



Shinichi Sunagawa

Assistant Professor of Microbiome Research

Department of Biology (D-BIOL)

Institute of Microbiology

ETH Zürich

Associated Group Leader – Swiss Institute of Bioinformatics

0. Workshop agenda

9:00 – 10:00

Session 1: Introductory lecture

10:00 – 10:30

Coffee break

10:30 – 12:00

Session 2: Preparation and installation

12:00 – 14:00

Lunch break

14:00 – 15:30

Session 3: Running mOTUs

15:30 – 16:00

Coffee break

16:30 – 17:30

Session 4: Analyzing data in [R]

Instructors



Hans Ruscheweyh



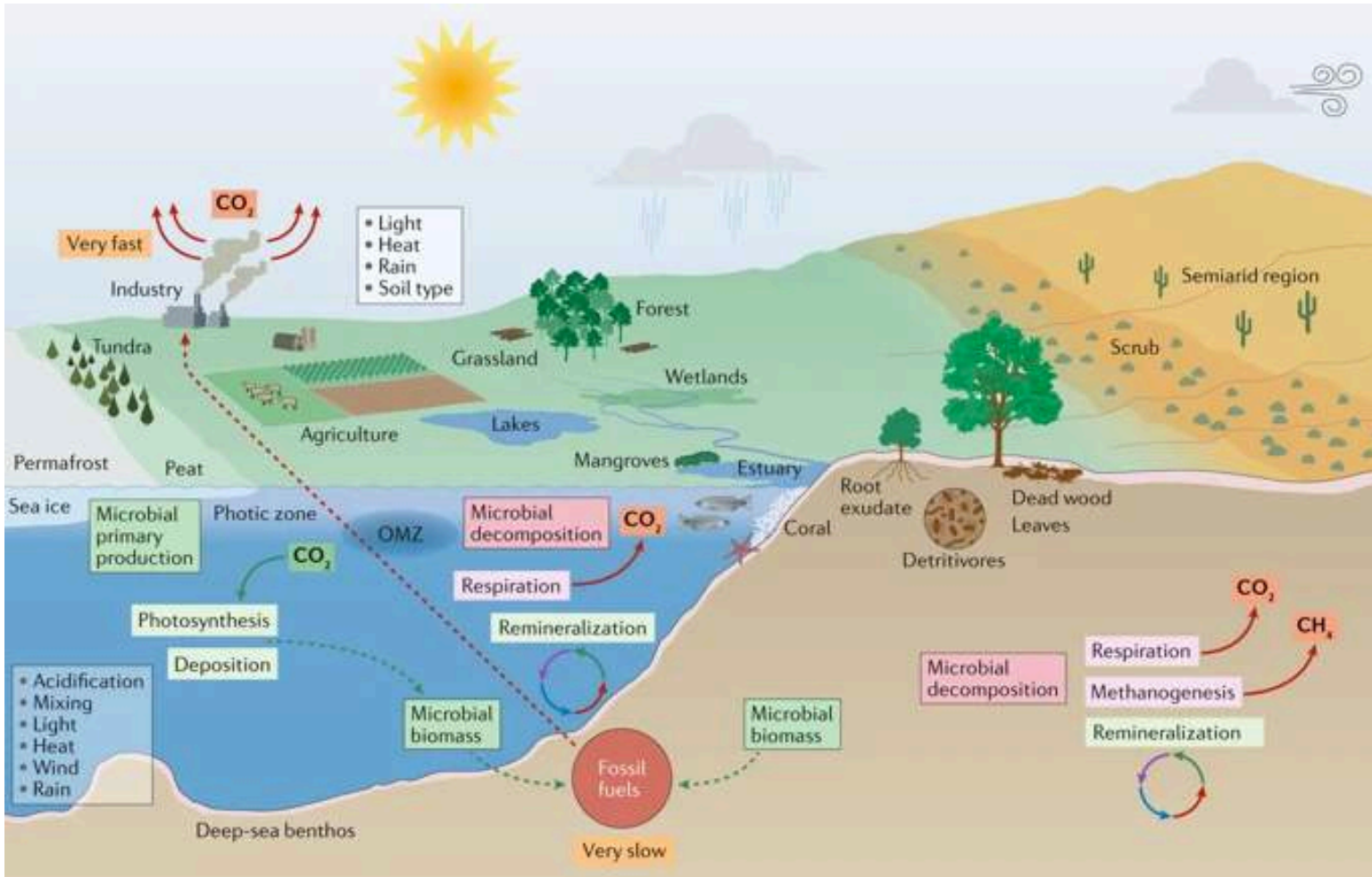
Chris Field

Overview

1. Introduction - Microbial communities / meta[gen | transcript]omics
2. Taxonomic profiling of metagenomes
3. mOTUs: protein-coding single-copy marker gene-based Operational Taxonomic Units
4. Profiling with mOTUs
5. Summary + Outlook

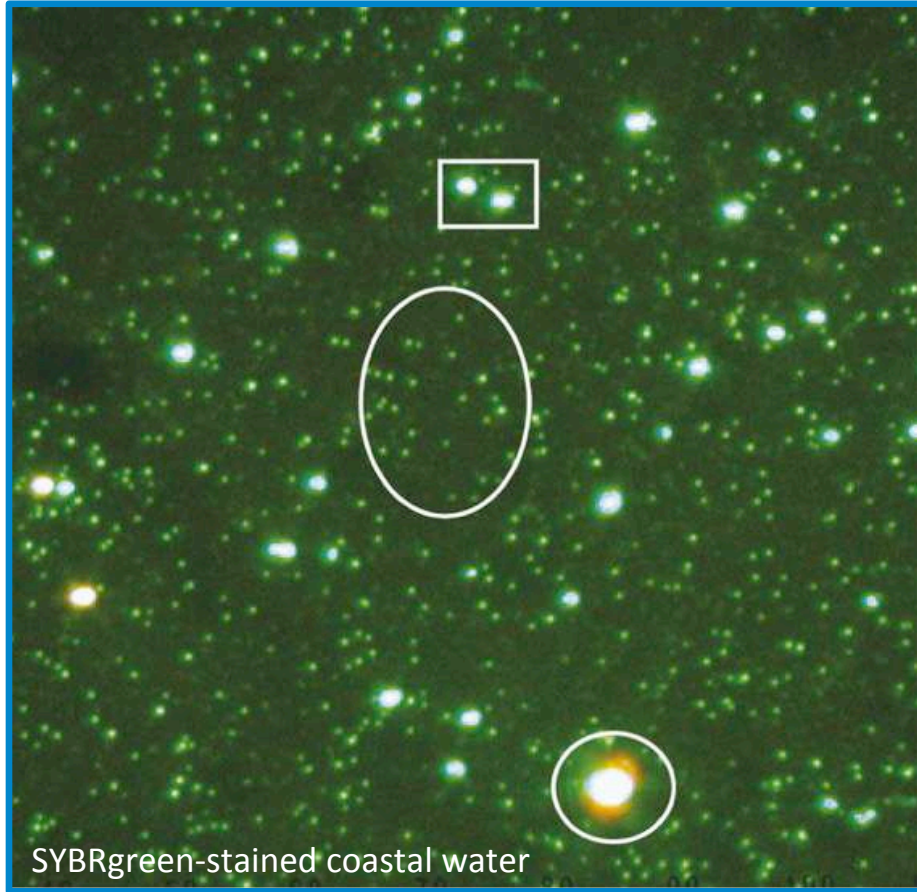
1. Introduction - Microbial communities

Microbial communities drive global biogeochemical cycles

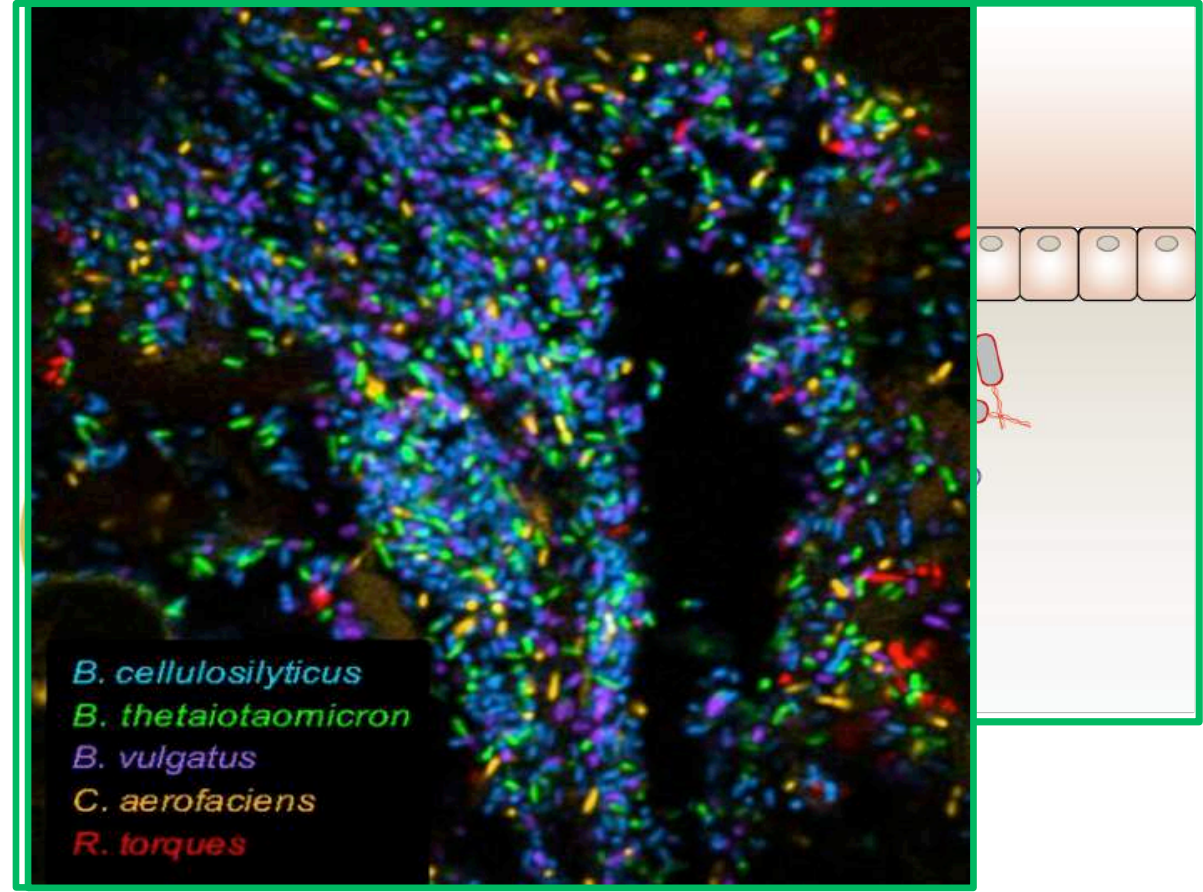


1. Introduction - Microbial communities

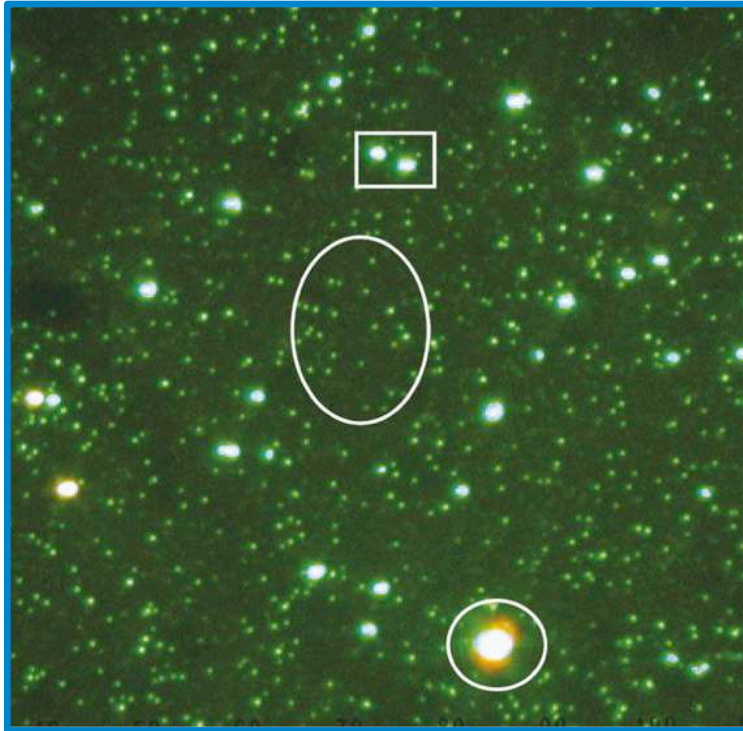
Ocean microbiome



Human gut microbiome



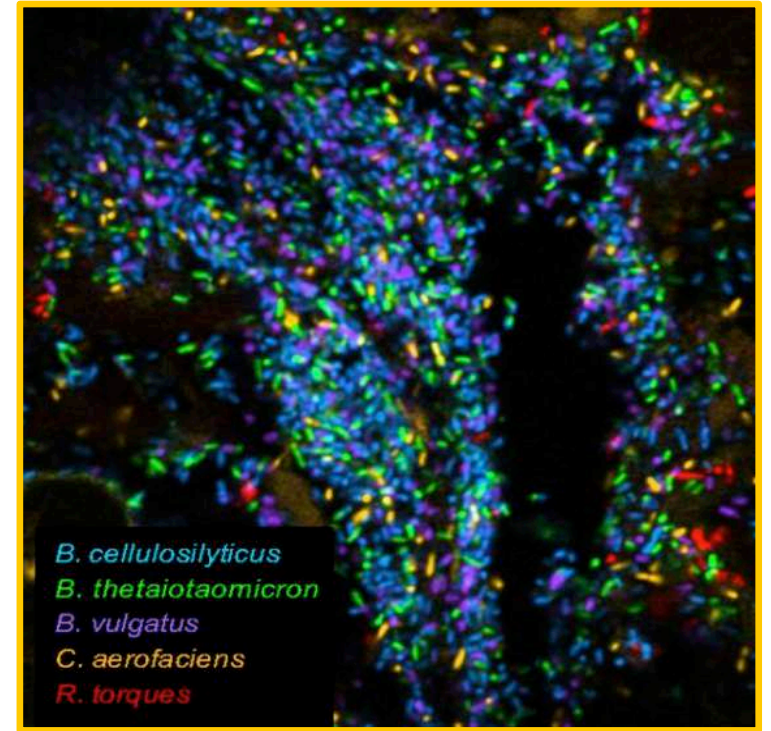
1. Introduction - Meta[gen | transcript]omics



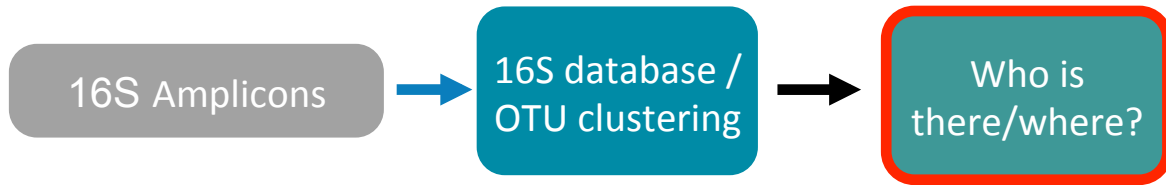
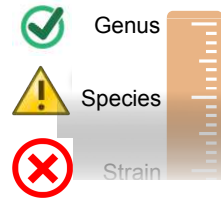
Who is there/where?

What can they do?

Who can do what?



1. Introduction - Meta[gen | transcript]omics

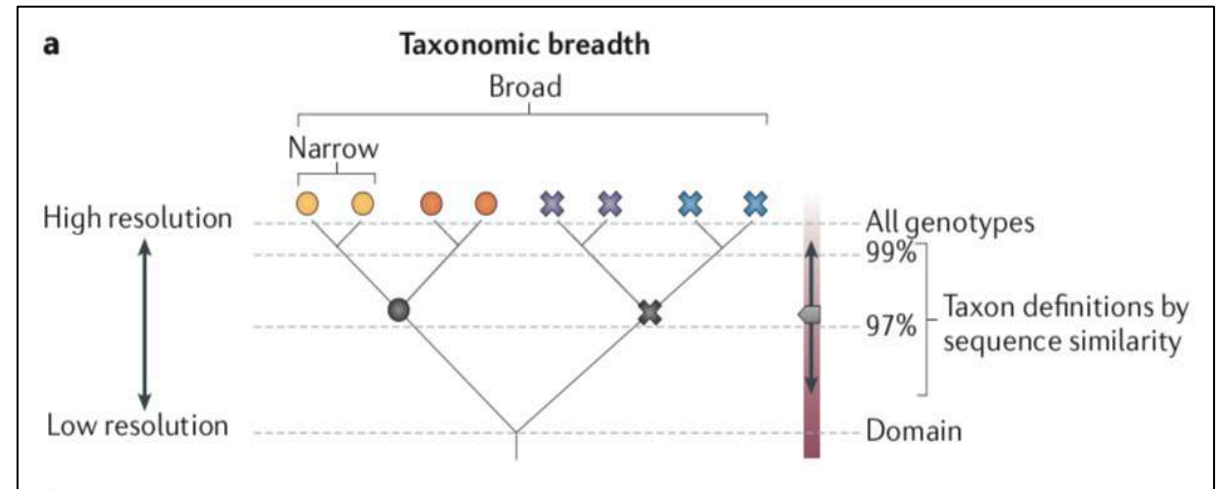


- low phylogenetic resolution
- copy number variation
- PCR primers not universal
- PCR region bias
- PCR chimera

Microbiologist have adopted the concept of taxonomic ranks

- Species *Escherichia coli*
- Genus *Escherichia*
- Family Enterobacteriaceae
- Order Enerobacterales
- Class gamma-Proteobacteria
- Phylum Proteobacteria
- Domain Bacteria

and (now) use DNA sequence similarity as the method of choice to determine genetic relatedness:



Hanson et al., NRM, 2012

Mende, Sunagawa et al., Nat Methods, 2013

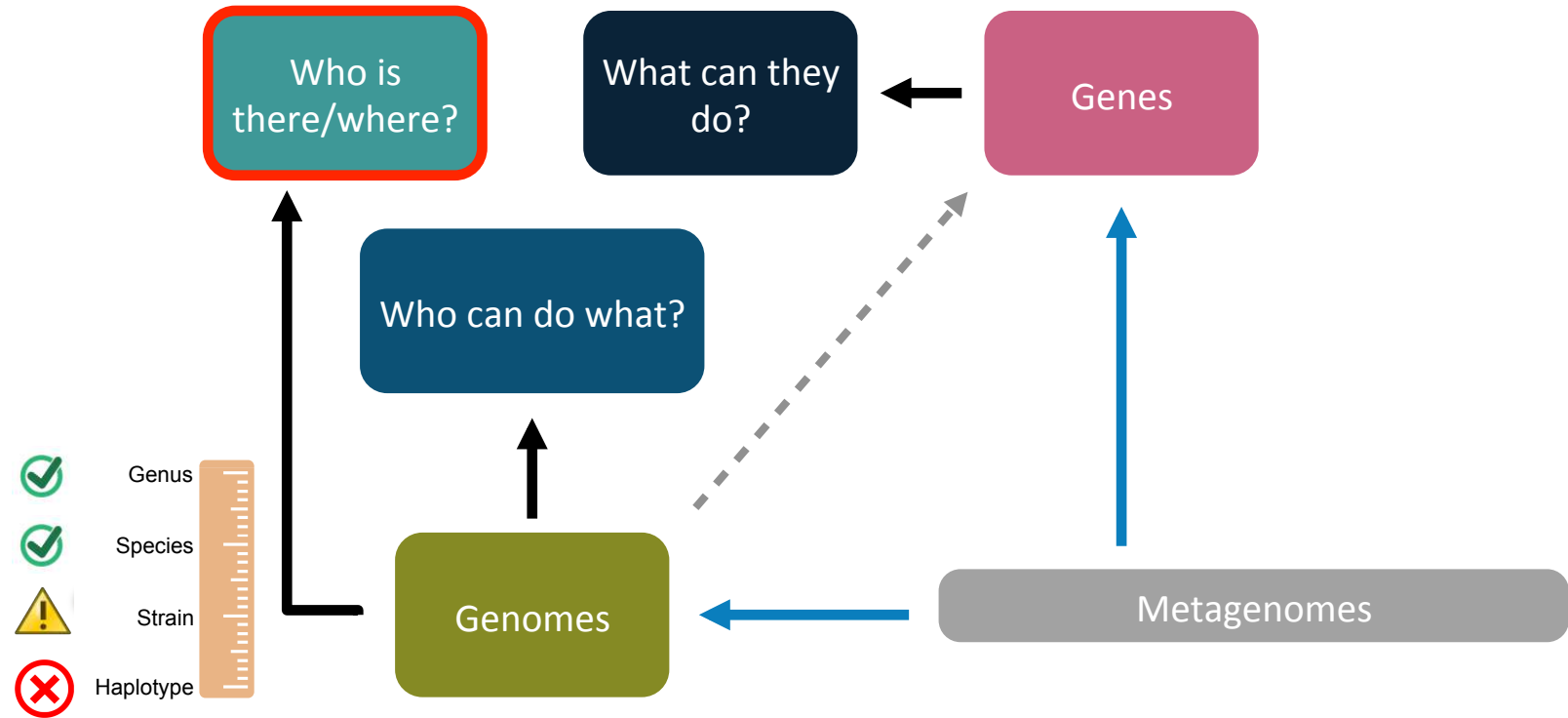
Klappenbach et al., NAR, 2001

Parada et al., Environ Microbiol, 2016

Claesson et al., NAR, 2010

Haas et al., Genome Res, 2011

1. Introduction - Meta[gen | transcript]omics



Taxonomic profiling of metagenomes

Truong et al., Nat Meth, 2015

Lu et al., PeerJ Comp Sci, 2017

Functional profiling of metagenomes

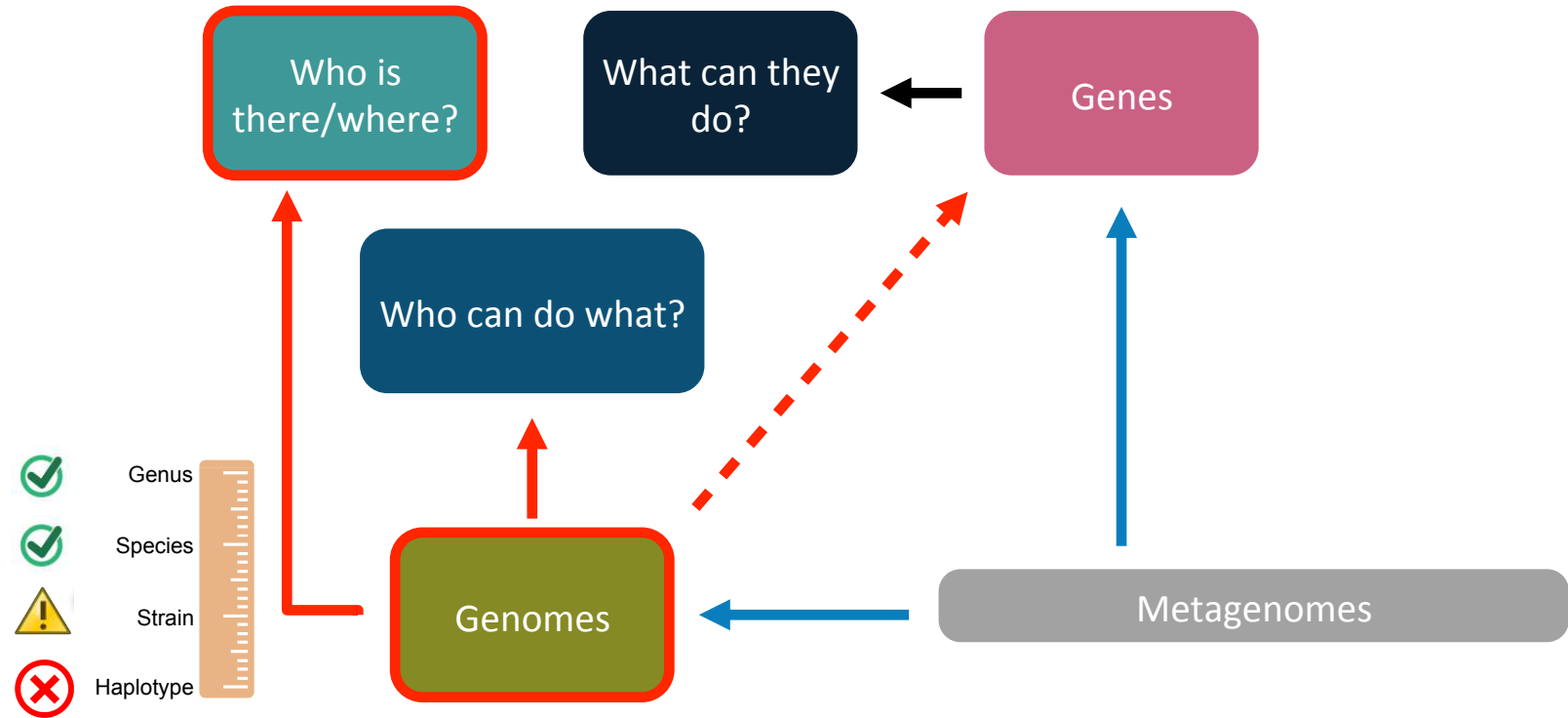
Qin et al., Nature, 2010

Arumugam et al., Nature, 2011

Huson et al., Genome Res, 2007

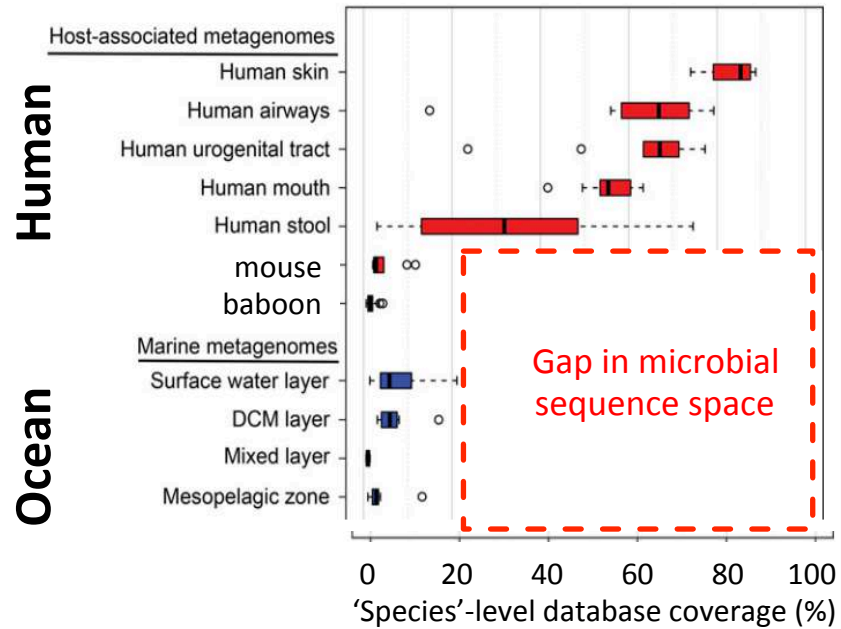
Kultima, Sunagawa, et al., PLoS One, 2012

1. Introduction - Meta[gen | transcript]omics

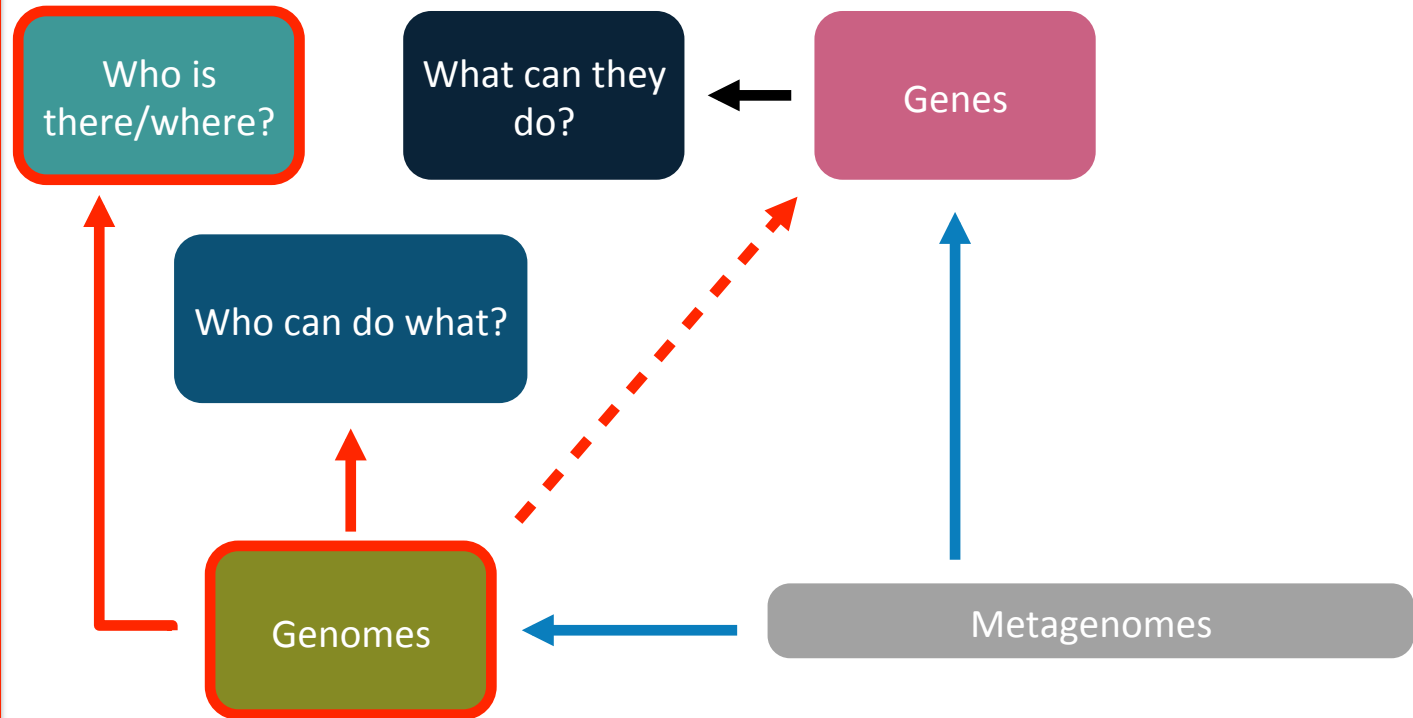


1. Introduction - Meta[gen | transcript]omics

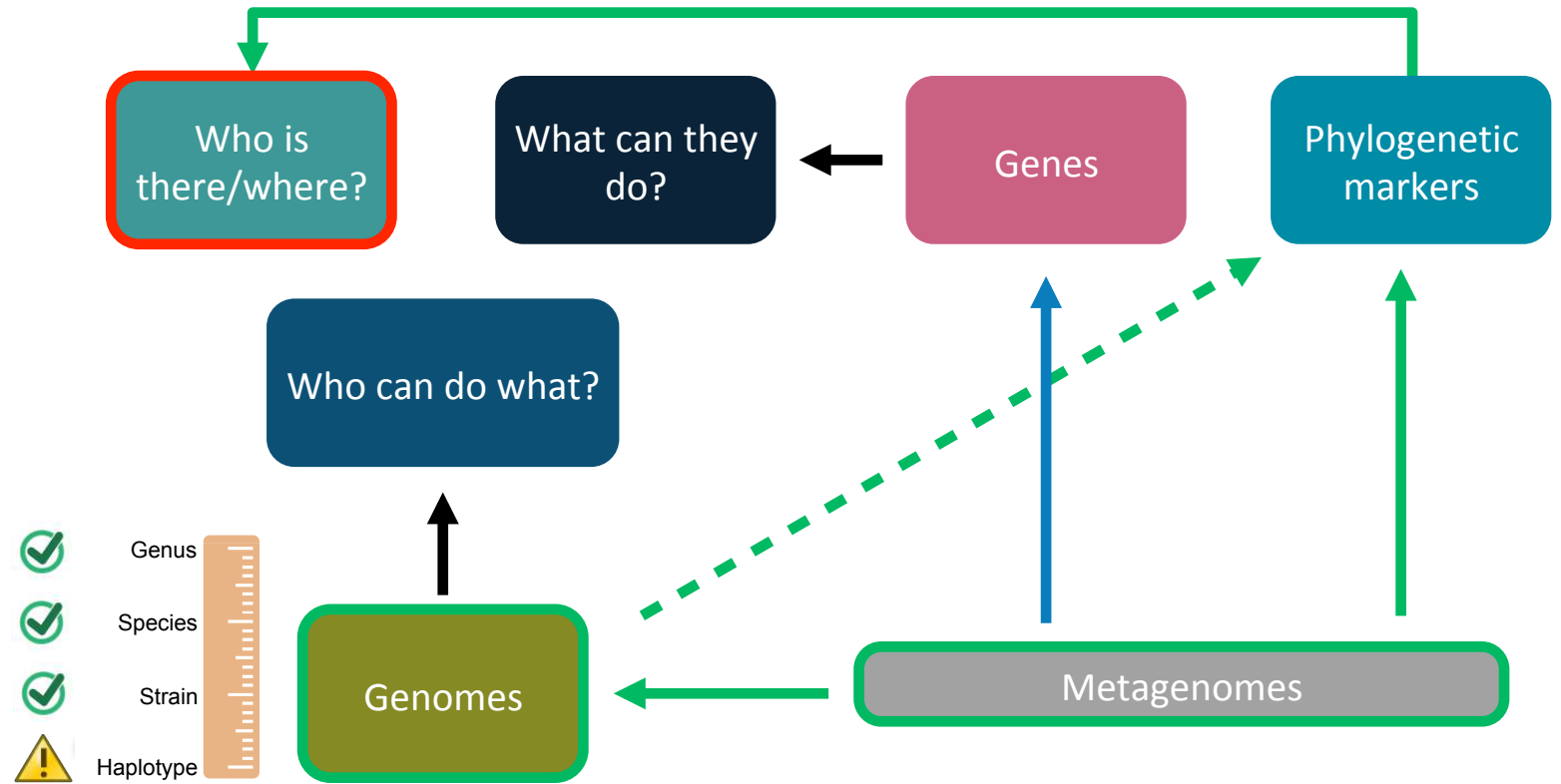
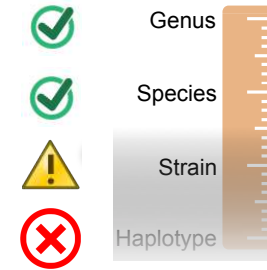
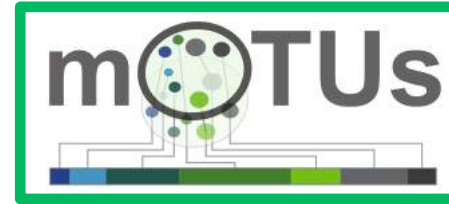
Metagenomic profiling of diverse microbiomes



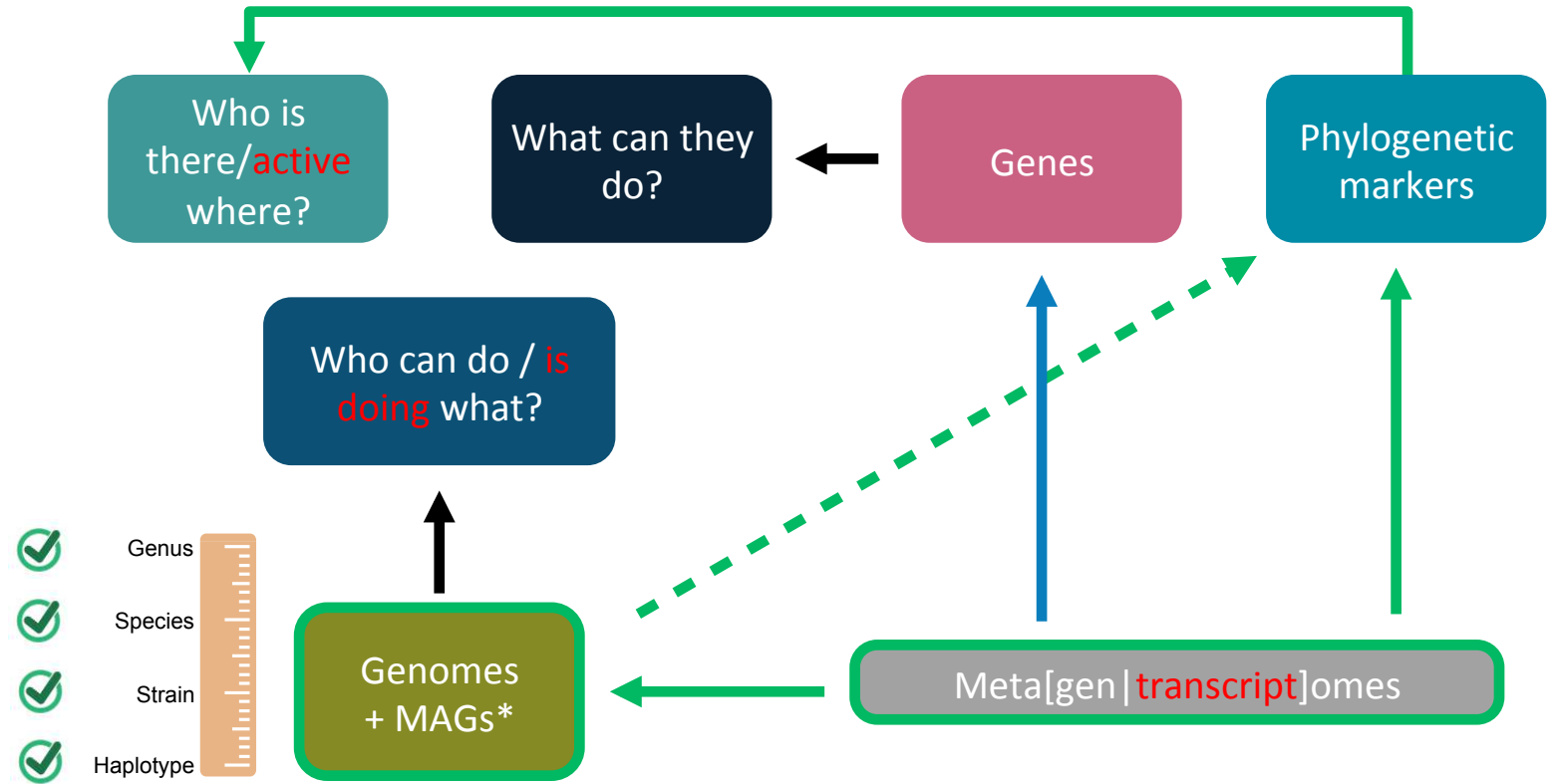
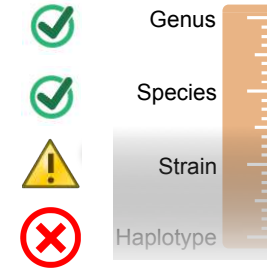
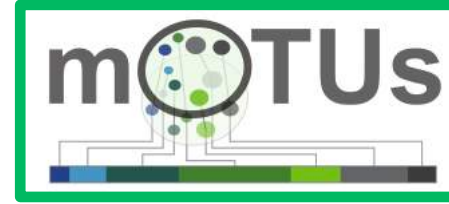
- Fraction of sequence space covered when relying on sequenced reference genomes
- Unknown sequence space is particularly large in the oceans (also for mice, apes, soil, etc.)



1. Introduction - Meta[gen | transcript]omics

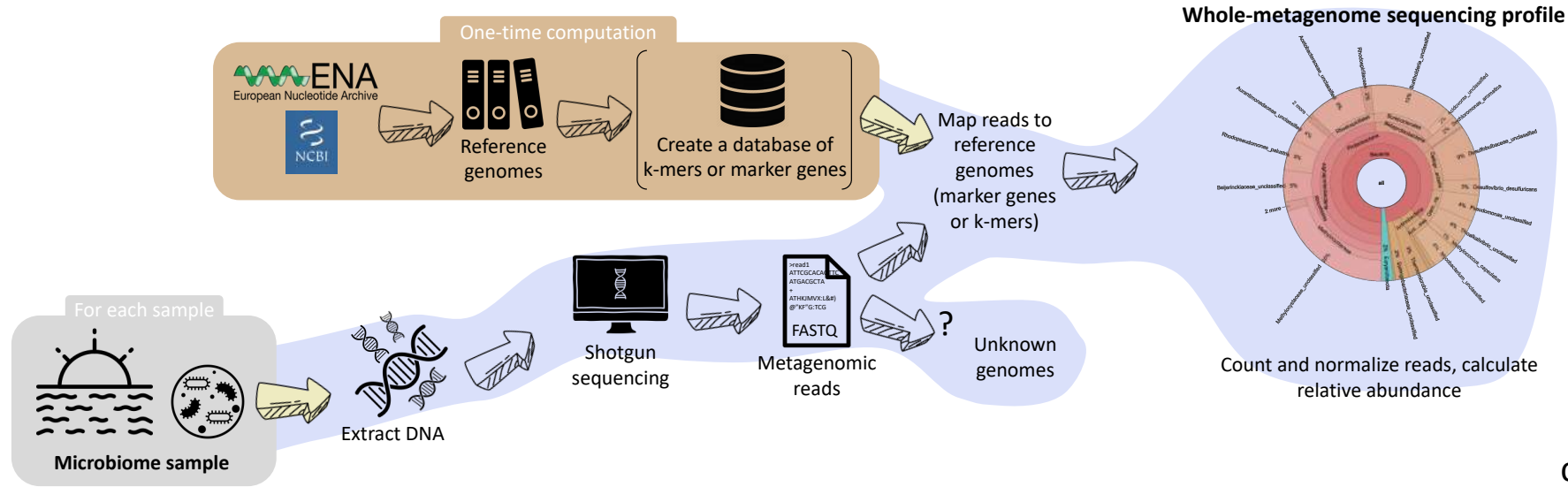


1. Introduction - Meta[gen | transcript]omics



*Metagenome-assembled genomes (MAGs)

2. Taxonomic profiling of metagenomes



Main strategies

- Use of universal marker genes
- Use of k-mers of reference genomes
- Use of clade-specific marker genes

universal	k-mers	clade-specific
mOTUs 2 [1]	mOTUs 1 [2]	Kraken [3]
		MetaPhlan2 [4]
		16S OTUs

[1] Milanese et al., *Nat Commun*, 2019
 [2] Sunagawa et al., *Nat Methods*, 2013
 [3] Wood et al., *Genome Biol*, 2014
 [4] Truong et al., *Nat Methods*, 2015

3. mOTUs: protein-coding single-copy marker gene-based Operational Taxonomic Units

Molecular definition[1,2]

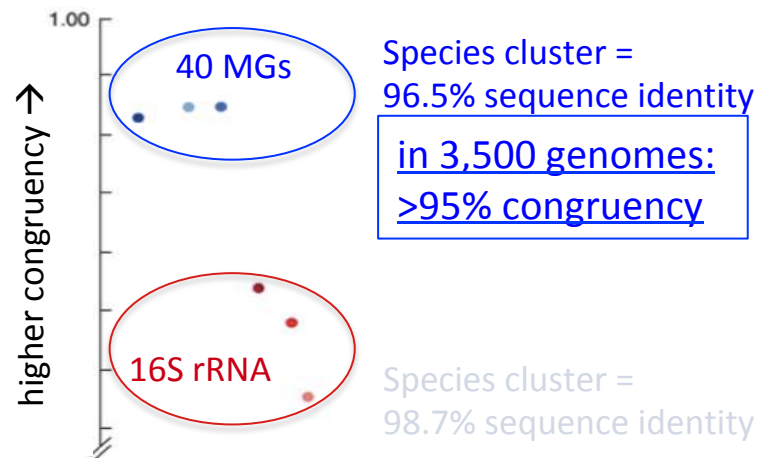
16S rRNA (single marker gene)

- 16S OTUs (97%) can include several species even from different genera

40 aa-coding single-copy marker genes (MG)

- universal across domains of life single-copy and low rate of HGT

Define taxonomic congruency-maximized clustering cutoffs



Use for metagenomics [3]

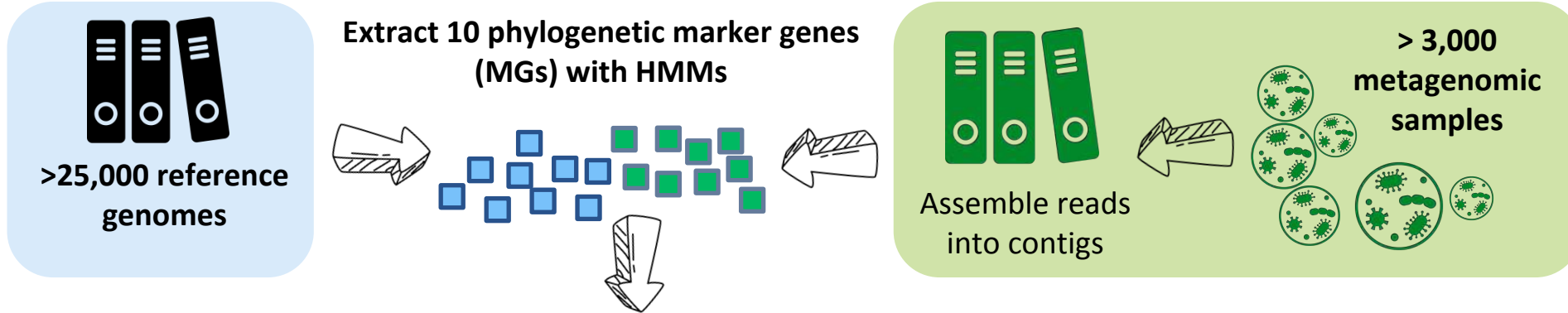
MG-based OTUs: mOTUs

10 out of the 40 MGs → good candidates

COG0012	Ribosome-binding ATPase YchF, GTP1/OBG family
COG0016	Phenylalanyl-tRNA synthetase alpha subunit
COG0018	Arginyl-tRNA synthetase
COG0172	Seryl-tRNA synthetase
COG0215	Cysteinyl-tRNA synthetase
COG0495	Leucyl-tRNA synthetase
COG0525	Valyl-tRNA synthetase
COG0533	tRNA A37 threonylcarbamoyltransferase TsaD
COG0541	Signal recognition particle GTPase (Ffh)
COG0552	Signal recognition particle GTPase (FtsY)

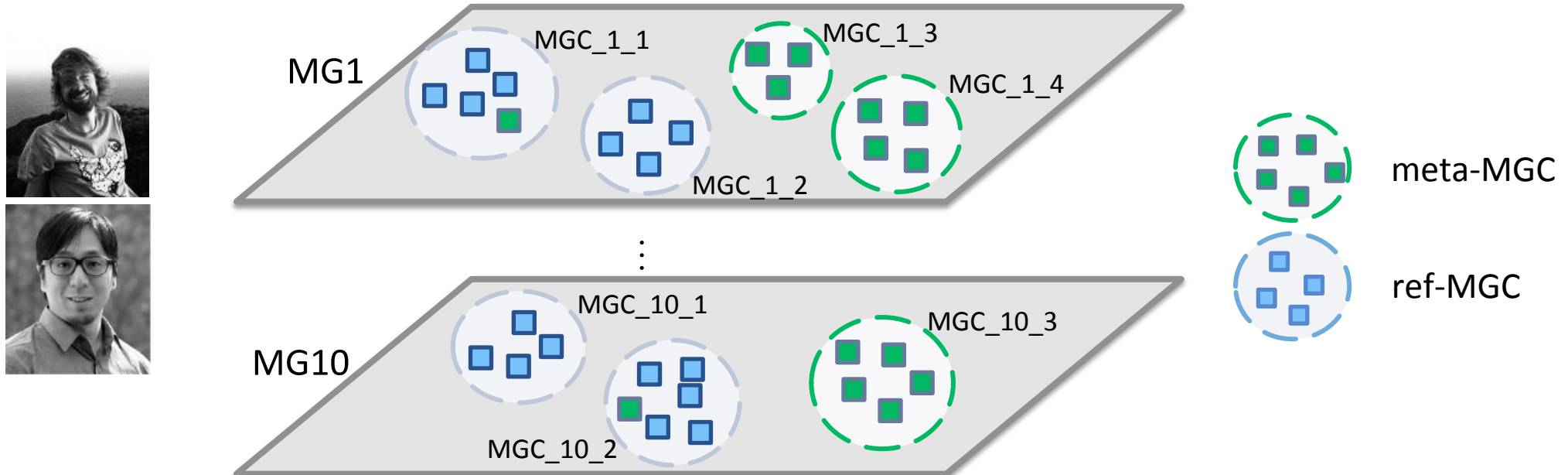
3. mOTUs: protein-coding single-copy marker gene-based Operational Taxonomic Units

Step 1



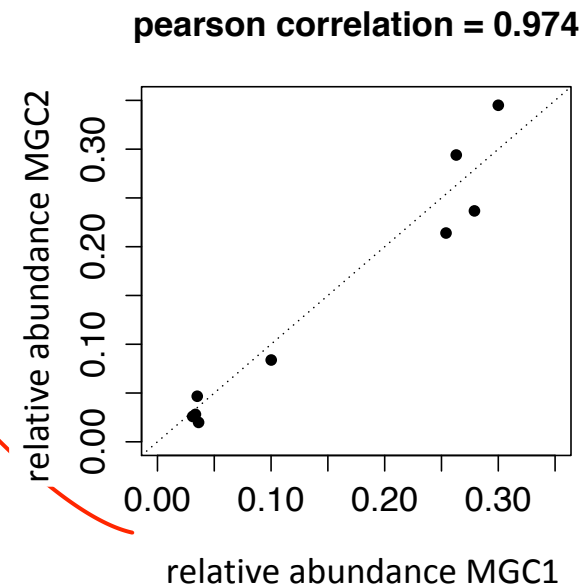
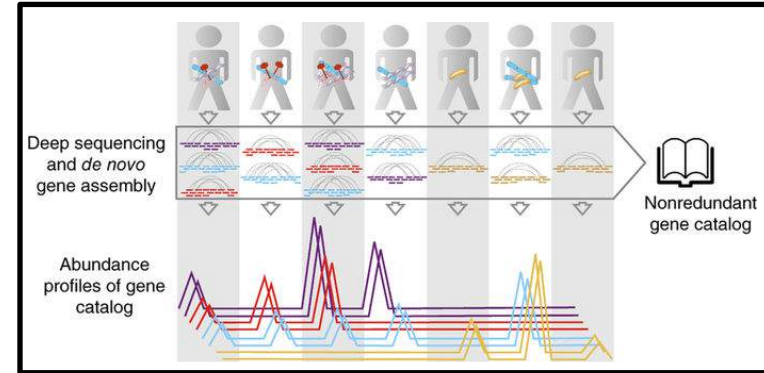
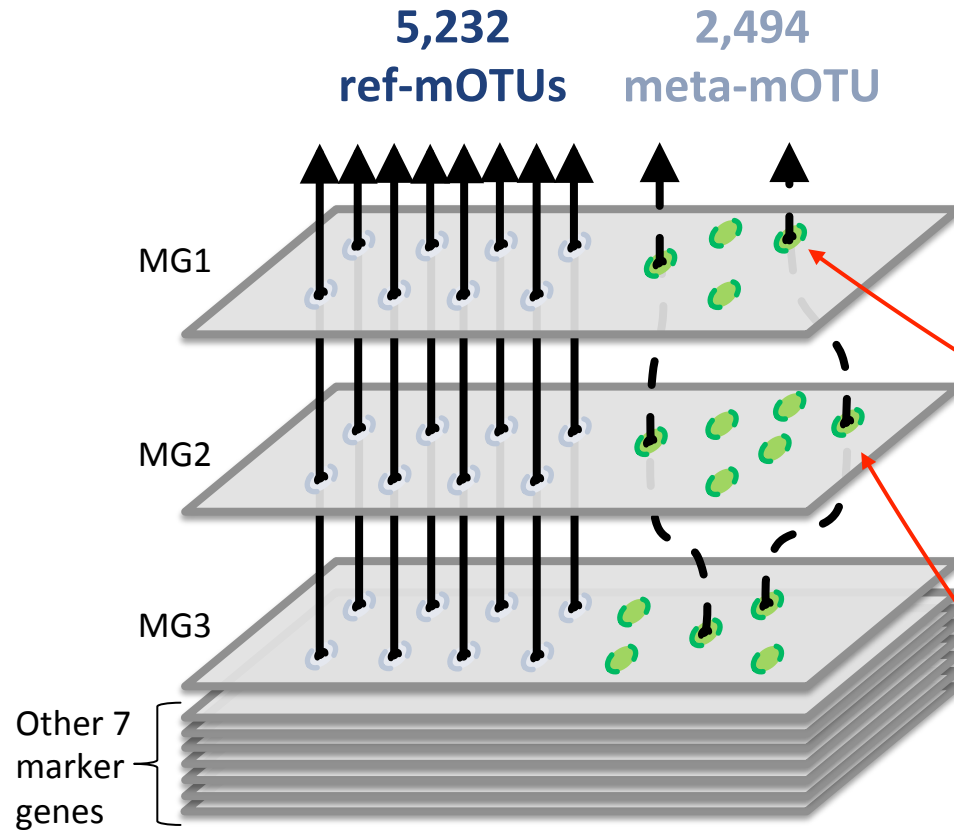
Step 2

Cluster similar gene sequences to create marker gene clusters (MGCs), independently for each MG (MG1 to MG10)



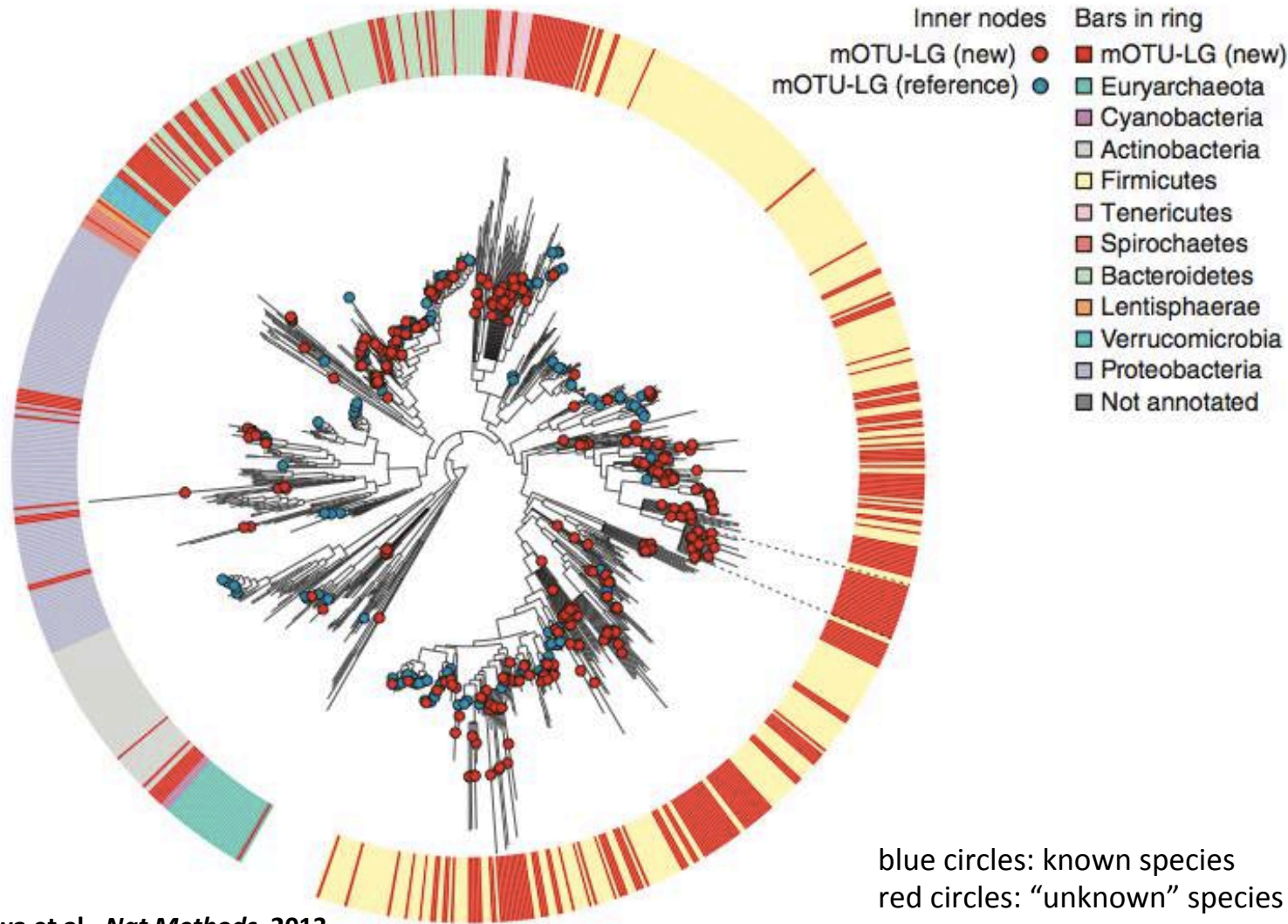
3. mOTUs: protein-coding single-copy marker gene-based Operational Taxonomic Units

Need to link MGCs from different marker genes



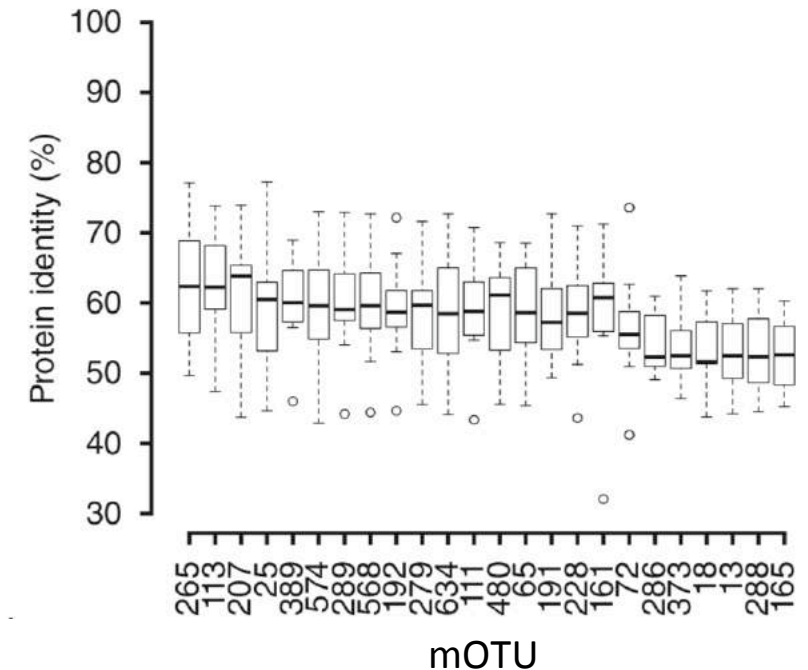
3. mOTUs: protein-coding single-copy marker gene-based Operational Taxonomic Units

“Unknown” species mapped onto prokaryotic Tree of Life



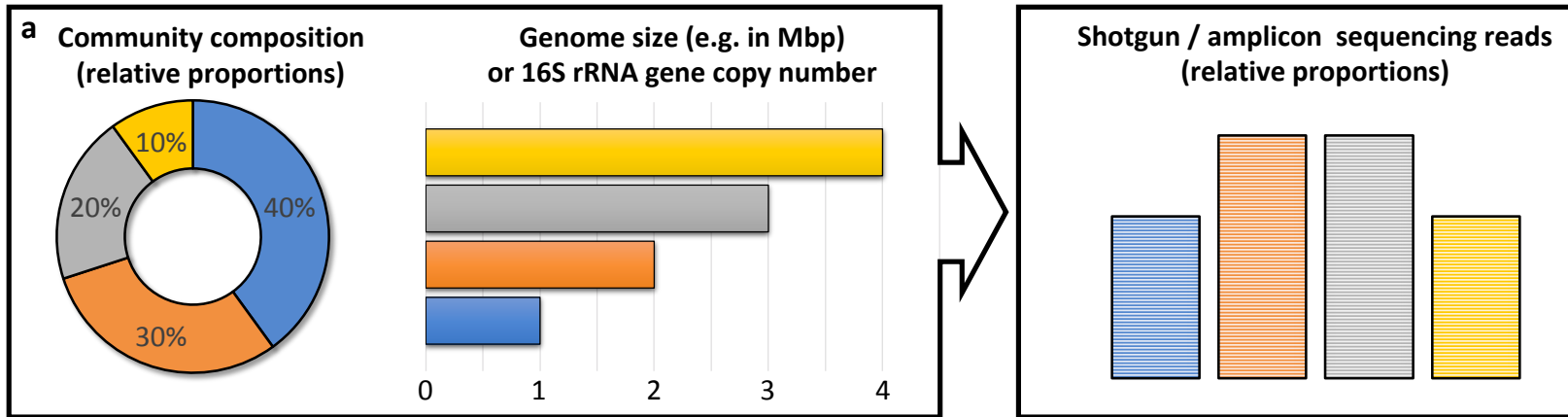
Taxonomic novelty

mOTUs without phylum level annotation:
~ 60% BLASTp identity to known species



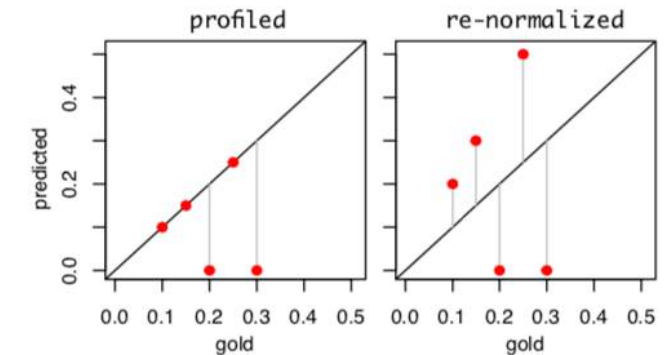
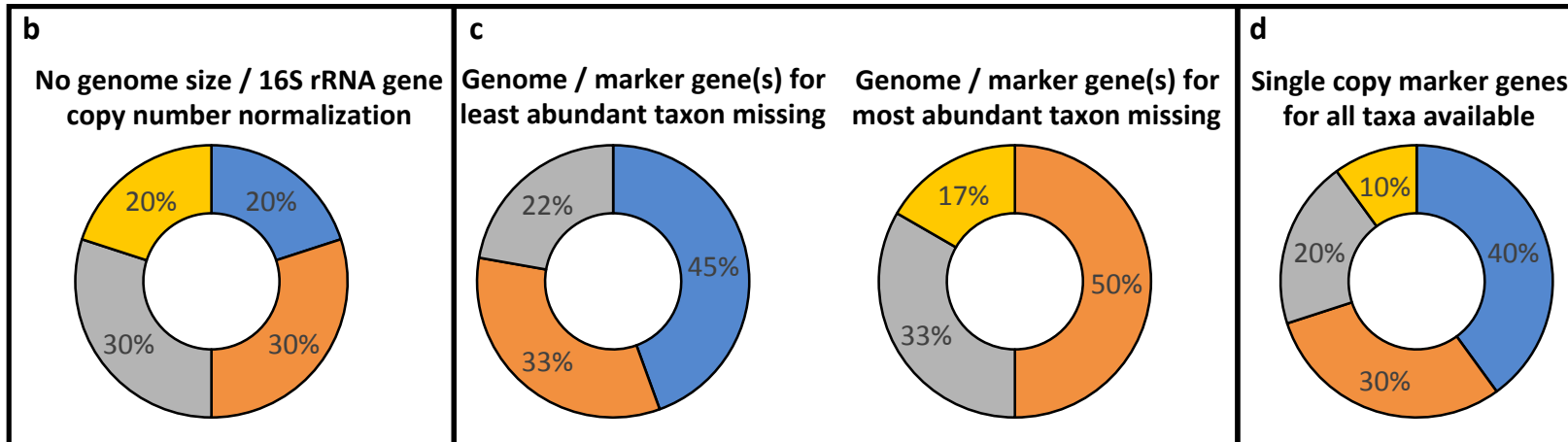
4. Profiling with mOTUs

Issue of compositionality / re-normalization of relative abundance data



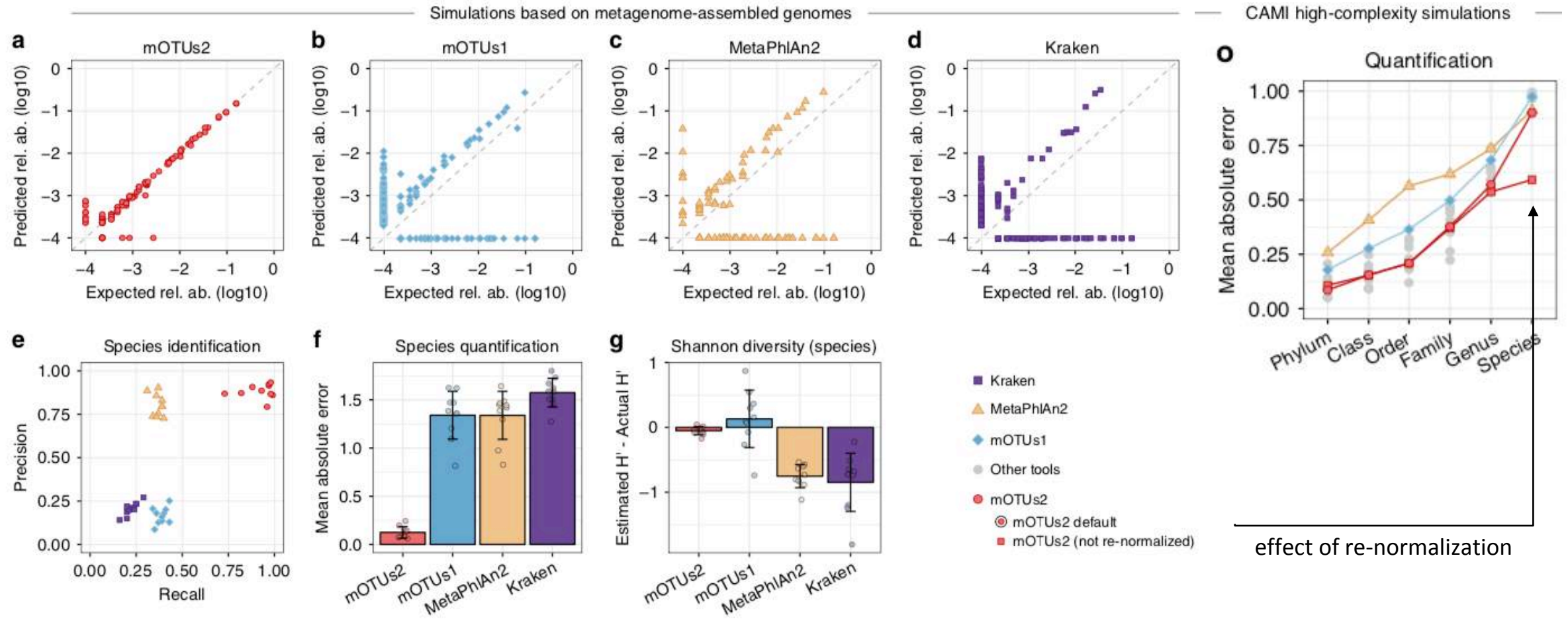
Effect of re-normalization (example)

Taxon	gold	profiled	re-normalized
Species 1	0.3	0	0
Species 2	0.25	0.25	0.5
Species 3	0.2	0	0
Species 4	0.15	0.15	0.3
Species 5	0.1	0.1	0.2
unknown		(0.5)	



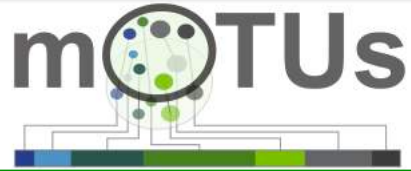
→ mOTUs can estimate the proportion of unknown taxa, providing less biased abundance profiles

4. Profiling with mOTUs



5. Summary + Outlook

Reference genome-independent taxonomic profiling of microbial communities



<http://motu-tool.org>

Profiles of 7000 genes, 3100 clusters of strains at higher sensitivity and precision than other taxonomic tools

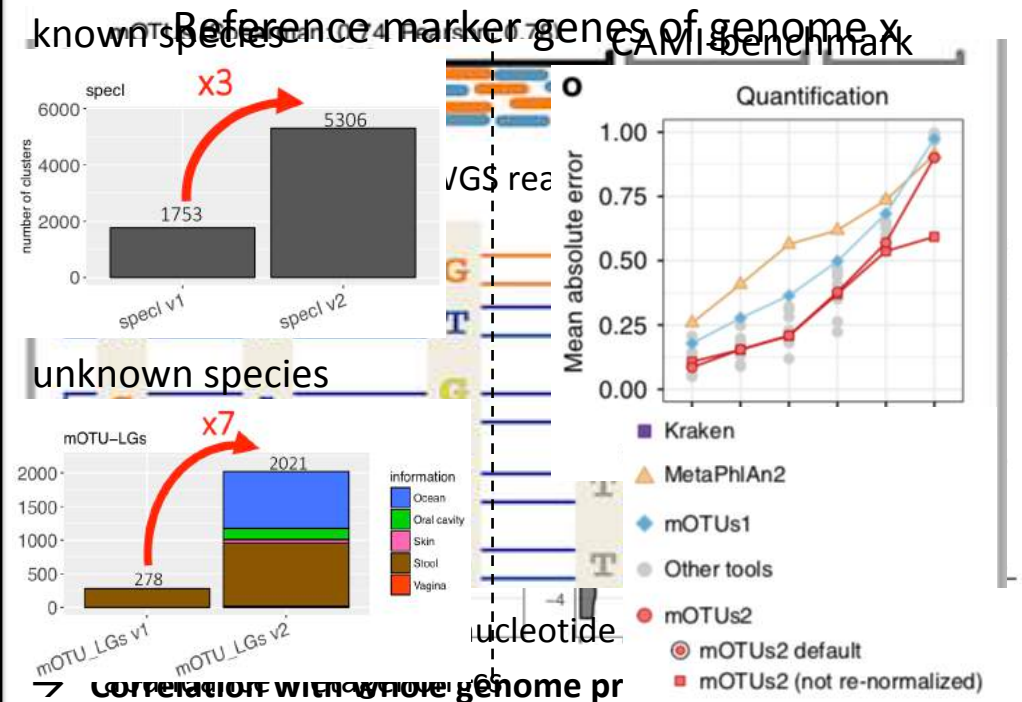
Use of universal, single copy marker genes from reference genomes and *denovo* assembled metagenomes

Ideal features of marker genes

- Universal present in all organisms
- Single copy quantitative cell abundances - unlike 16S/18S rRNA
- Housekeeping constitutively expressed
- Low HGT vertical descent → phylogeny
- Protein-coding 20,000 bp (rather than 1,500)

Original: Sunagawa et al., Nat Methods, 2013; New: Milanese et al., Nat Commun, 2019

‘How similar are populations?’



5. Summary + Outlook

Reference genome-independent taxonomic profiling of microbial communities



<http://motu-tool.org>

Outlook

Use of universal, single copy marker genes from reference genomes and *denovo* assembled metagenomes

Ideal features of marker genes

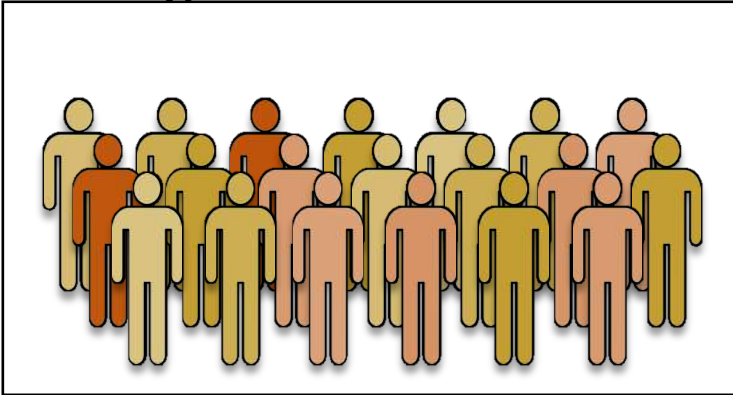
- Universal present in all organisms
- Single copy quantitative cell abundances
 - unlike 16S/18S rRNA
- Housekeeping constitutively expressed
- Low HGT vertical descent → phylogeny
- Protein-coding 20,000 bp (rather than 1,500)

- **Extension to other biomes**
 - mouse
 - soil
 - lakes
 - air
- **Link taxa to functions through genomes & MAGs**
- **Use as classification tool for new genomes**
- **Democratize tool through custom extendibility**
 - tutorial online now!

MGWAS: microbiome-wide genome wide association study

Microbiome-wide GWAS: linking phenotypes to microbiome features

Phenotype



Readout



Features

- Species diversity / richness
- Presence/absence/abundance of:
 - viral + microbial taxa
 - genes, transcripts, functions
 - methylation patterns, etc.

Analogous to GWAS, microbiome-features can be linked to

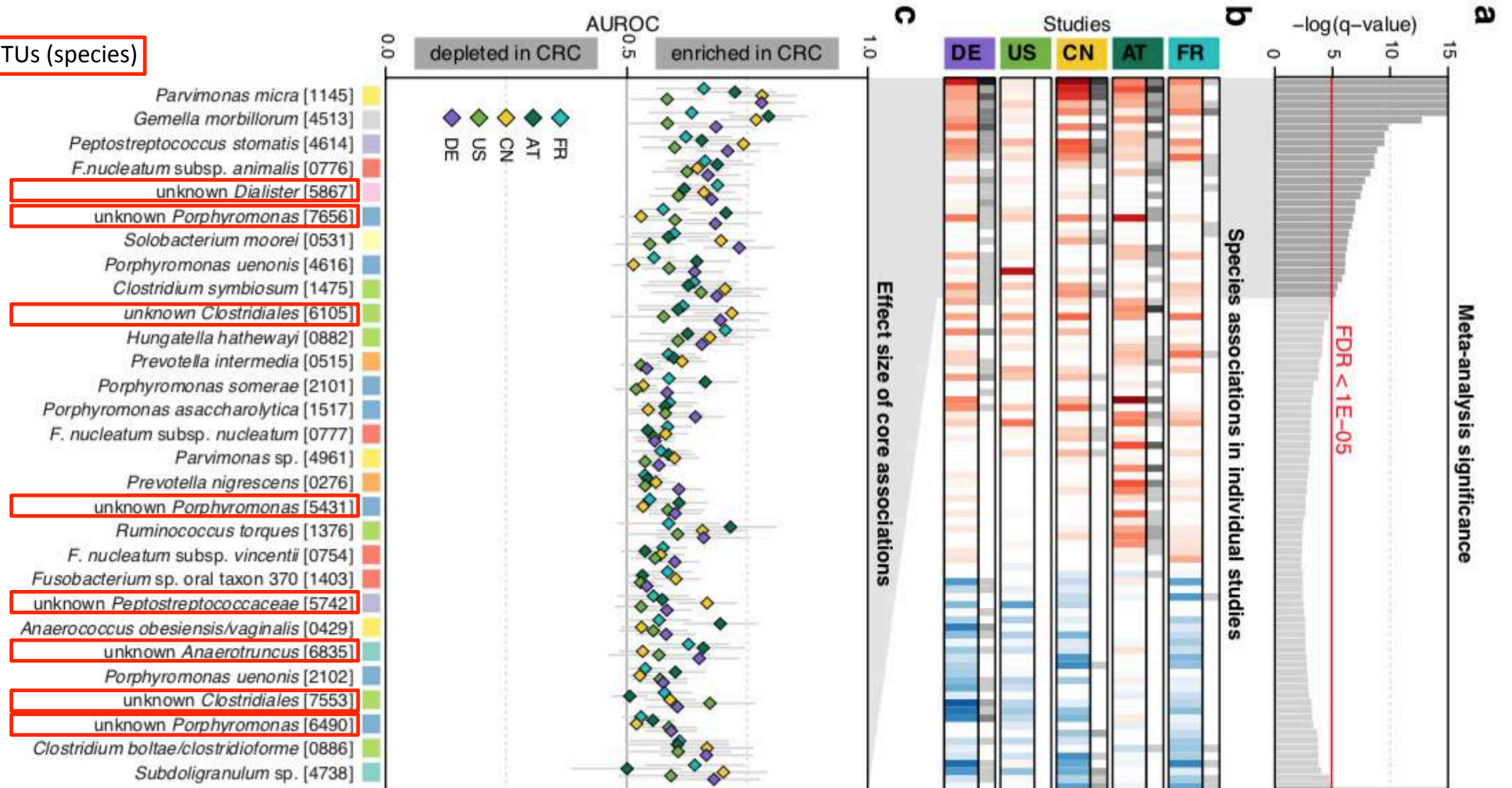
- groups of individuals and/or health states
- differential response to drugs (or nutrition)

Examples

- lean vs obese, **healthy vs colorectal cancer**
European patent: EP14172205.8
- cardiac drug digoxin inactivation by *E. lenta* strains
- role of gut microbiome in cancer immune therapy

MGWAS: microbiome-wide genome wide association study

'unknown' mOTUs (species)



6. Workshop agenda

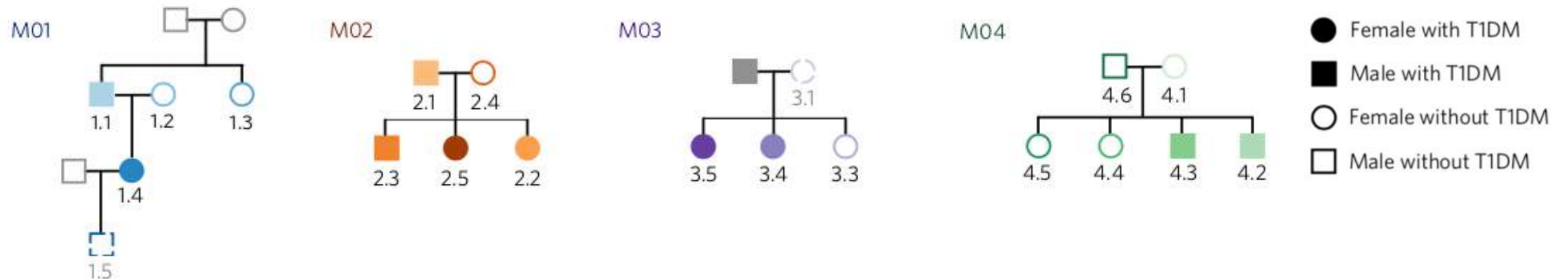
Example data used during the workshop



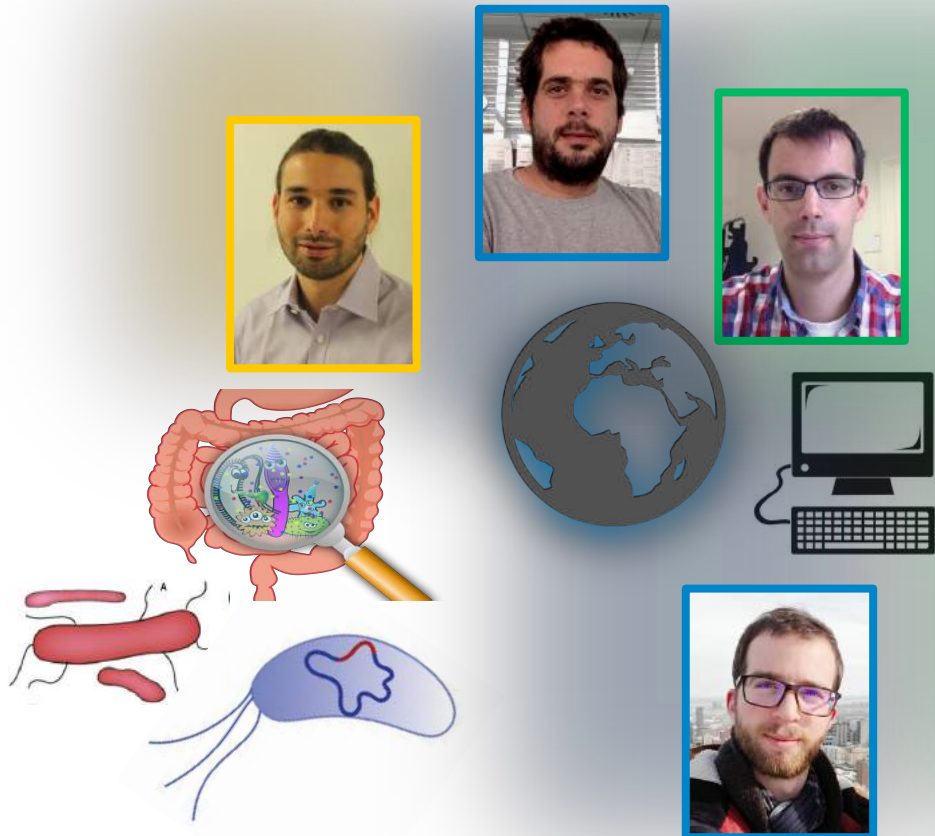
Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes

Anna Heintz-Buschart^{1*}, Patrick May¹, Cédric C. Laczny¹, Laura A. Lebrun¹, Camille Bellora²,
Abhimanyu Krishna¹, Linda Wampach¹, Jochen G. Schneider^{1,3,4}, Angela Hogan²,
Carine de Beaufort^{1,5} and Paul Wilmes^{1*}

Metagenomic and metatranscriptomic data from stool samples from 4 families (including time-series)



Acknowledgements



Sunagawa lab

Miguel Cuenca
Guillem Salazar
Hans Ruscheweyh
Lucas Paoli
Chris Field

Zeller group (EMBL)

Alessio Milanese



Daniel Mende (U Hawaii)

Bork group (EMBL)

Finn group (EBI)