Assembly-free shotgun metagenomic profiling

15 November 2022

Alessio Milanese

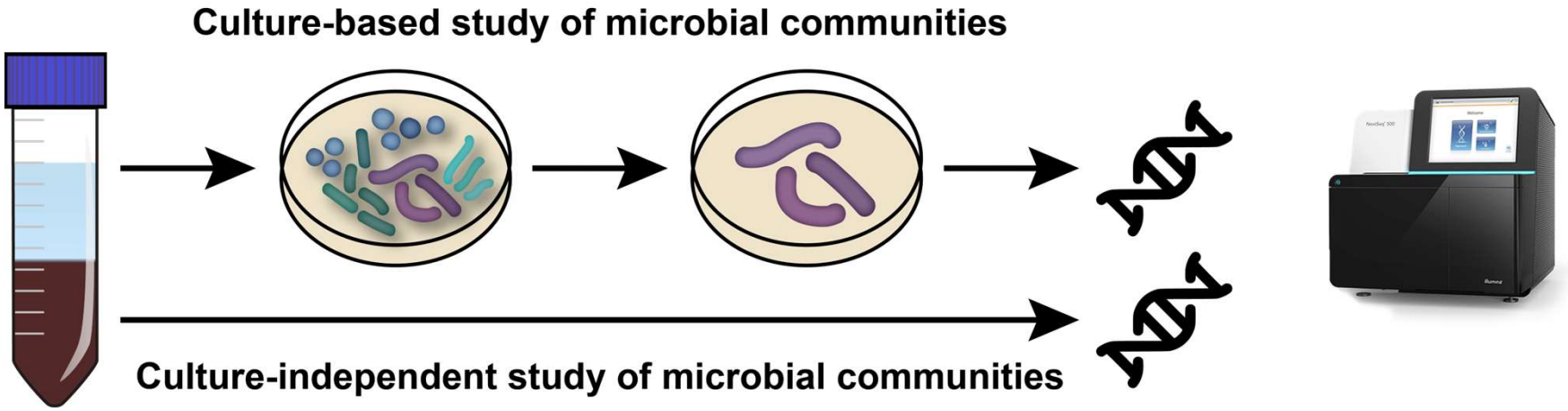
Postdoc in the Sunagawa group (ETH)

alessio_mila

Aims of this lecture

- Introduction of the **taxonomic profiling problem**
- **Conceptual overview** over approaches for taxonomic profiling
- **Sources of bias** in taxonomic profiling
- Introduction of the **mOTU profiling tool**
- **Comparative evaluation** of existing tools on community benchmarks
- How to account for **unknown species**, i.e. ones for which an isolate genome sequence is not available

Culture-independent sequencing of microbial communities

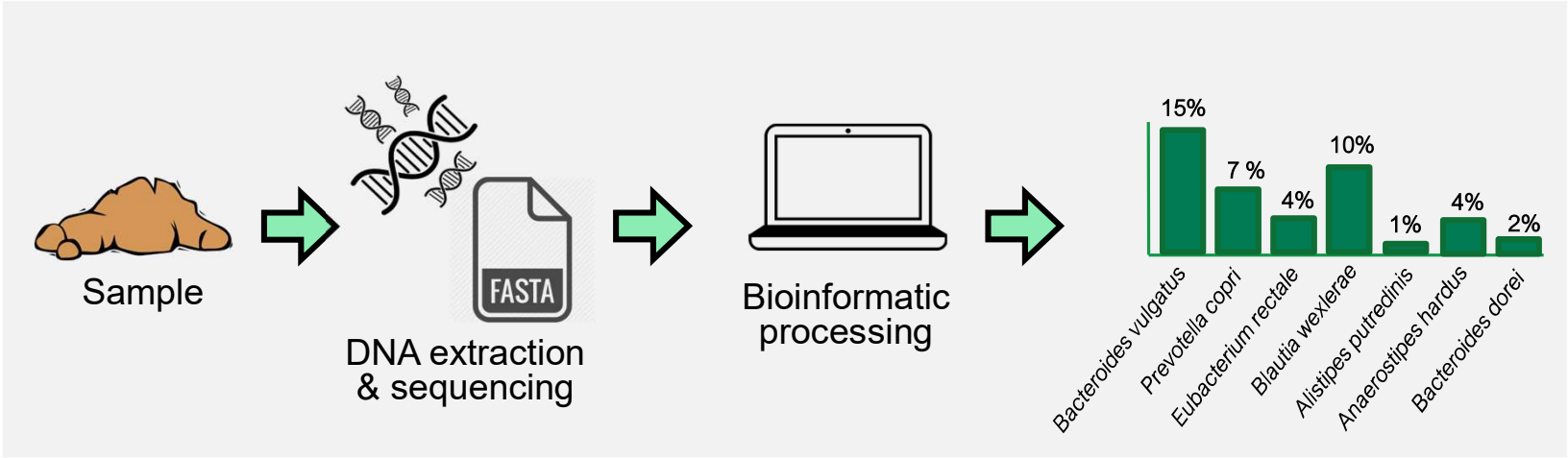


- Who's there?
  - ➢ Taxonomic profiling

- Cave: DNA extraction and bioinformatics analysis are not standardized (technical biases)
- Key limitation: technology provides only snapshots, spatial information lost, causalities unclear

[Costea et al., *Nat. Biotechnol.*, 2017]

Taxonomic profiling – what is it?

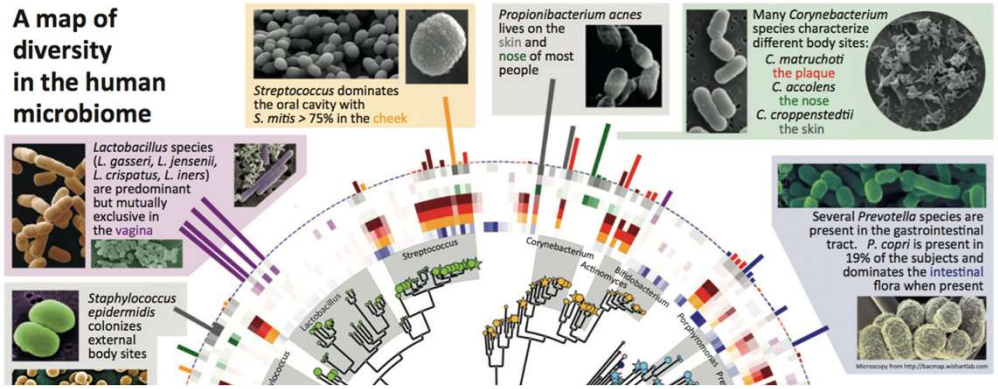


Taxonomic Profiling:
Estimate relative cell counts in a microbiome sample from metagenomic sequencing

Taxonomic profiling – why it is important?

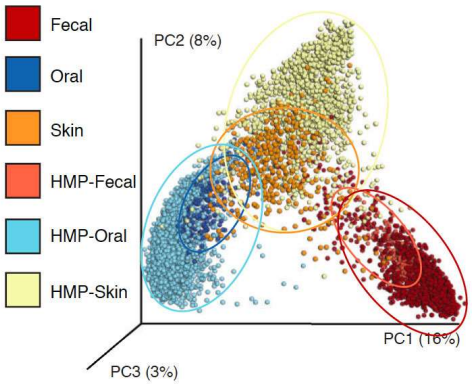
Taxonomic analysis is fundamental to the analysis of microbial communities

Describing the microbial community under study



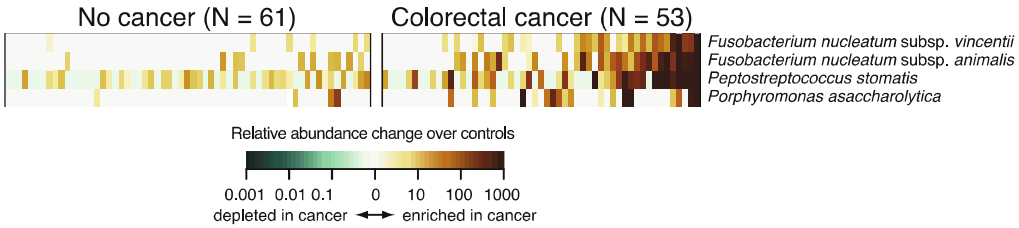
[Morgan et al., *Trends in Genetics*, 2013]

Comparing different microbial communities



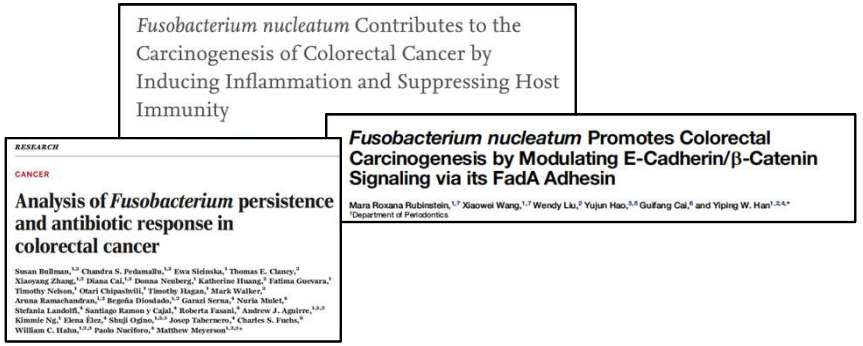
[McDonald et al., mSystems, 2018]

Correlating environm. or host features to microbes

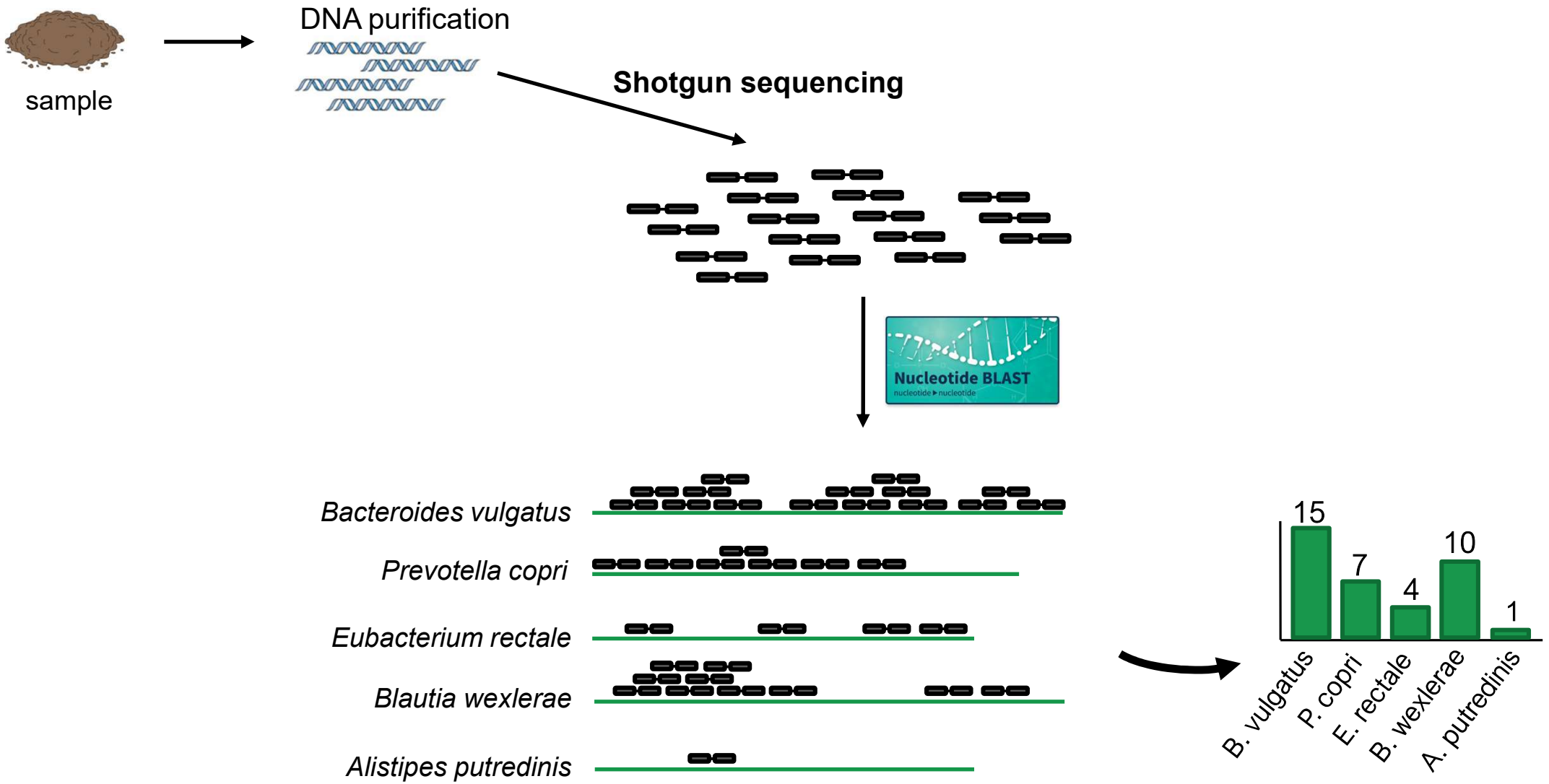


[Zeller et al., MSB, 2014]

Comparing findings to literature



Taxonomic profiling – how it is done?

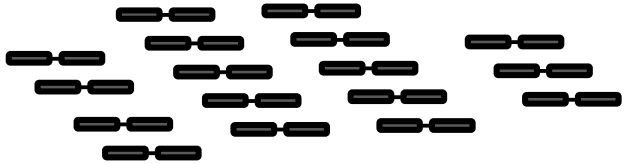


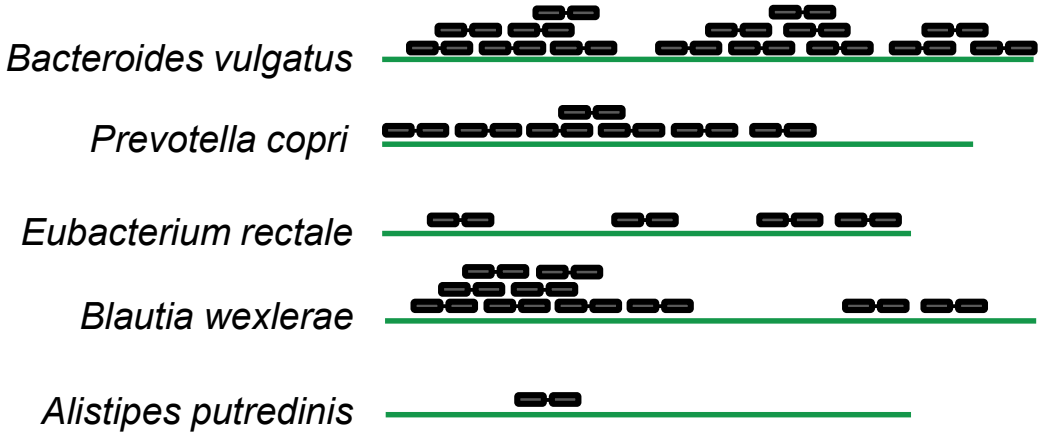
Taxonomic profiling – how it is done?

sample

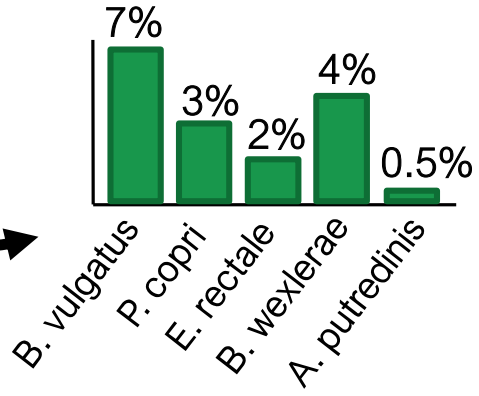
DNA purification

Shotgun sequencing





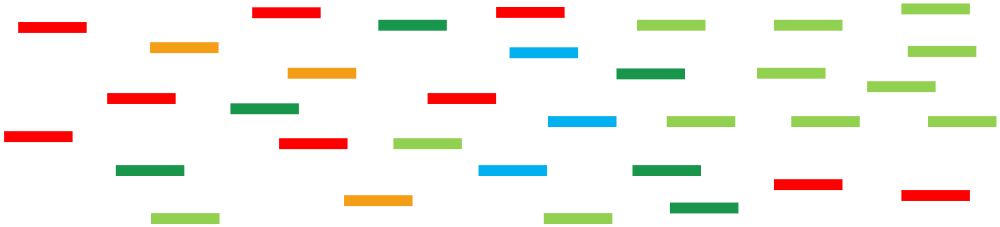
- Data is compositional



Taxonomic profiling approaches – whole-genome mapping

Environmental sample

Shotgun sequencing



- DNA extraction bias
- sequencing biases
- sampling noise

Taxonomic profiling approaches – whole-genome mapping

Environmental sample

Shotgun sequencing

- DNA extraction bias
- sequencing biases
- sampling noise

Taxonomic profiling approaches – whole-genome mapping

Environmental sample

Shotgun sequencing

- DNA extraction bias
- sequencing biases
- sampling noise

True taxonomic annotation

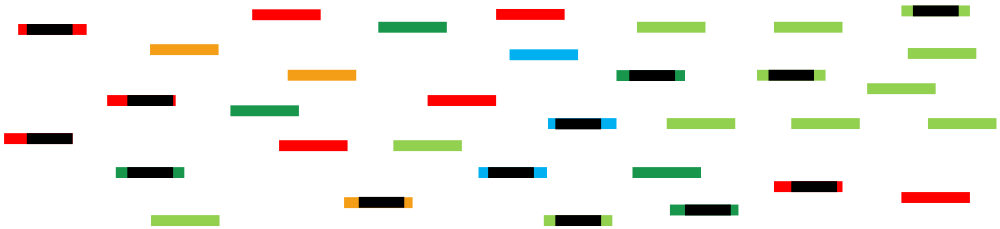
Estimated by whole-genome mapping

- genome size issue

Taxonomic profiling approaches – marker gene mapping

Environmental sample

Shotgun sequencing



- DNA extraction bias
- sequencing biases
- sampling noise

Taxonomic profiling approaches – marker gene mapping

Environmental sample

Shotgun sequencing

- DNA extraction bias
- sequencing biases
- sampling noise

Taxonomic profiling approaches – marker gene mapping

Environmental sample

Shotgun sequencing

- DNA extraction bias
- sequencing biases
- sampling noise

Taxonomic profiling approaches – marker gene mapping

Environmental sample

Shotgun sequencing

- DNA extraction bias
- sequencing biases
- sampling noise

True taxonomic annotation

Estimated by whole-genome mapping

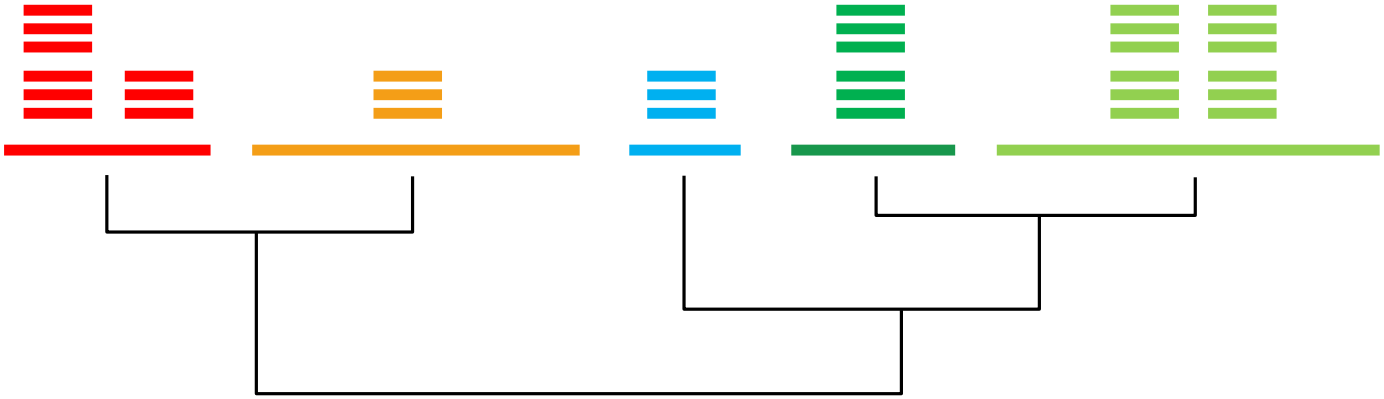
- genome size issue

Estimated by universal marker

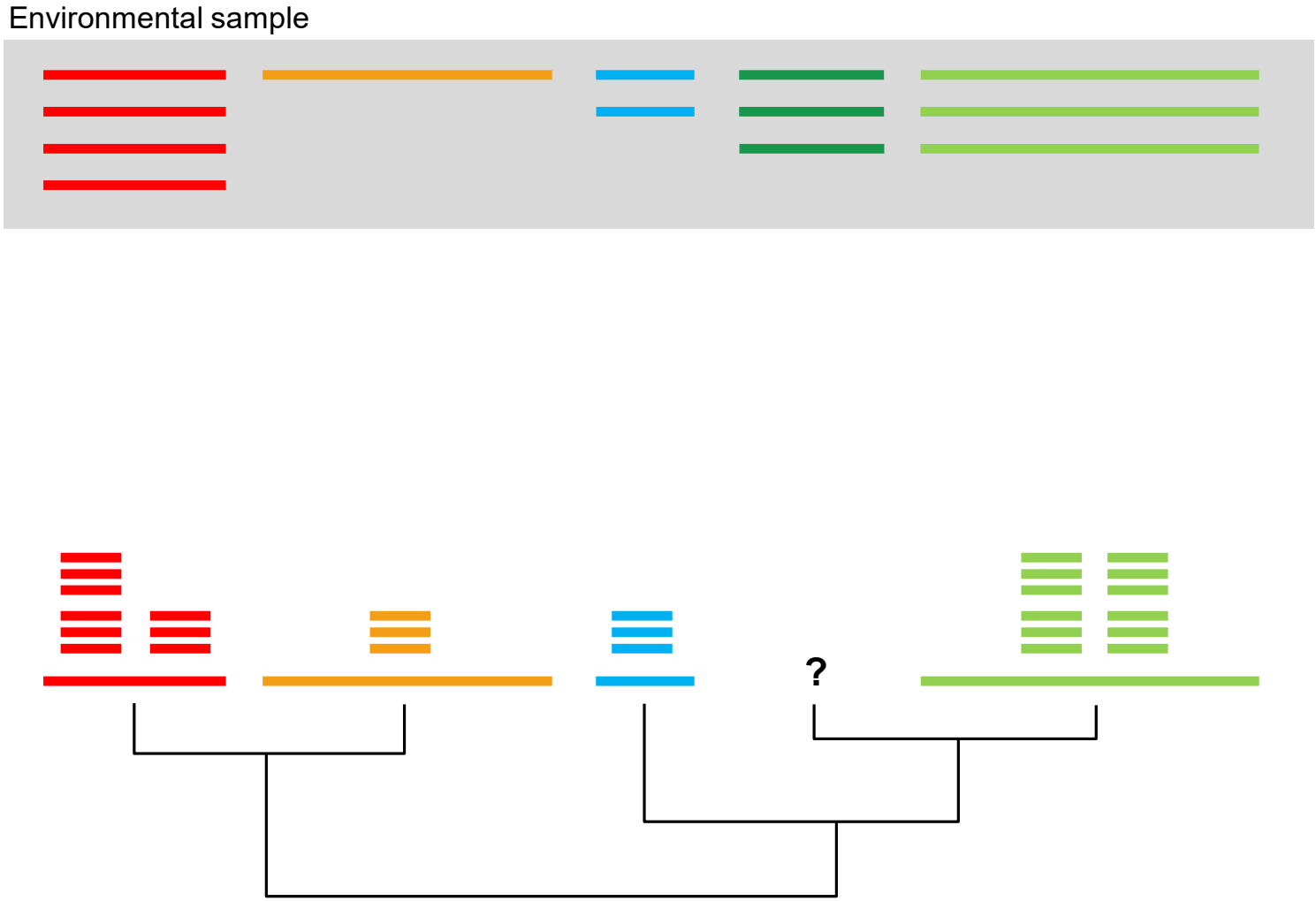
Taxonomic profiling – mapping reads to genomes

Environmental sample

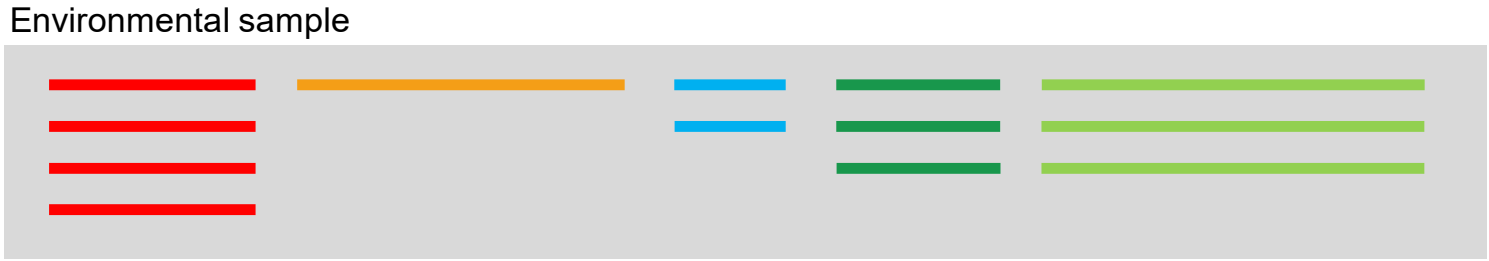
Shotgun sequencing

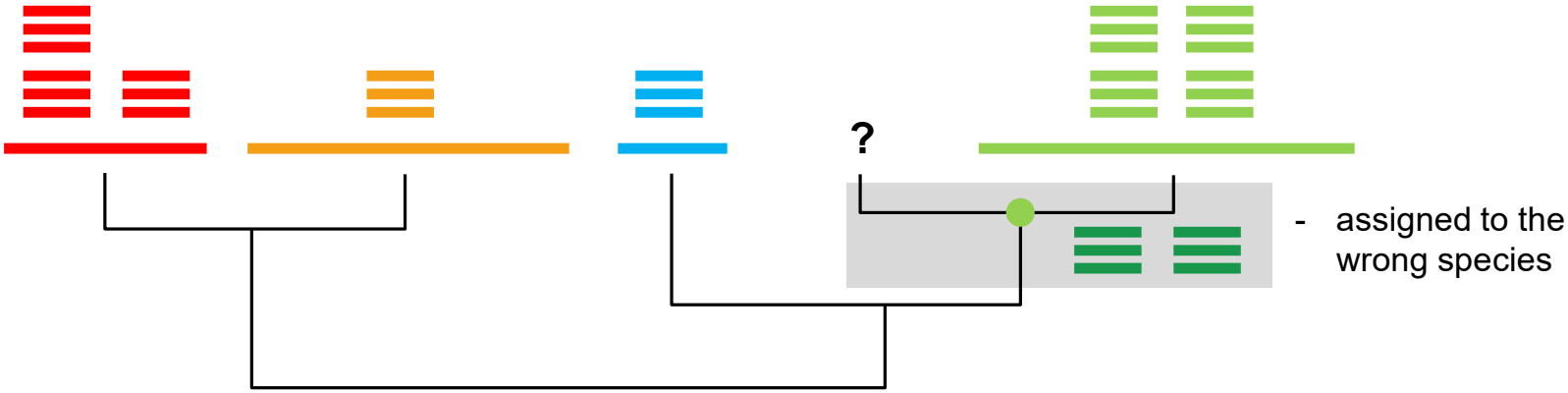


Taxonomic profiling – incomplete reference databases

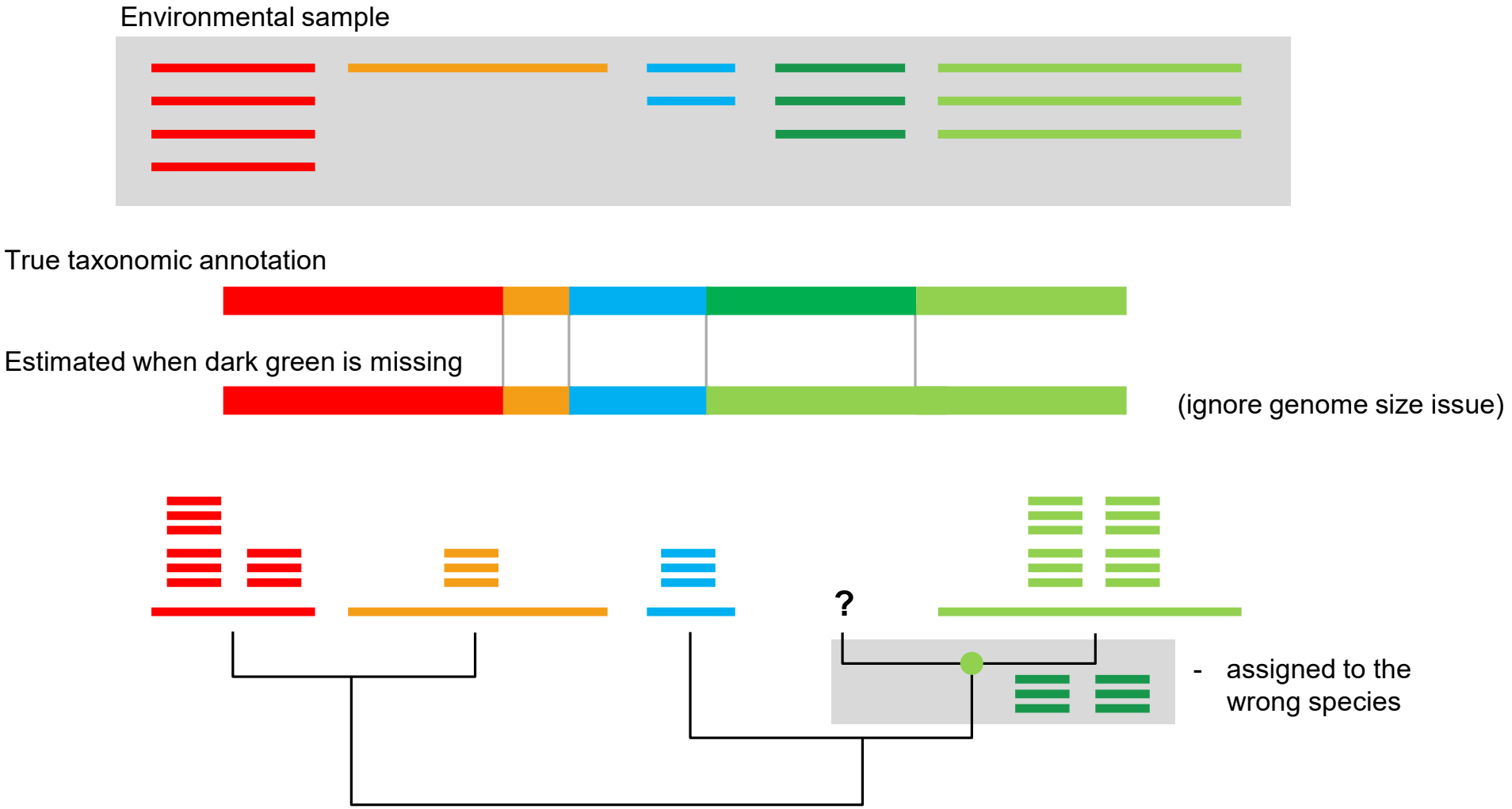


Taxonomic profiling – incomplete reference databases





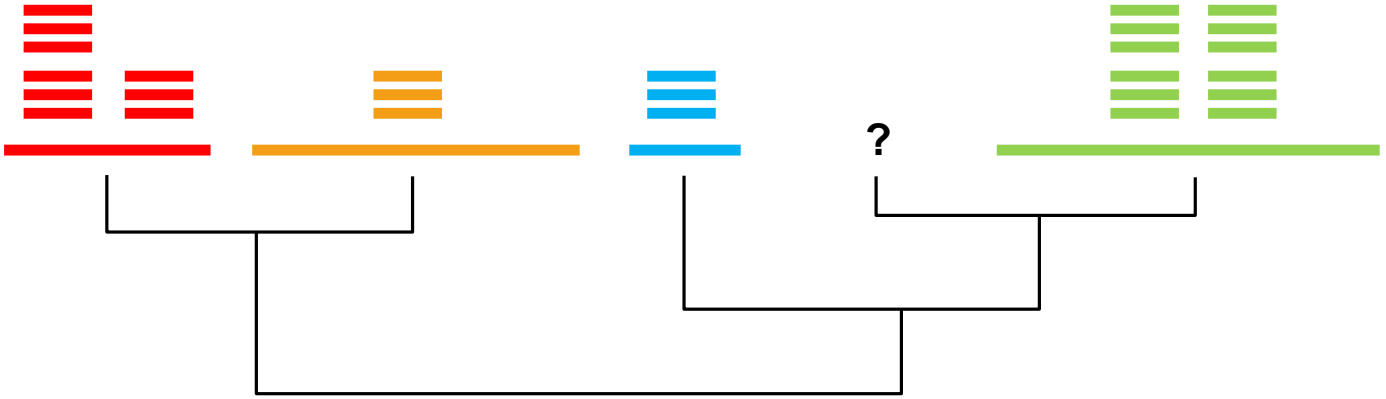
Taxonomic profiling – incomplete reference databases



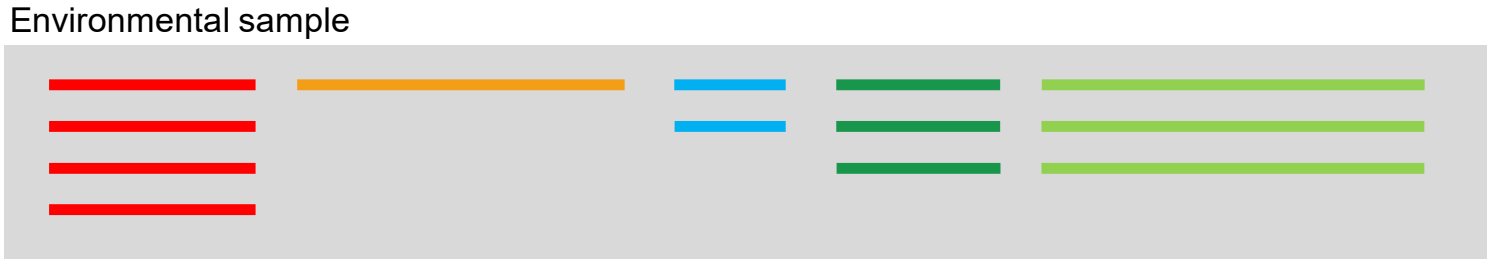
Taxonomic profiling – incomplete reference databases

Environmental sample

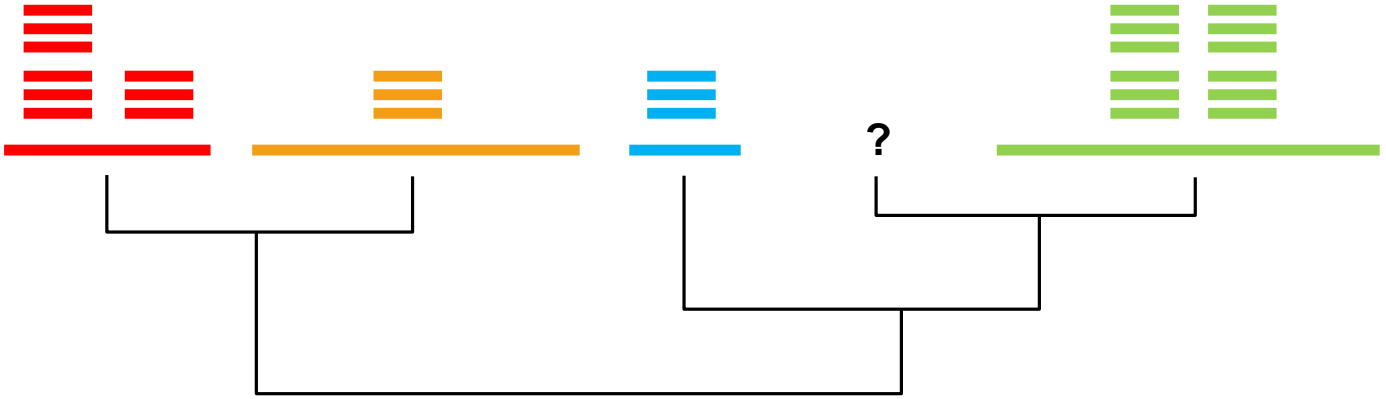
- ignore the reads



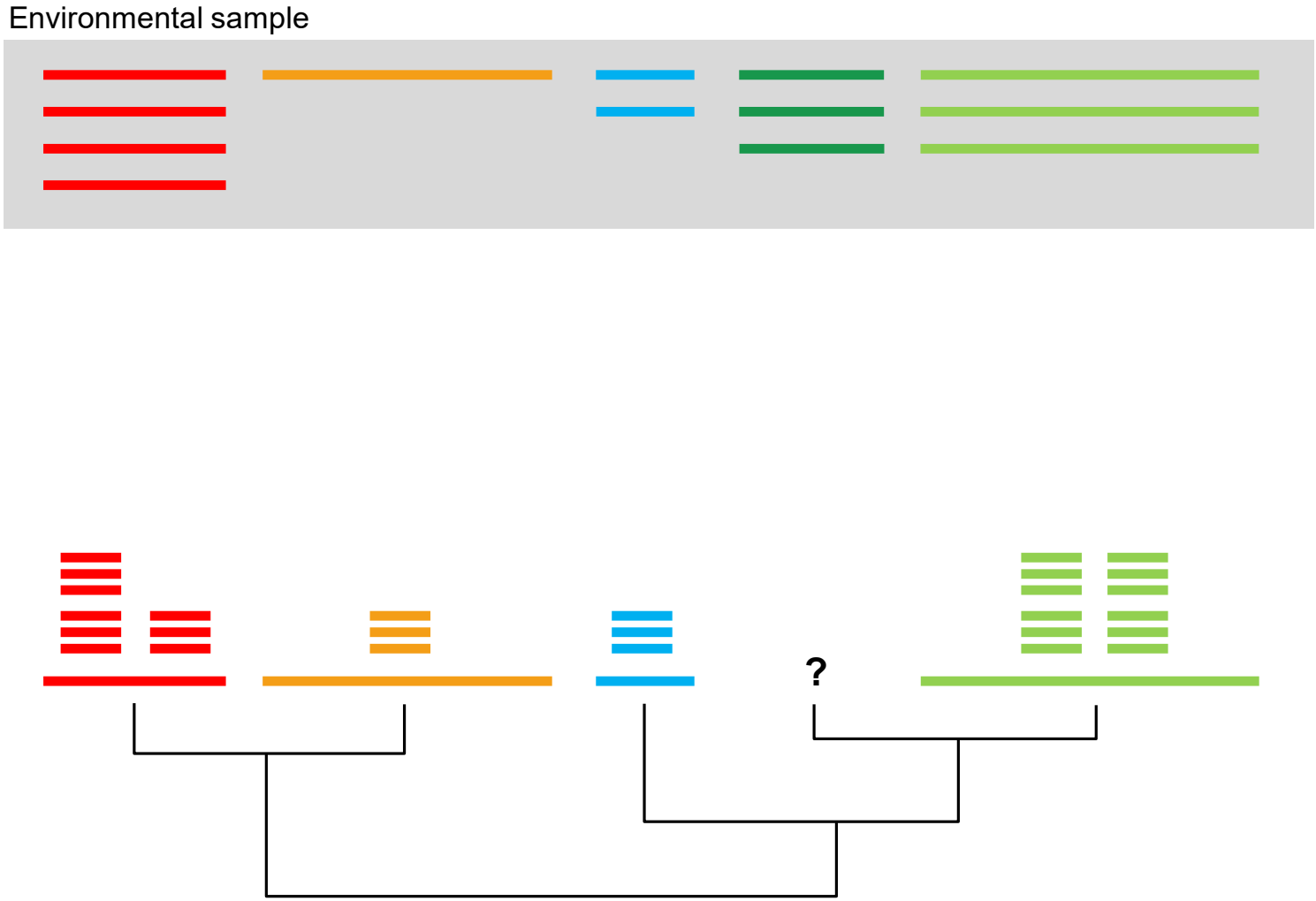
Taxonomic profiling – incomplete reference databases



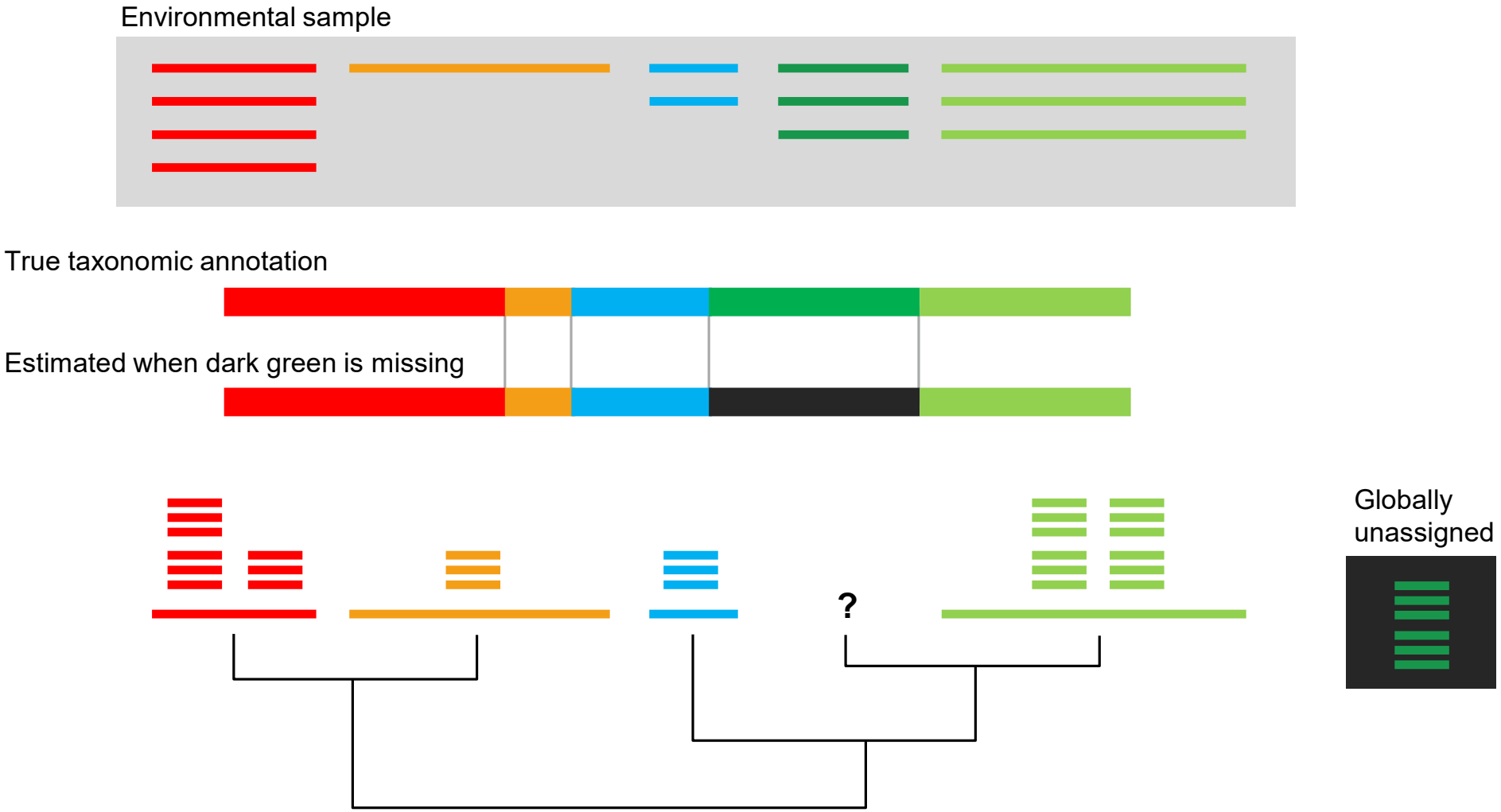
- ignore the reads



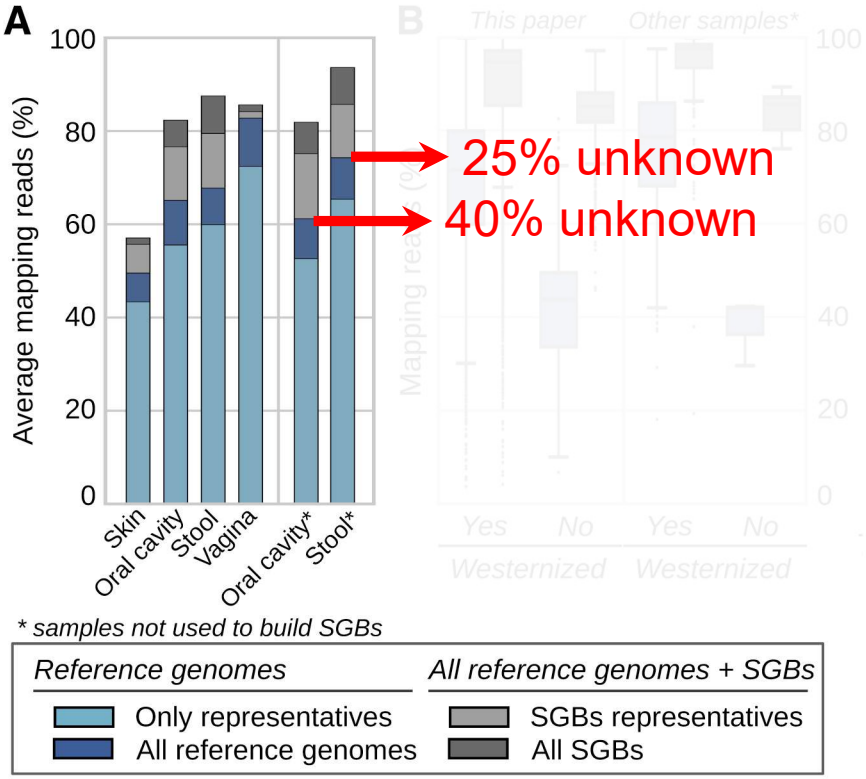
Taxonomic profiling – incomplete reference databases



Taxonomic profiling – incomplete reference databases

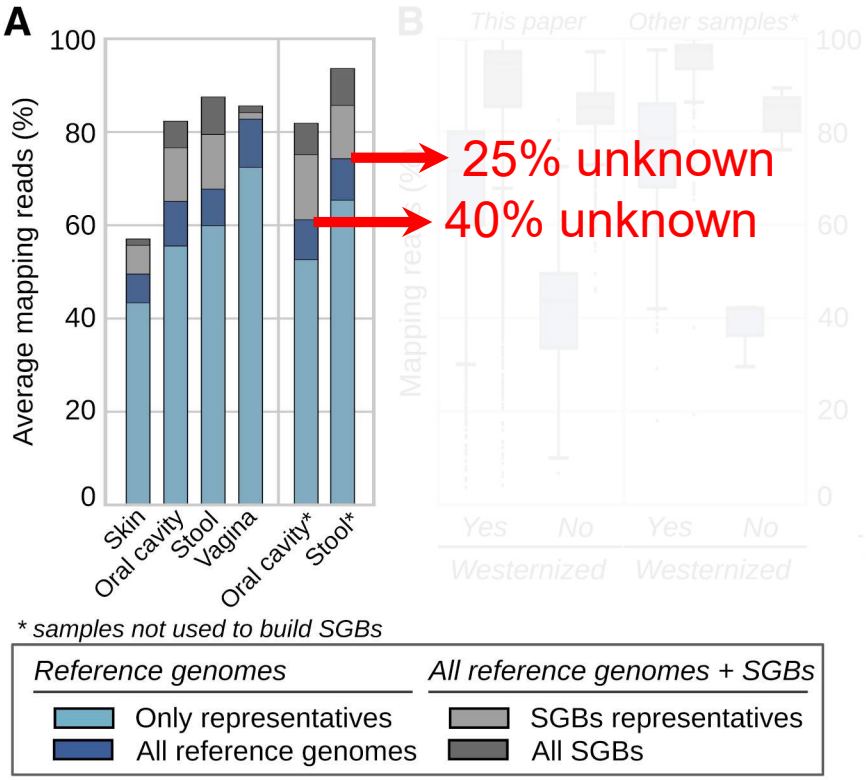


How much is unknown?

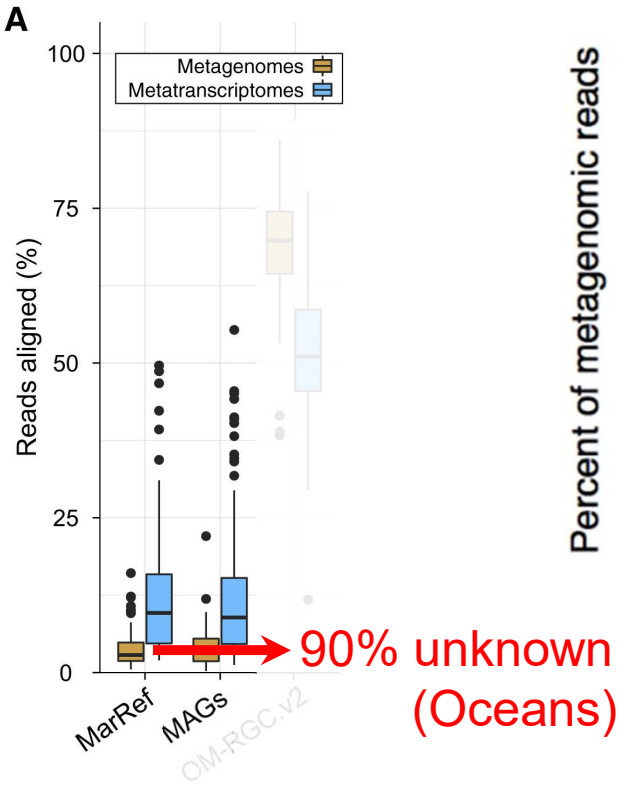


[Pasolli et al., *Cell*, 2019]

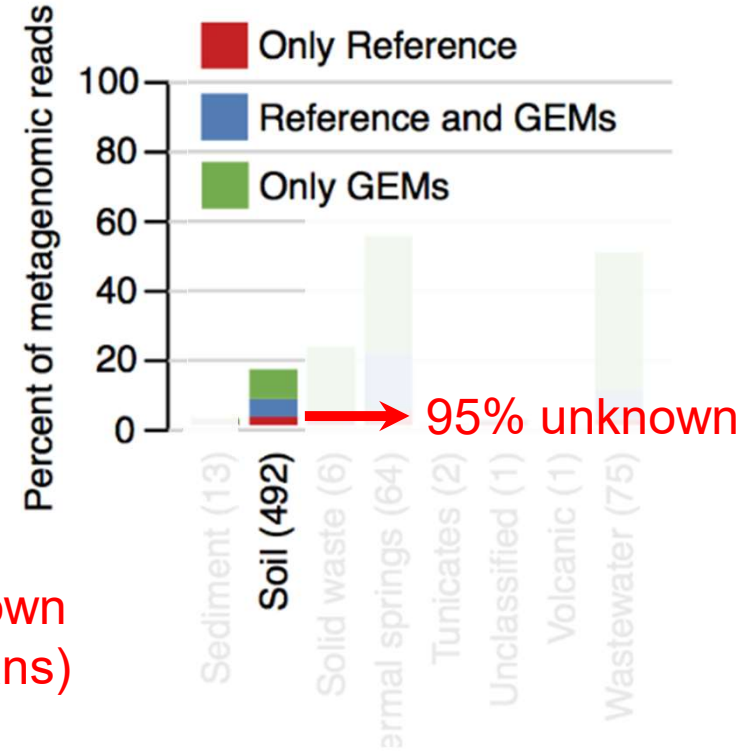
How much is unknown?



[Pasolli et al., *Cell*, 2019]



[Salazar et al., *Cell*, 2019]



[Nayfach et al., *Nat. Biotechnol.*, 2021]

Taxonomic profiling approaches – marker gene mapping

Environmental sample

Shotgun sequencing

- DNA extraction bias
- sequencing biases
- sampling noise

True taxonomic annotation

Estimated by whole-genome mapping

- genome size issue

Estimated by universal marker

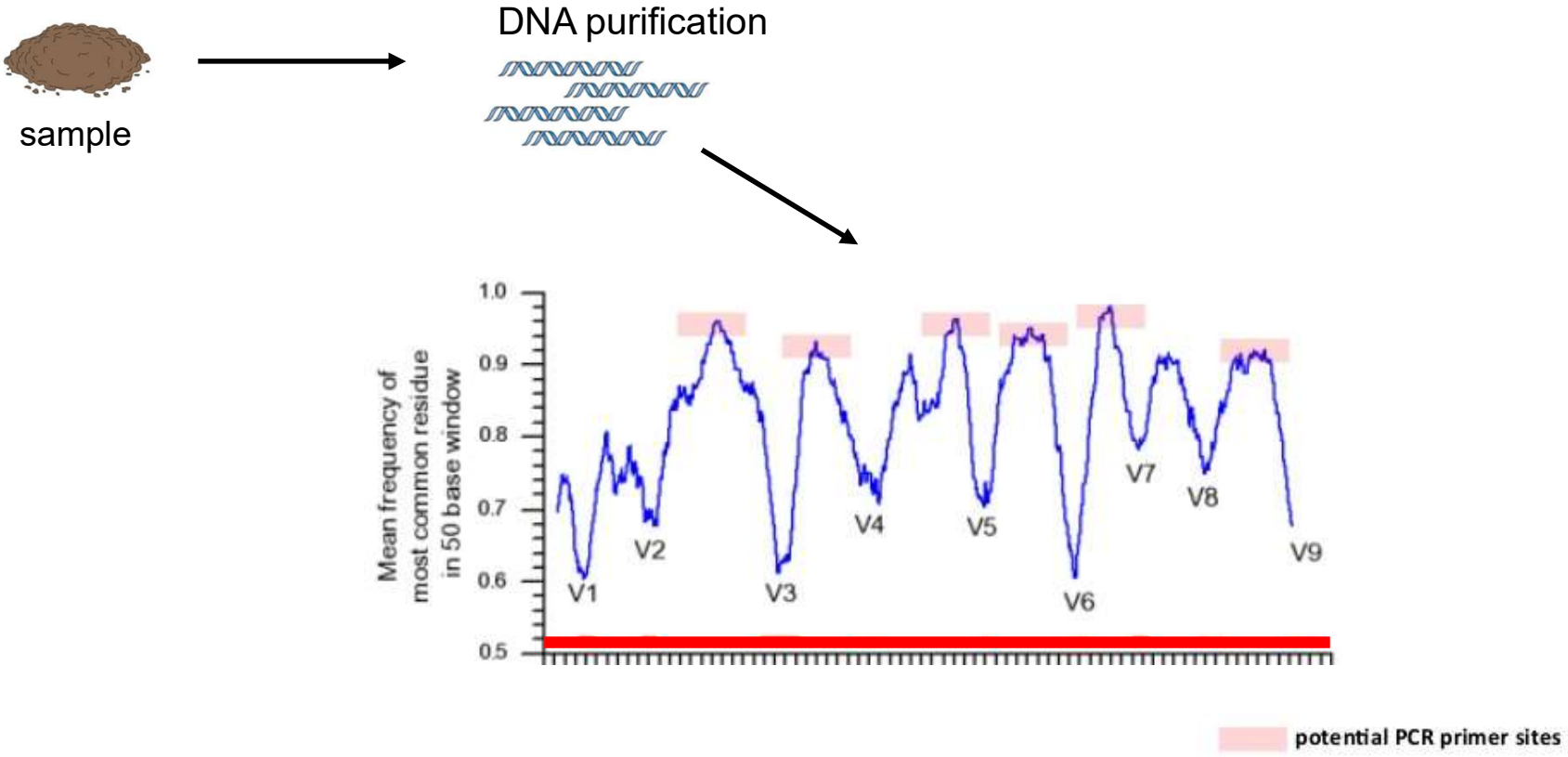
Taxonomic profiling approaches – 16S rRNA amplicon

sample

DNA purification

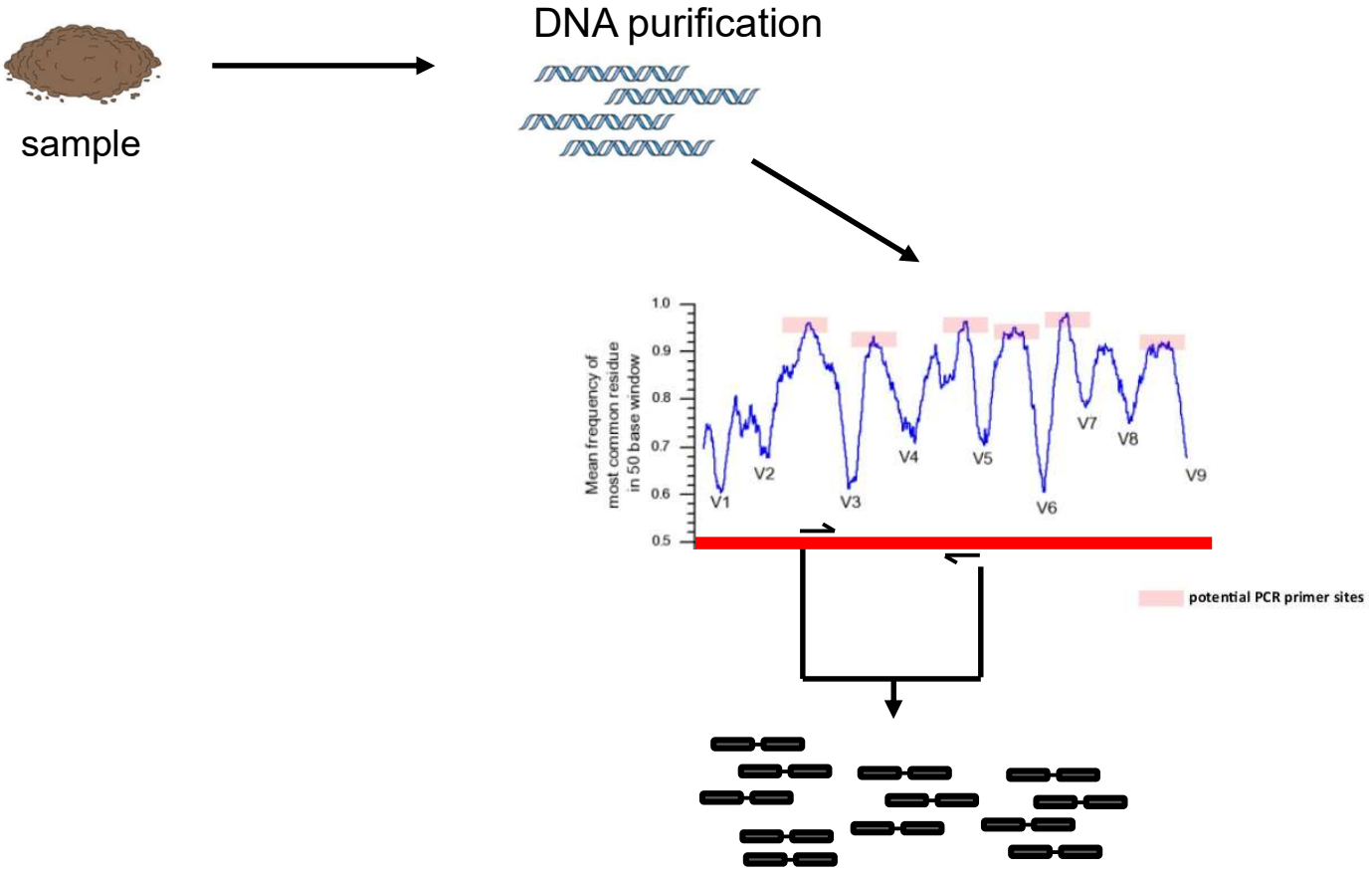
[Meisel et al., J Invest Dermatol, 2016]

Taxonomic profiling approaches – 16S rRNA amplicon



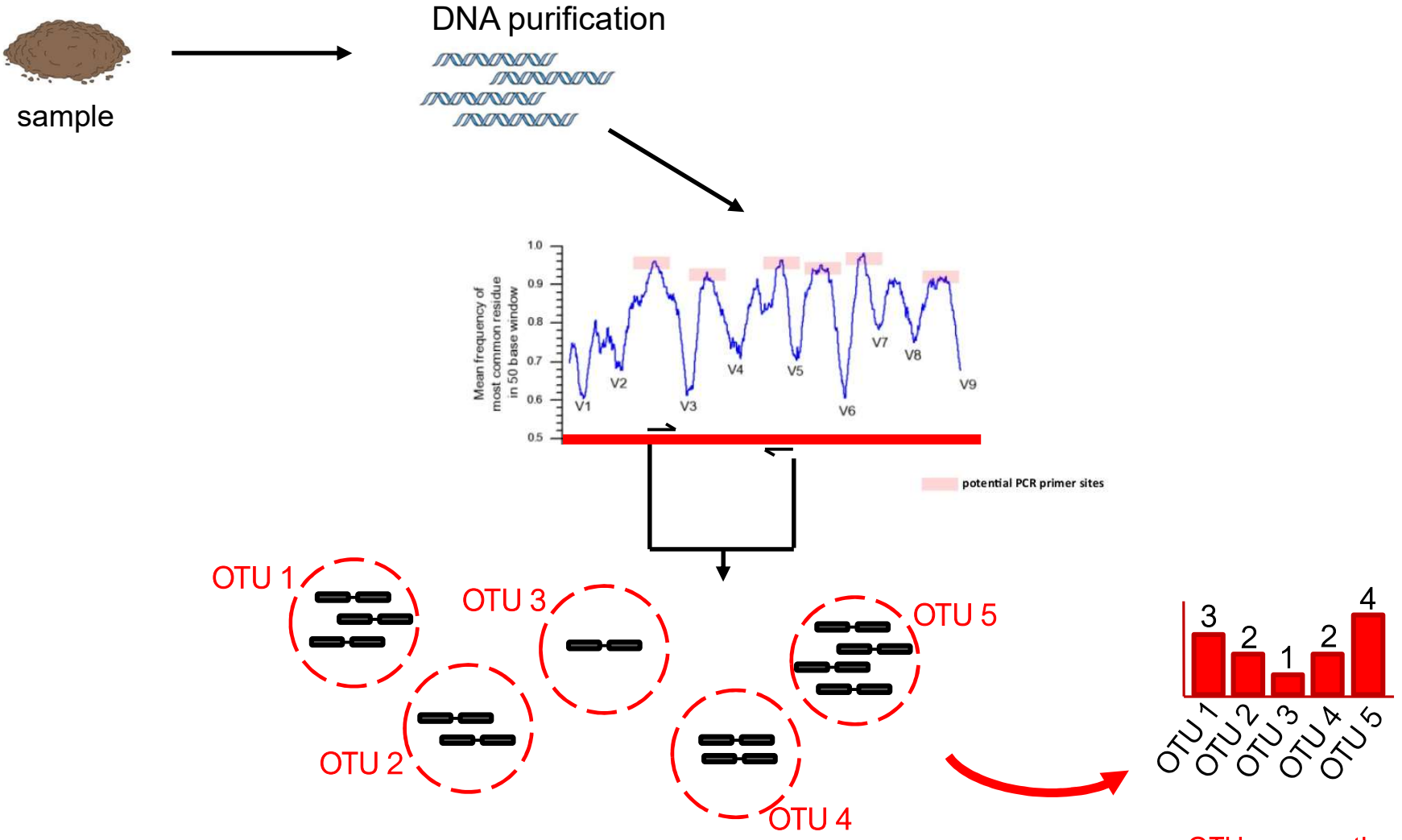
[Meisel et al., J Invest Dermatol, 2016]

Taxonomic profiling approaches – 16S rRNA amplicon



[Meisel et al., J Invest Dermatol, 2016]

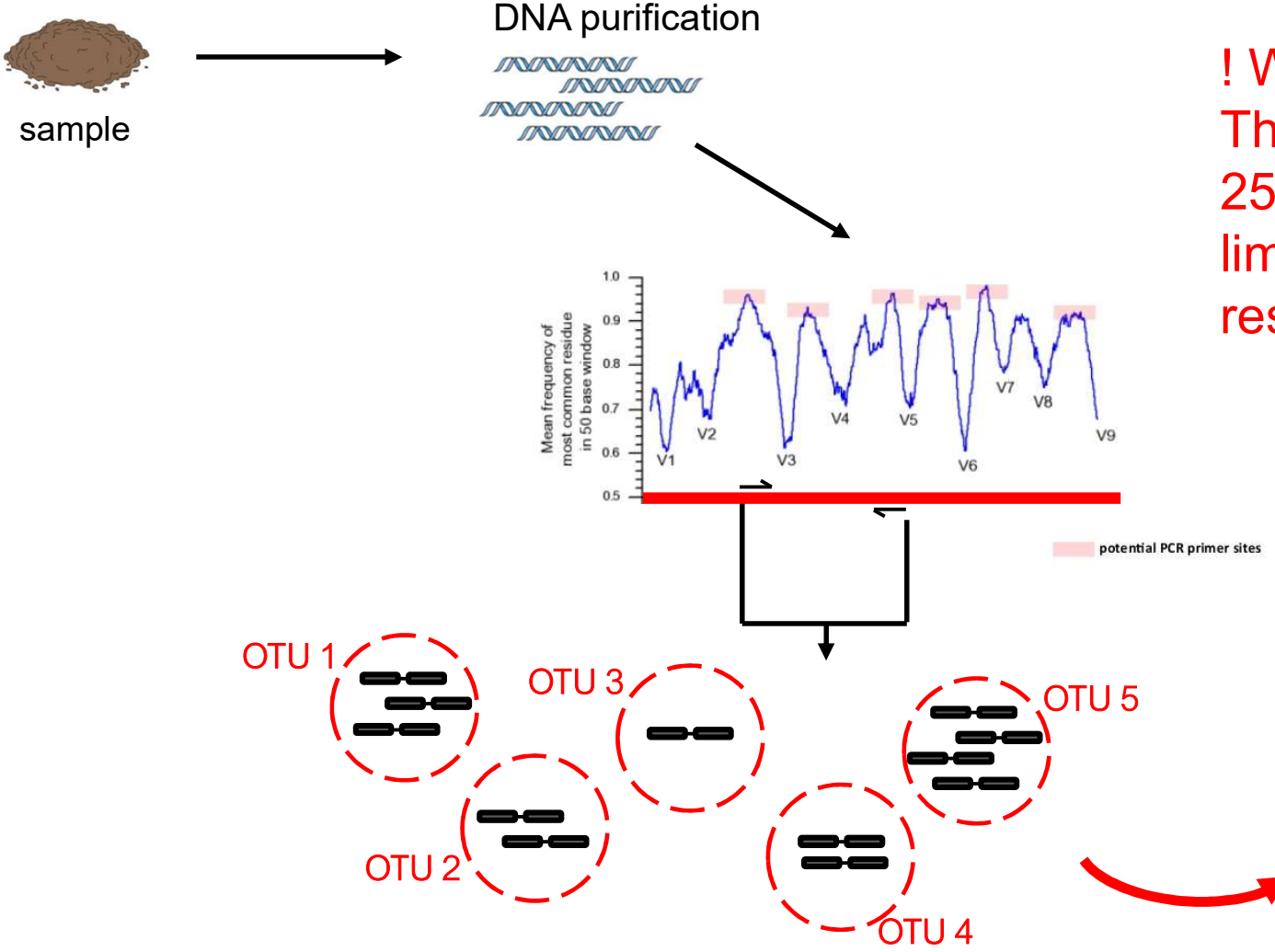
Taxonomic profiling approaches – 16S rRNA amplicon



[Meisel et al., J Invest Dermatol, 2016]

OTU = „operational taxonomic unit“

Taxonomic profiling approaches – 16S rRNA amplicon



! WARNING:
The OTUs have a length of 250/400 which results in limited taxonomic resolution

[Meisel et al., J Invest Dermatol, 2016]

OTU = „operational taxonomic unit“

nature methods

Explore content | About the journal | Publish with us

nature > nature methods > brief communications > article

Published: 20 October 2013

Metagenomic species profiling using universal phylogenetic marker genes

Shinichi Sunagawa, Daniel R Mende, Georg Zeller, Fernando Izquierdo-Carrasco, Simon A Berger, Jens Roat Kultima, Luis Pedro Coelho, Manimozhiyan Arumugam, Julien Tap, Henrik Bjørn Nielsen, Simon Rasmussen, Søren Brunak, Oluf Pedersen, Francisco Guarner, Willem M de Vos, Jun Wang, Junhua Li, Joël Doré, S Dusko Ehrlich, Alexandros Stamatakis & Peer Bork

Nature Methods **10**, 1196–1199 (2013) | Cite this article

16k Accesses | **294** Citations | **62** Altmetric | Metrics

nature communications

Explore content | About the journal | Publish with us

nature > nature communications > articles > article

Article | Open Access | Published: 04 March 2019

Microbial abundance, activity and population genomic profiling with mOTUs2

Alessio Milanese, Daniel R Mende, Lucas Paoli, Guillem Salazar, Hans-Joachim Ruscheweyh, Miguelangel Cuenca, Pascal Hingamp, Renato Alves, Paul I Costea, Luis Pedro Coelho, Thomas S. B. Schmidt, Alexandre Almeida, Alex L Mitchell, Robert D. Finn, Jaime Huerta-Cepas, Peer Bork, Georg Zeller & Shinichi Sunagawa

Nature Communications **10**, Article number: 1014 (2019) | Cite this article

24k Accesses | **103** Citations | **77** Altmetric | Metrics

CURRENT PROTOCOLS

PROTOCOL | Open Access

mOTUs: Profiling Taxonomic Composition, Transcriptional Activity and Strain Populations of Microbial Communities

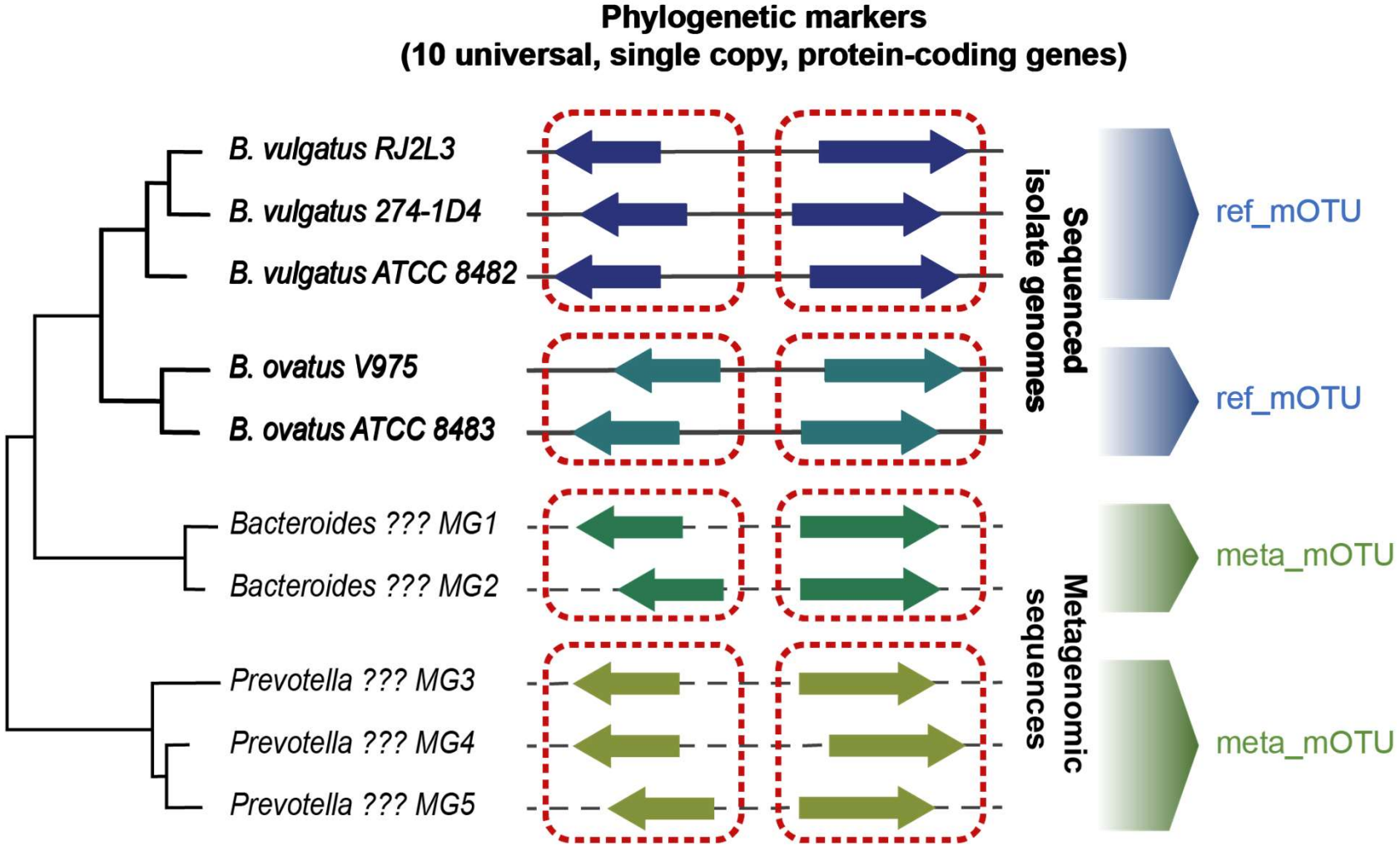
Hans-Joachim Ruscheweyh, Alessio Milanese, Lucas Paoli, Anna Sintsova, Daniel R. Mende, Georg Zeller, Shinichi Sunagawa

First published: 13 August 2021 | https://doi.org/10.1002/cpz1.218

sfx.ethz.ch

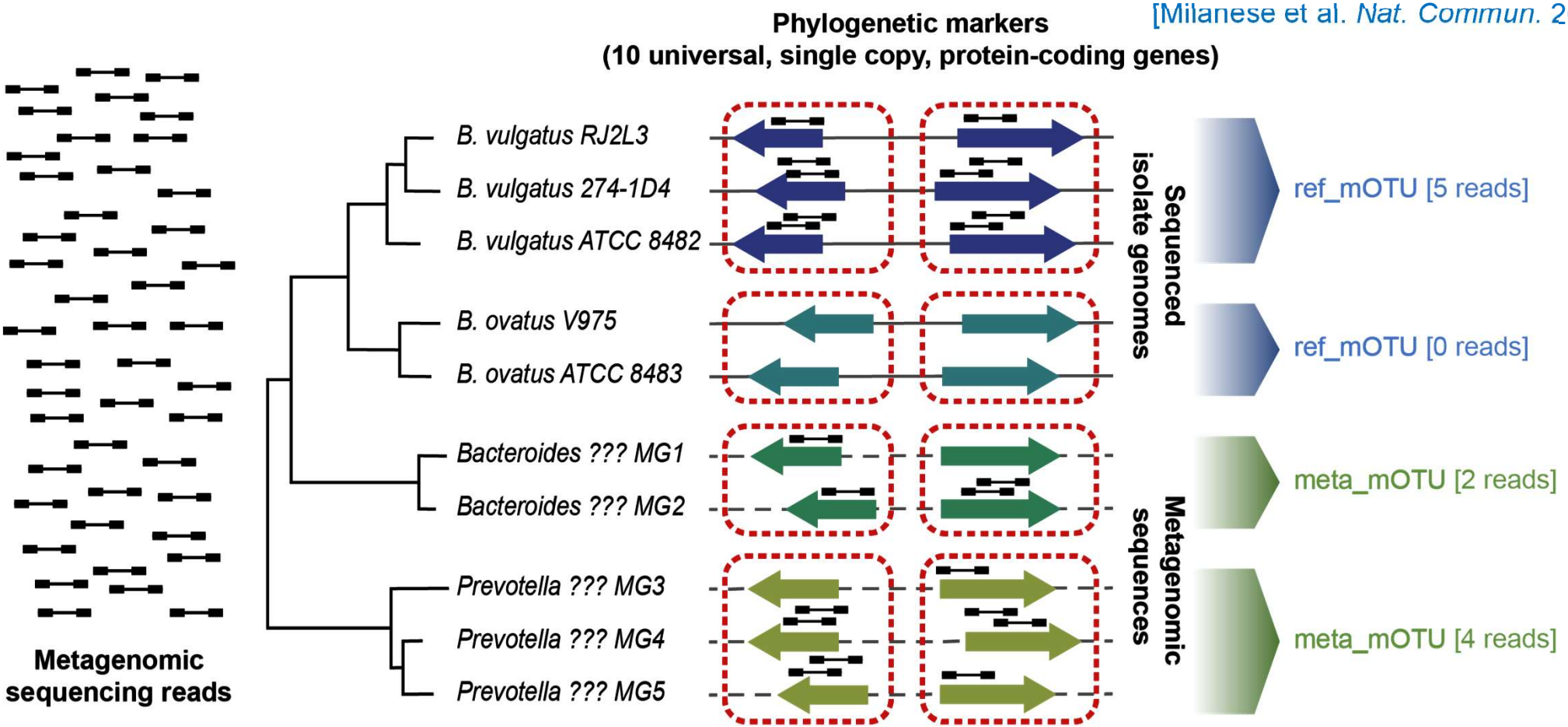
The mOTUs framework – DB construction

[Ciccarelli et al. *Science* 2006]
[Sunagawa et al. *Nat. Methods* 2013]
[Milanese et al. *Nat. Commun.* 2019]



The mOTUs framework – mapping / profiling

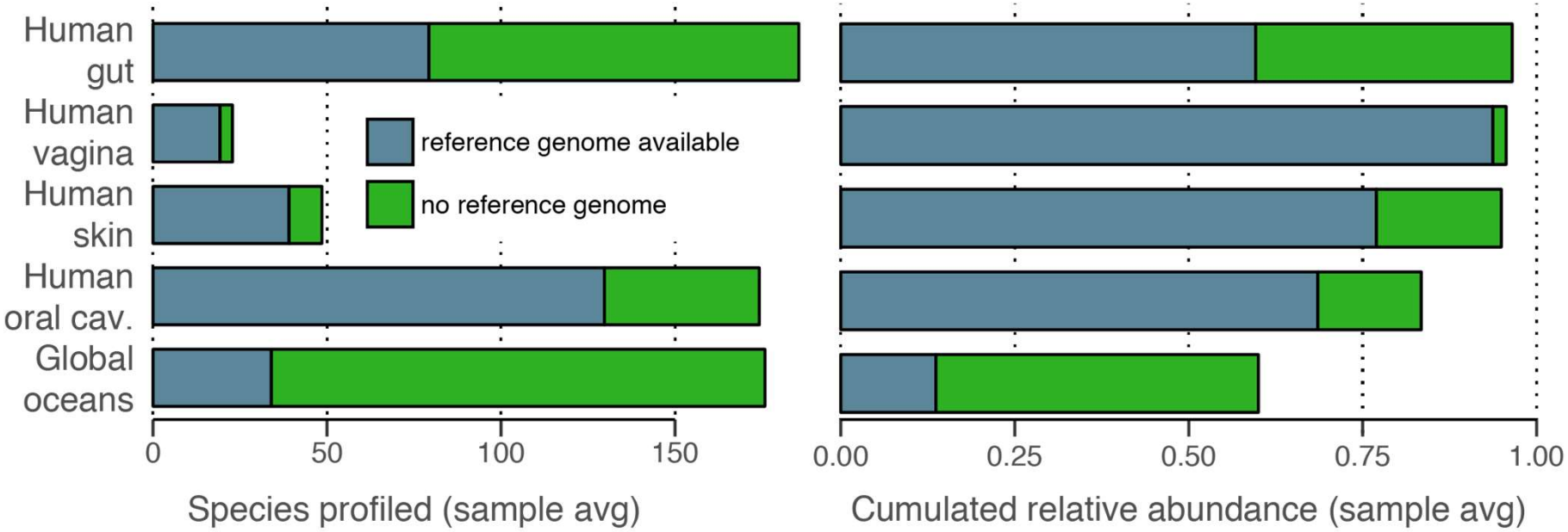
[Sunagawa et al. *Nat. Methods* 2013]
[Milanese et al. *Nat. Commun.* 2019]



Co-abundance binning

Break till 14:20

High-accuracy profiling of uncharacterized microbial species



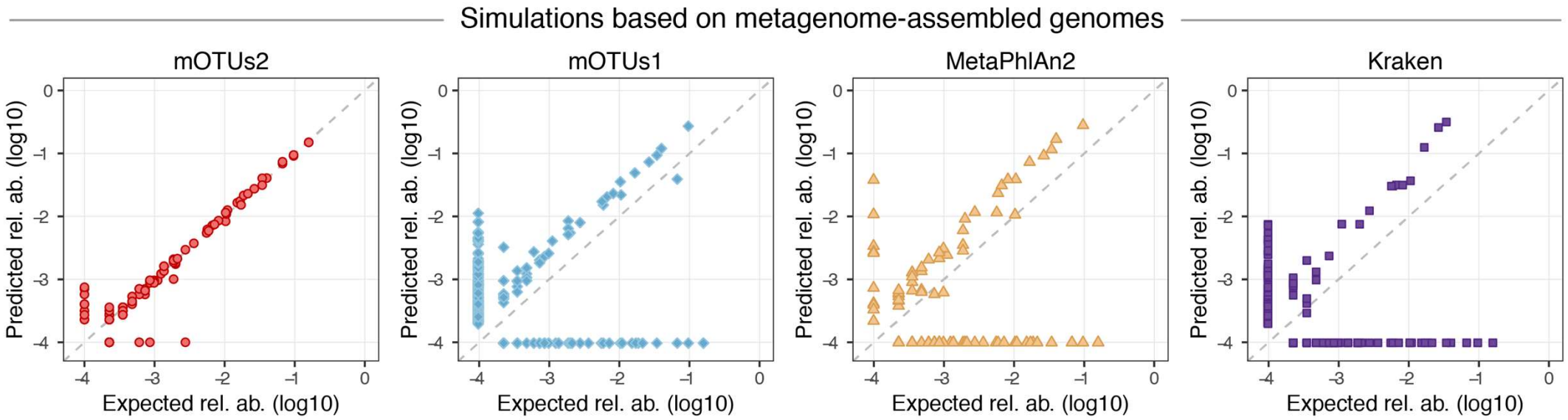
mOTUs

- High recall (non-reference species)
- High accuracy
- High resolution (up to subspecies)
- Also useful for meta-transcriptomics

[Milanese et al., *Nat. Commun.* 2019]

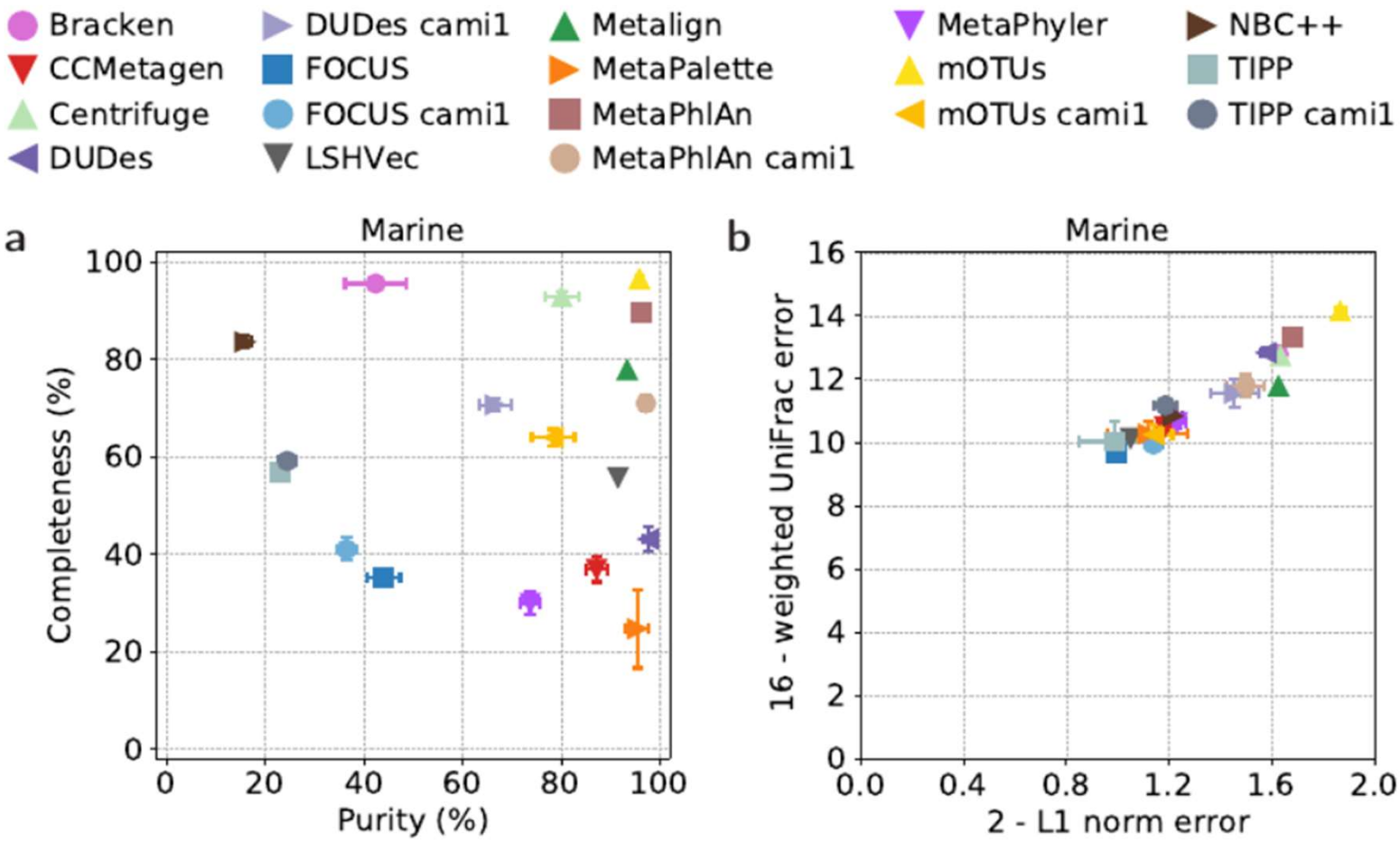
Improved quantification owing to comprehensive DB

Missing species in the reference database lead to **overestimating the abundance of known taxa**, i.e. to biased quantification.



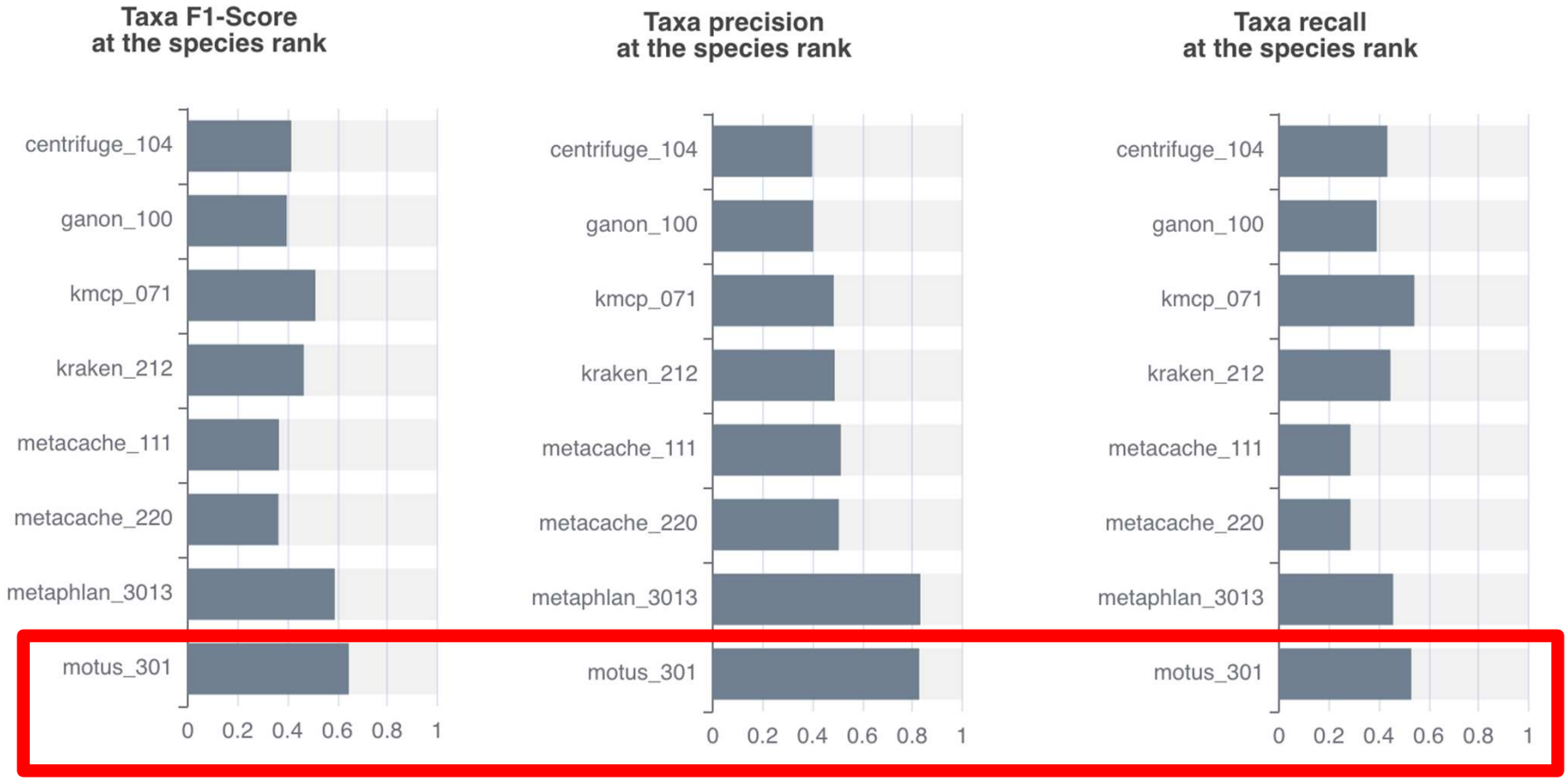
[Milanese et al., *Nat. Commun.* 2019]

High-accuracy profiling as evaluated by an independent benchmark - CAMI



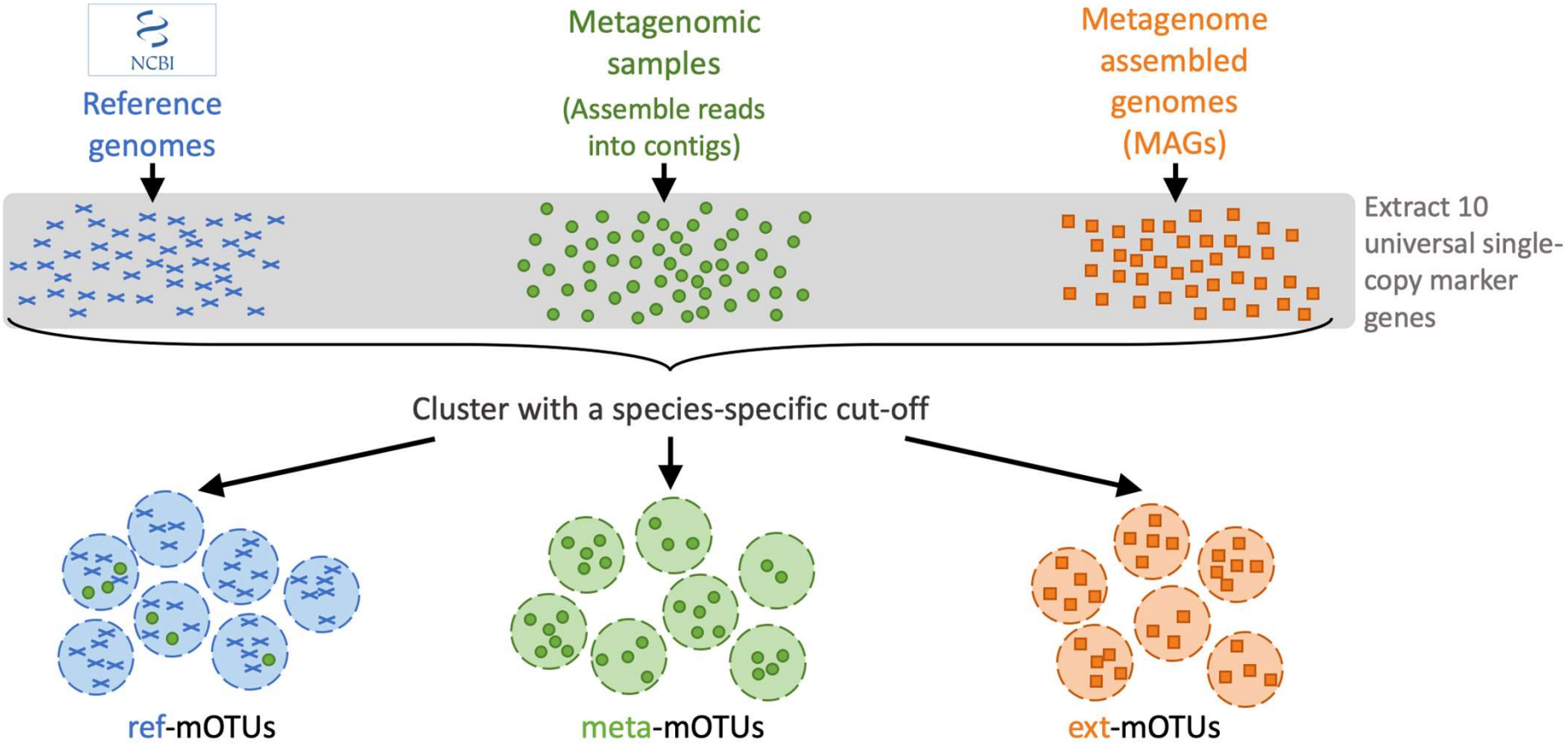
[Meyer et al., *Accepted in Nature Methods*, 2022]

High-accuracy profiling as evaluated by an independent benchmark - LEMMI



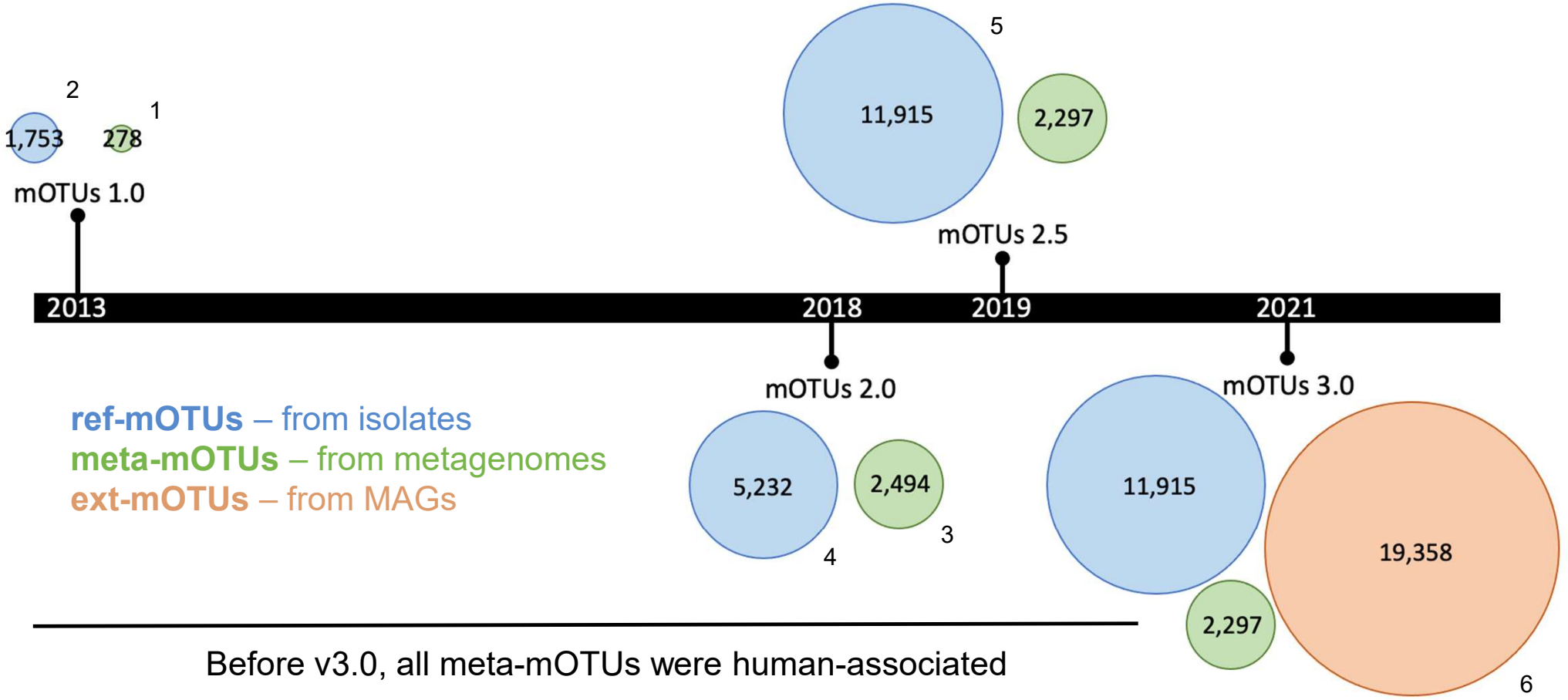
https://lemmi.ezlab.org/2022_01_PROK_NCBI

Incorporation of MAGs into the mOTUs3 database



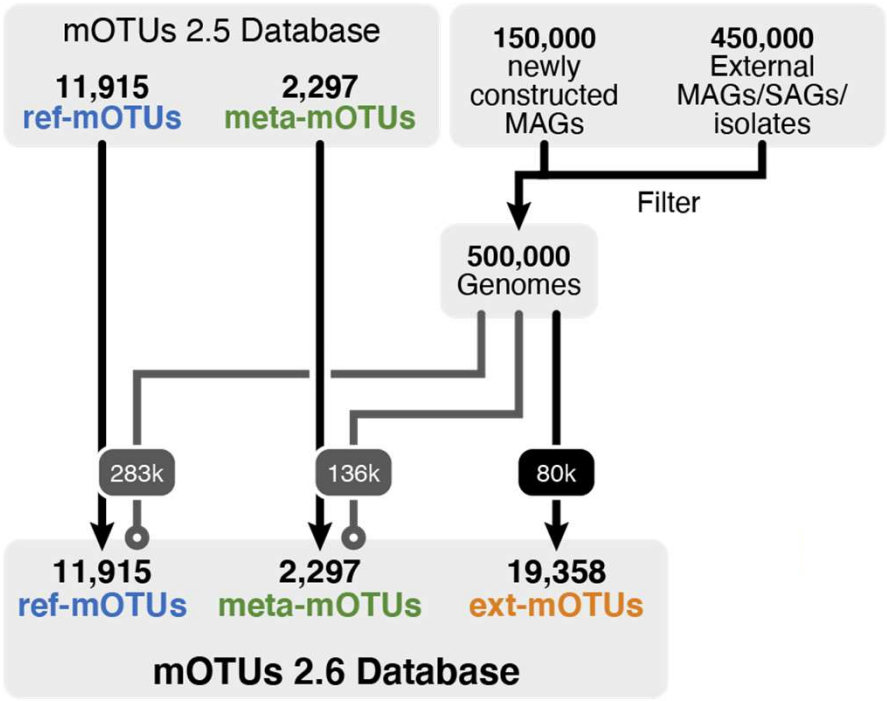
MAG-derived mOTUs are called ext_mOTUs

Improvement of scope in mOTUs since first version



1. [Sunagawa et al., *Nat. Methods* 2013]
2. [Mende et al. *Nat. Methods* 2013]
3. [Milanese et al., *Nat. Commun.* 2019]
4. [Mende et al., Nucleic Acids Res. 2017]
5. [Mende et al., Nucleic Acids Res. 2020]
6. [Ruscheweyh, Milanese et al. bioRxiv 2021]

mOTUs3 – database extension by marker genes from metagenome-assembled genomes (>500,000 MAGs)

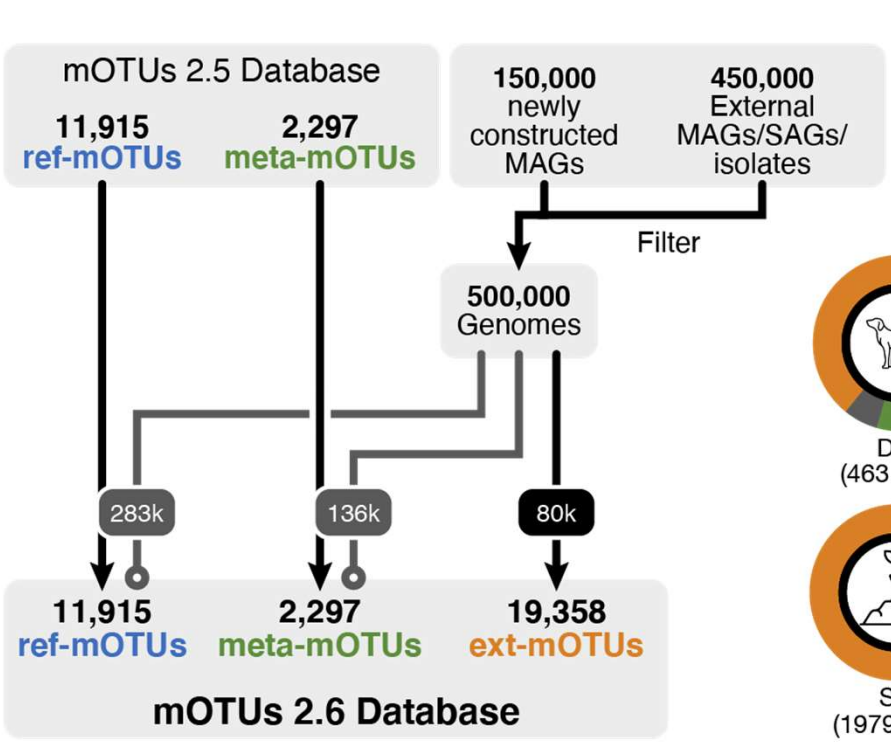


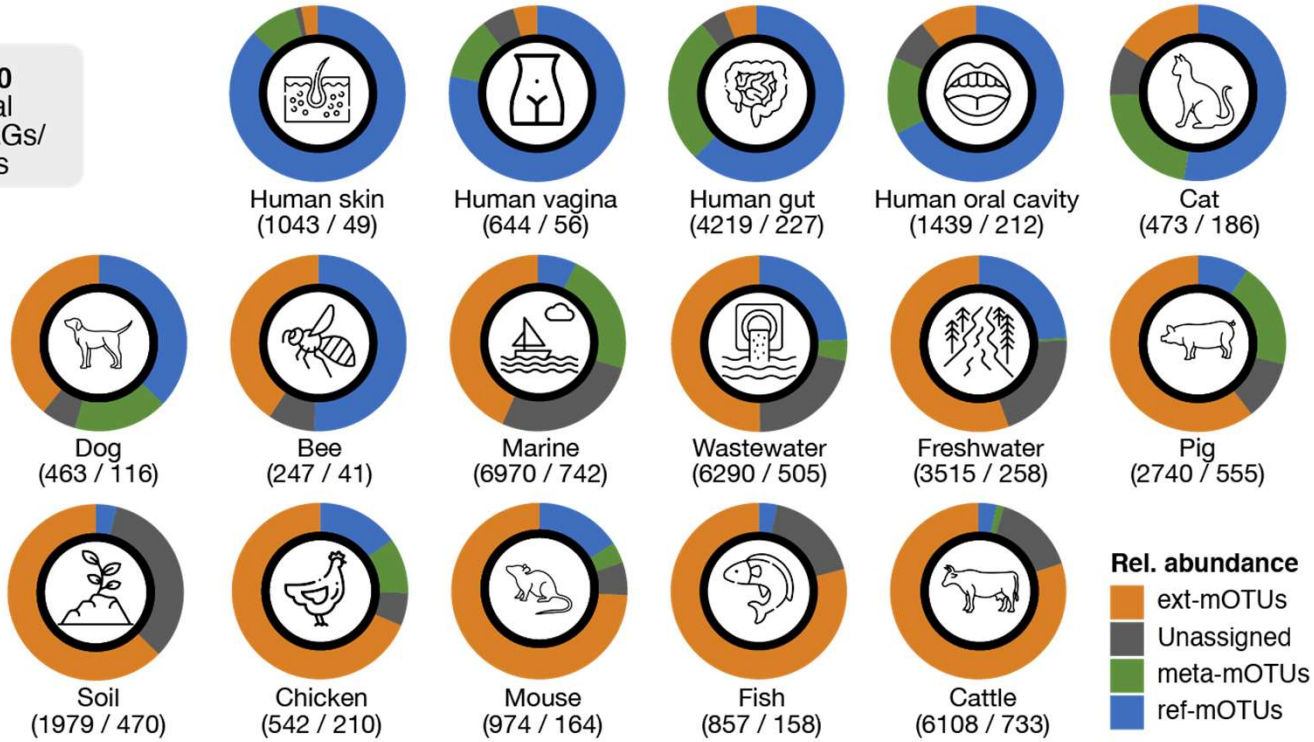
motu-tool.org

[Ruschewey, Milanese et al., bioRxiv 2021]

How do you map MAGs to mOTUs?

mOTUs3 – database extension by marker genes from metagenome-assembled genomes (>500,000 MAGs)





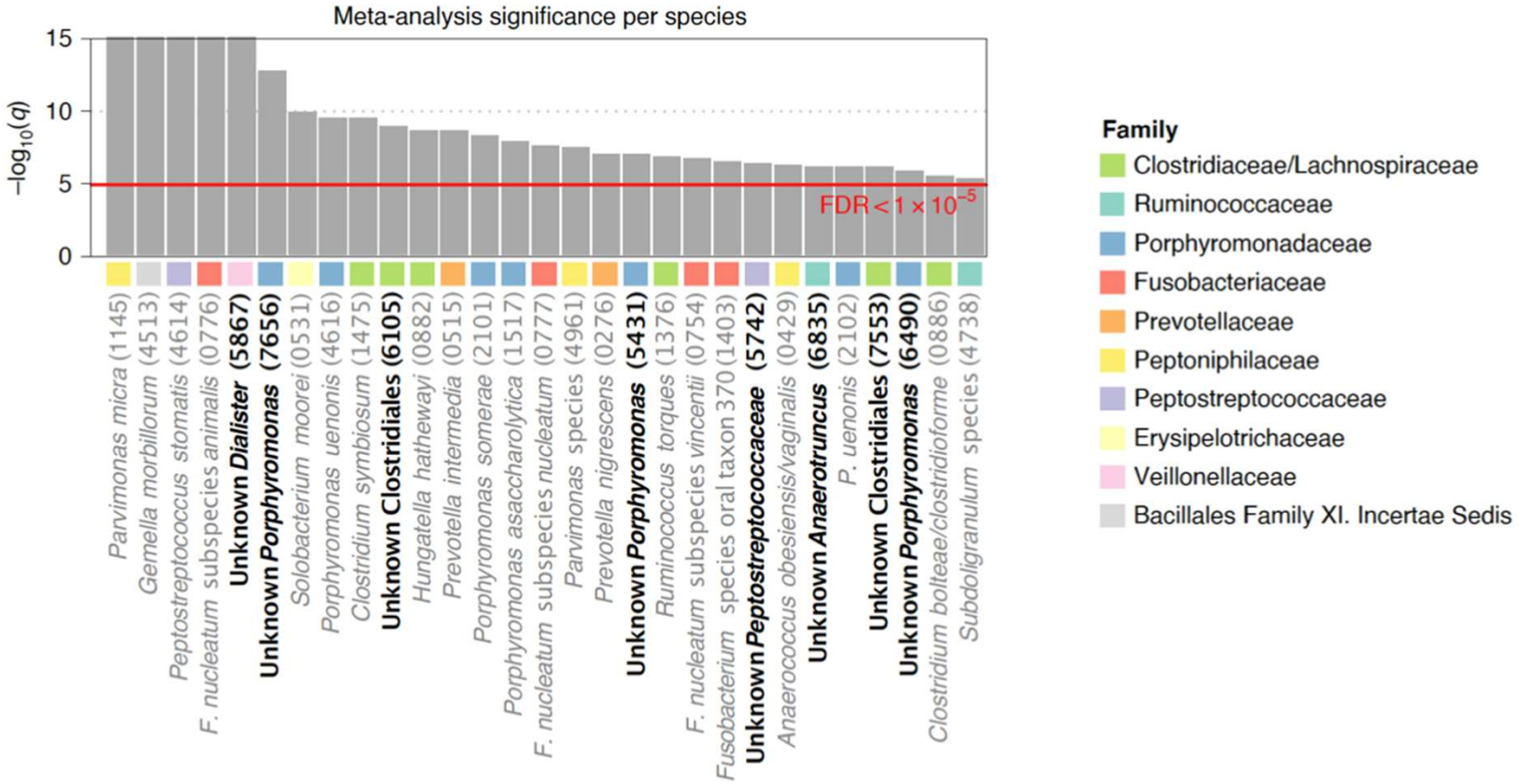
Enables profiling an unprecedented diversity of prokaryotes (33,570 species) across many environments.

motu-tool.org

[Ruschewey, Milanese et al., bioRxiv 2021]

What is the added value of using mOTUs?

Since we are able to profile unknown species, we can study associations of unknown species to disease



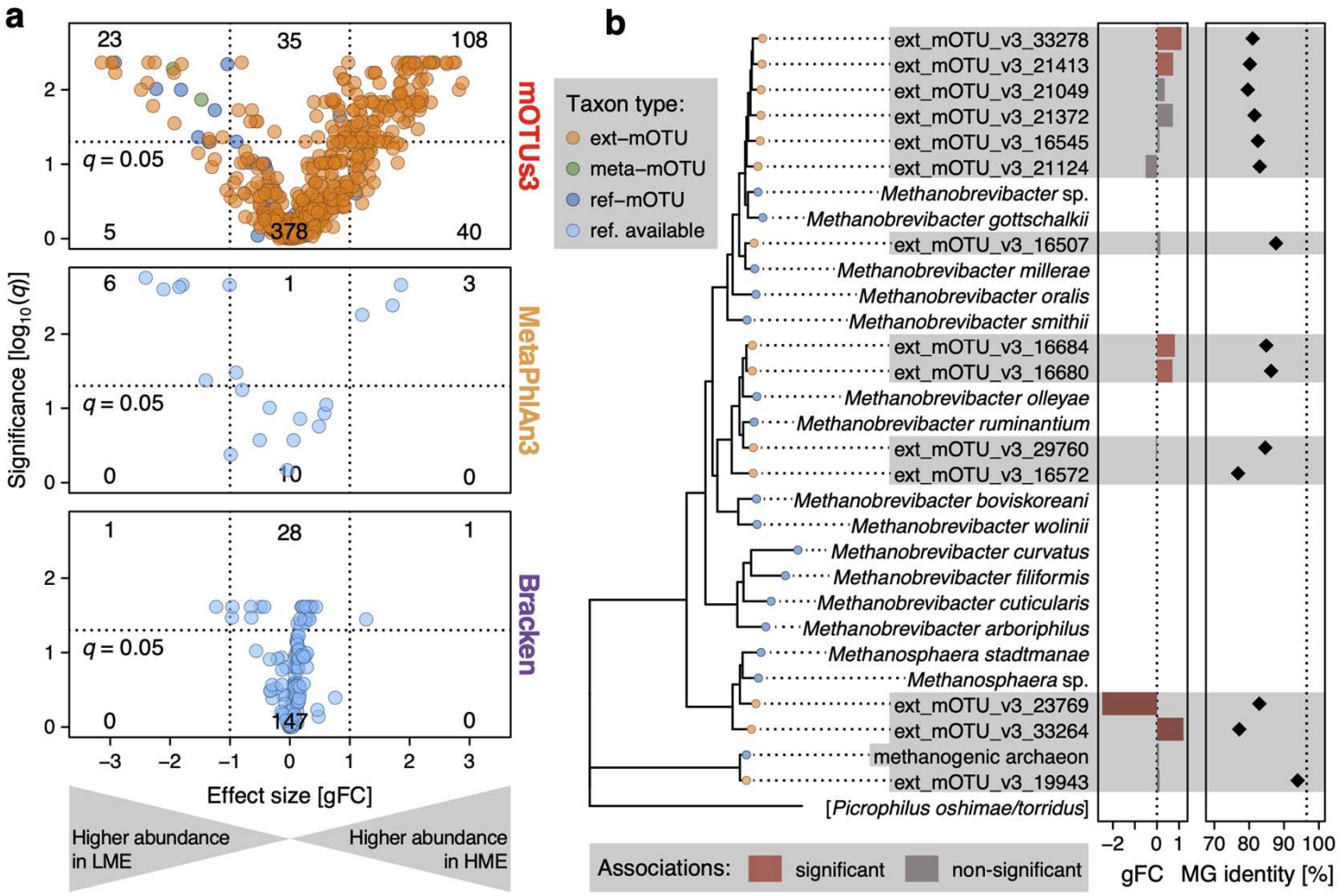
[Wirbel et al., Nature Medicine 2019]

What is the added value of using mOTUs?

- To demonstrate the utility of mOTUs3, we reanalysed rumen metagenomes from high- and low-methane emitting (HME and LME) sheep.

LME = low-methane emitting sheep

HME = high-methane emitting sheep

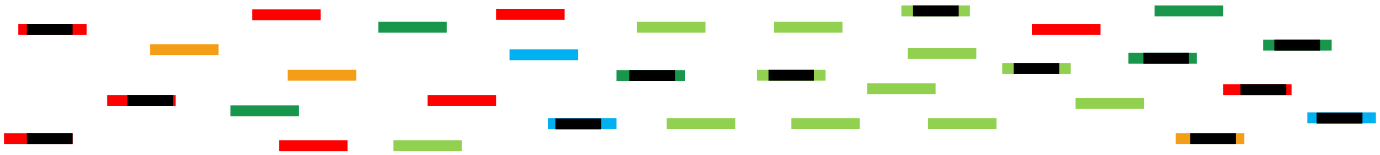


[Shi et al., Genome research 2014]
[Ruschewey, Milanese et al., bioRxiv 2021]

Functional profiling

Environmental sample

Shotgun sequencing



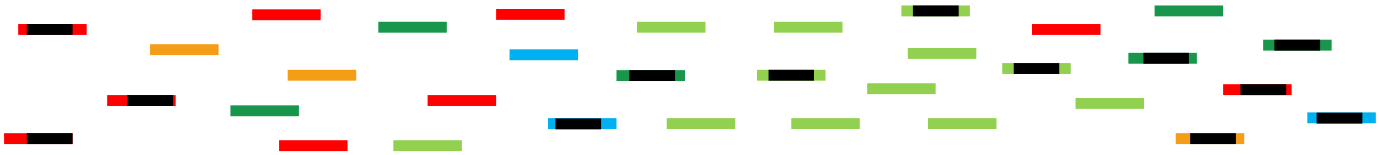
Taxonomic profiling

- Universal marker gene reads can be used for taxonomic profiling

Functional profiling

Environmental sample

Shotgun sequencing



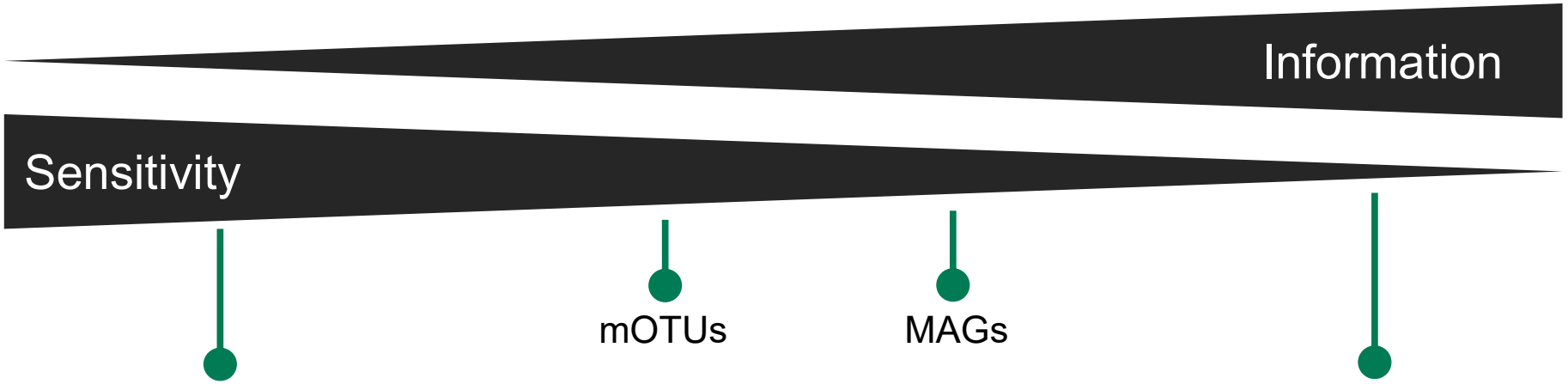
Taxonomic profiling

- Universal marker gene reads can be used for taxonomic profiling

Functional profiling

- The other gene can be used to understand the functional potential

Strengths and weaknesses of different approaches



16S amplicon profiling

- Taxonomic profiling
- High sensitivity (discover the unknown)
- No functional profiling
- Sequencing errors
- PCR amplification errors
- Copy number bias
- Lower resolution

whole-genome profiling

- Taxonomic profiling
- Need to estimate unknown species
- Functional profiling (potential)
- Sequencing errors
- High resolution

culture-based analysis

- Taxonomic profiling is limited to cultivable species
- Full potential for functional activity that can be tested
- Best resolution

Take home message

- **Taxonomic profiling** is a fundamental step in the analysis of microbial communities
- **Incomplete reference databases** yield incomplete pictures (missing taxa) and tend to overestimate the abundances of known taxa
- 16S rRNA amplicon, shotgun sequencing and culture based approaches **are complementary**
- **16S rRNA amplicon** is a cheap solution that allow for sensitive taxonomic profiling
- **Shotgun sequencing** allows for a more precise taxonomic profiling (species and subspecies resolution) and functional profiling
- **Culture based approaches** are limited on the number of species that can study, but allows to fully investigate species properties
- **mOTUs** is one of the most competitive tools on the market for taxonomic profiling of metagenomes