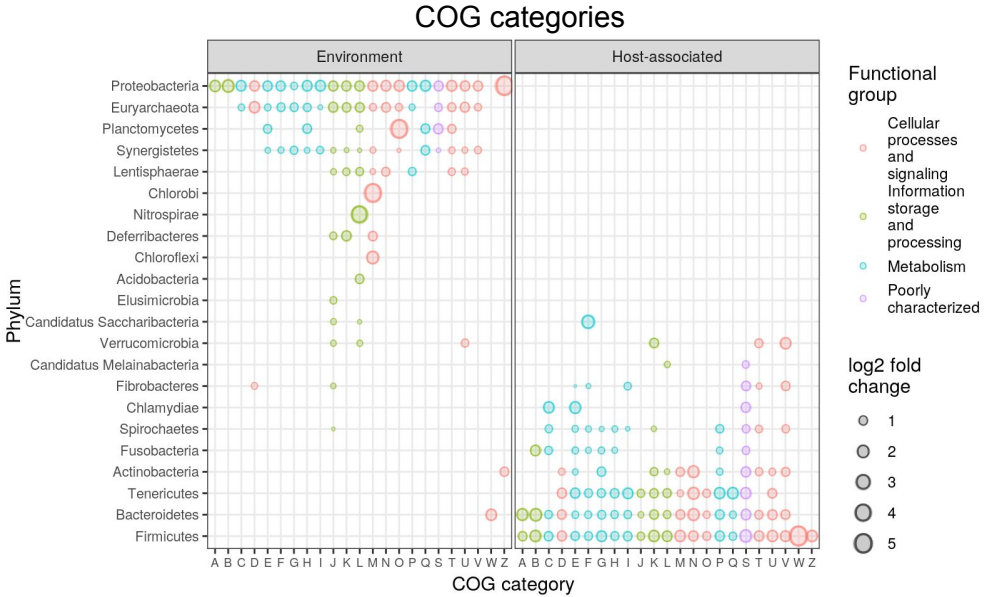
>150 million gene sequences

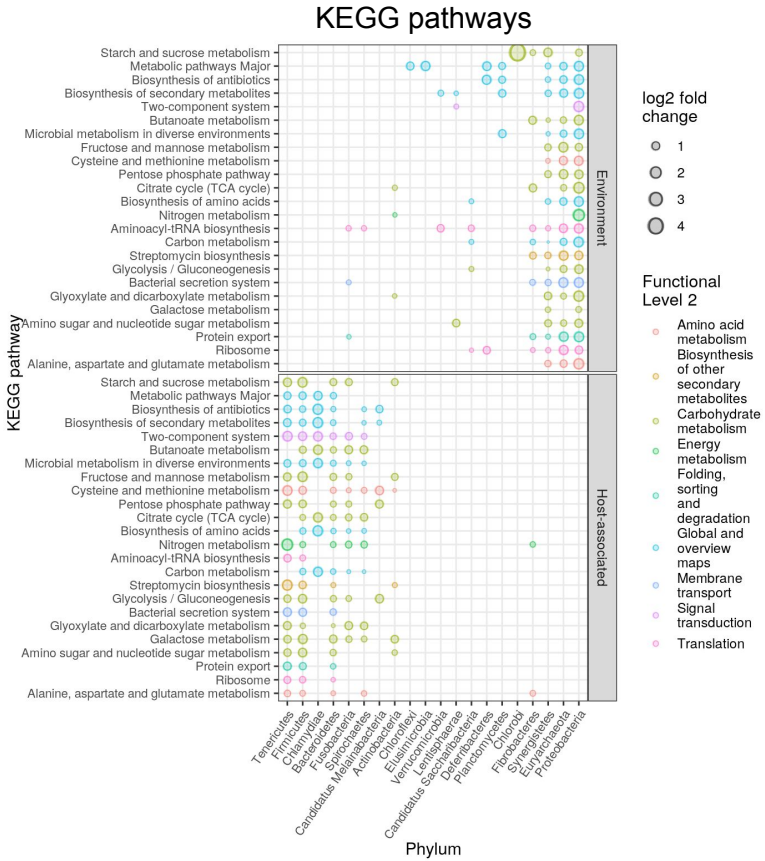


Youngblut et al., 2020. mSystems



Functional redundancy for major gene categories



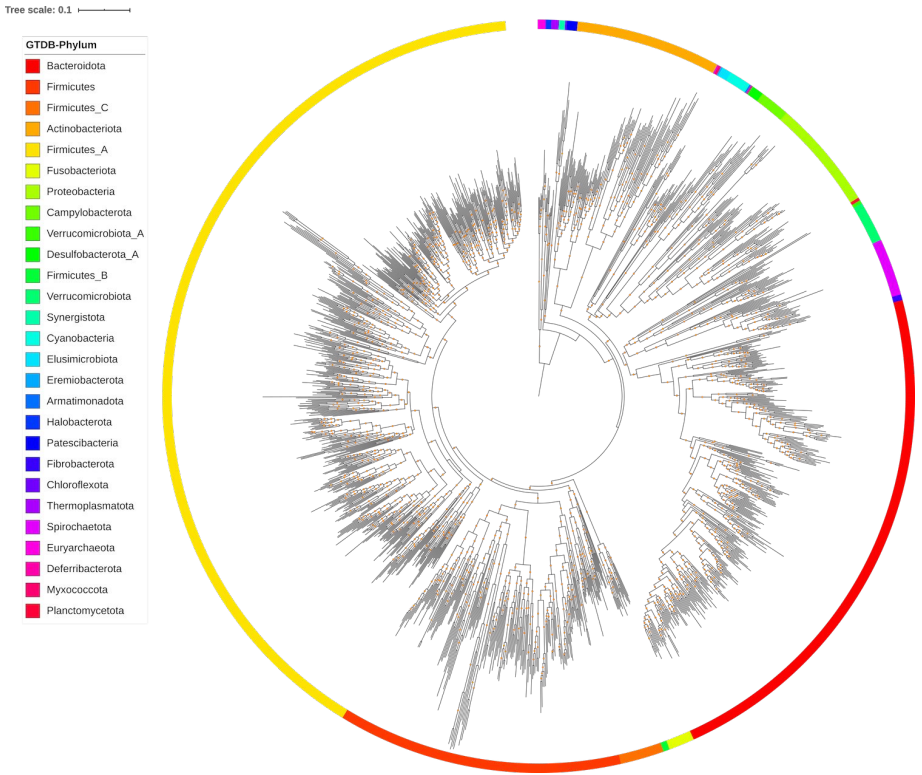


Youngblut et al., 2020. mSystems

Summary

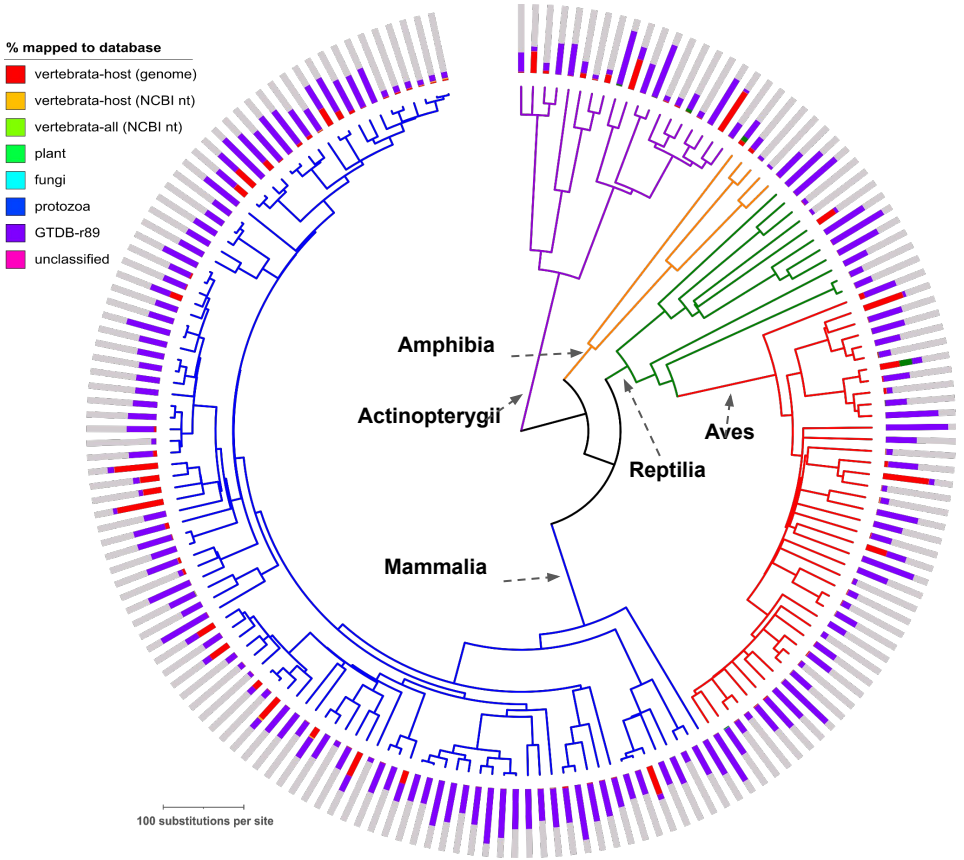
- Novel taxonomic diversity
- Most MAGs were host-associated
- Traits enriched in host- and environment-specific species
- Large BGC diversity
  - RiPPs & NRPS most prevalent
- Large gene diversity
  - Functional redundancy

1522 species-level MAGs
(5596 at the strain-level)



Future directions

- Species-level phylosymbiosis & cophylogeny
  - Genome phylogeny
- Phylosymbiosis at the functional level
  - Genes/pathways
- Diet-function associations
- Adaptation to the gut environment & symbiosis



Bioinformatics working group

What?

- Journal club
- Lecture series
- Round table discussions

When?

- Every other Thursday @ 4pm
- Zoom

Department of Microbiome Science (Ley Lab)


Georg Reischer


Andreas Farnleitner


Gabriella Stalder


Chris Walzer


Silke Dauser


Tony Walters


Jacobo de la Cuesta


Ruth Ley