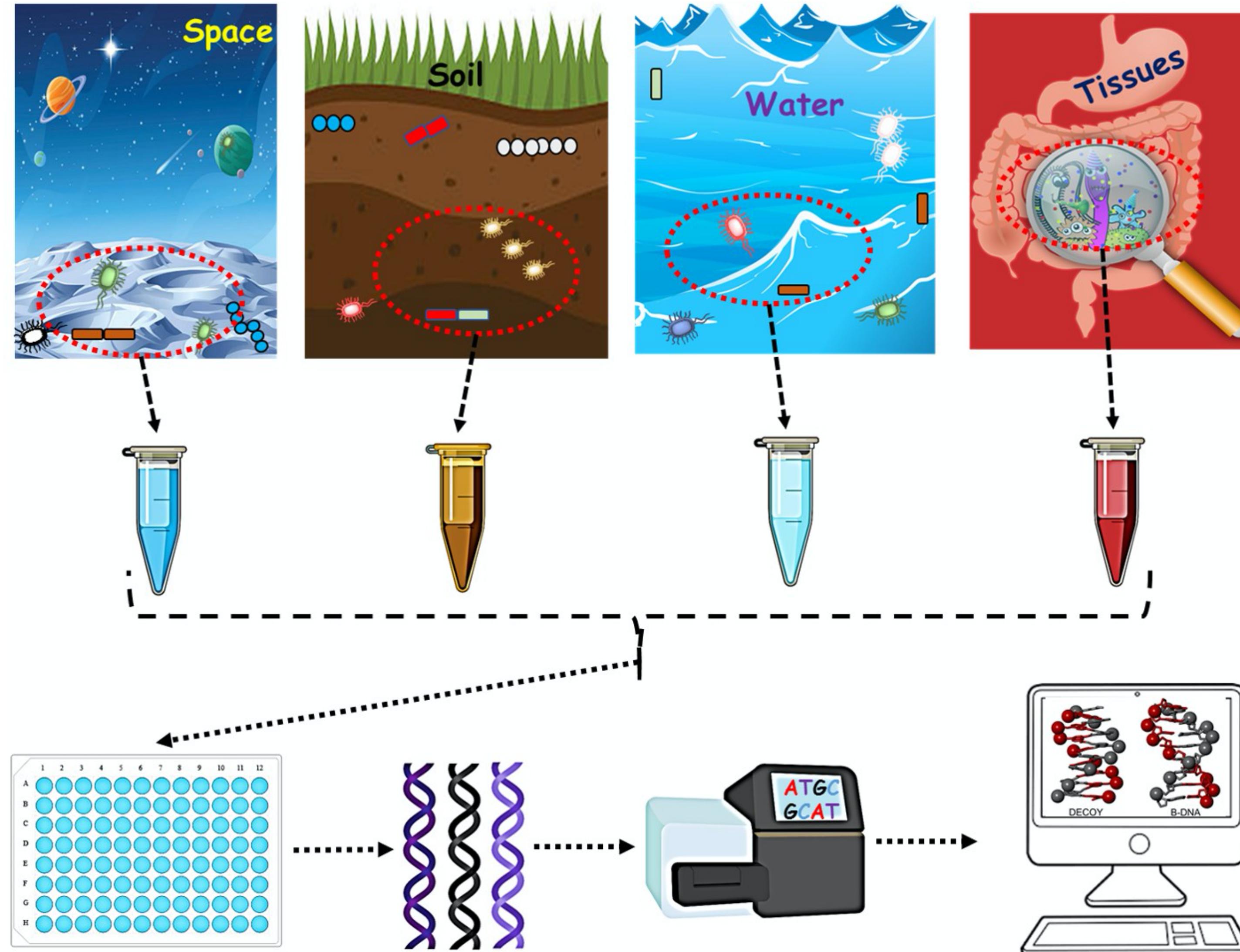


How to find fungi in RNA-seq data?

Valeria Flores Almaraz

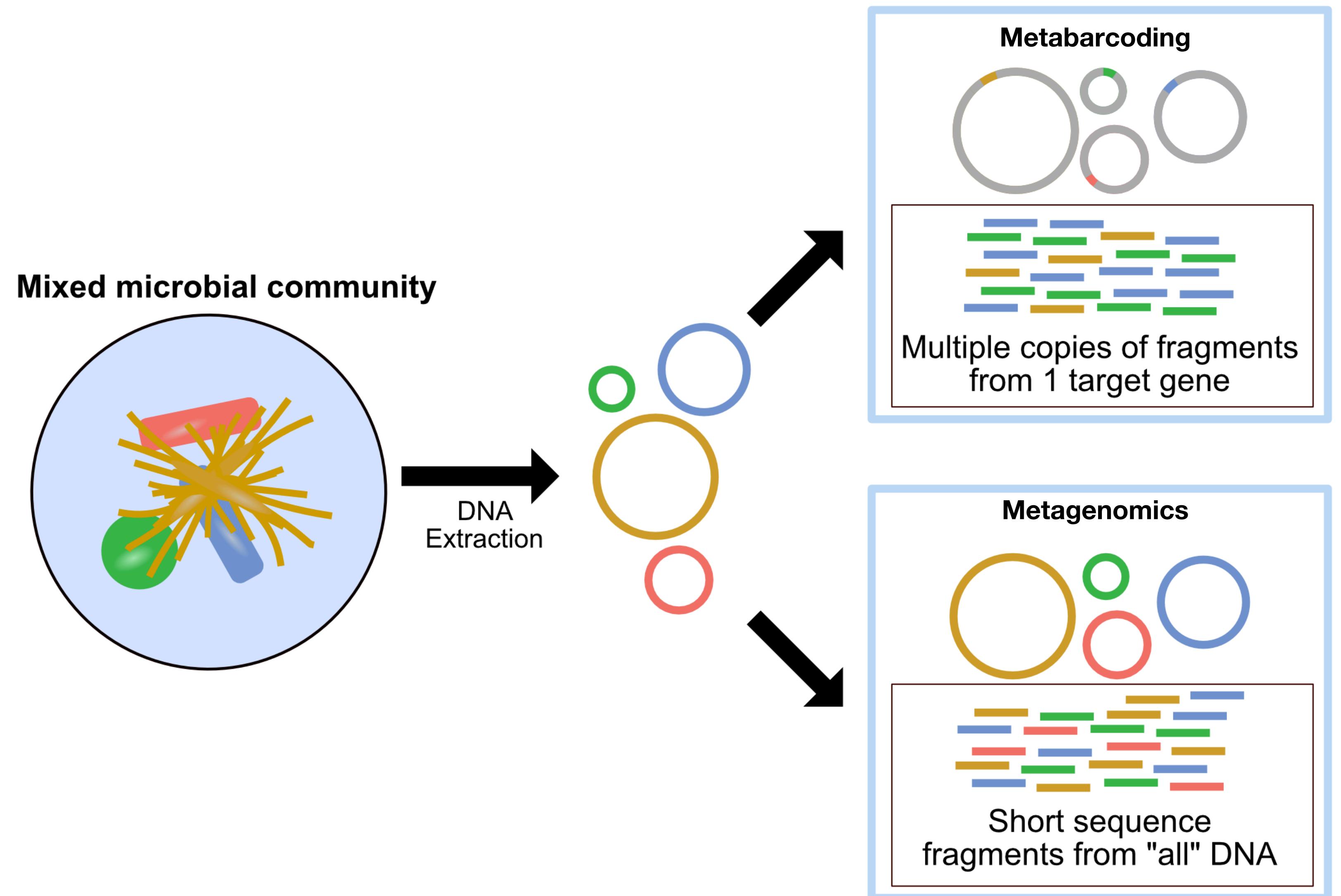
29-10-20

Fungi and HTS



Most of the HTS has focused on DNA sequencing of entire communities using either:

- **Targeted approaches**
- **Shotgun sequencing**



(Shakya *et al.* 2019)

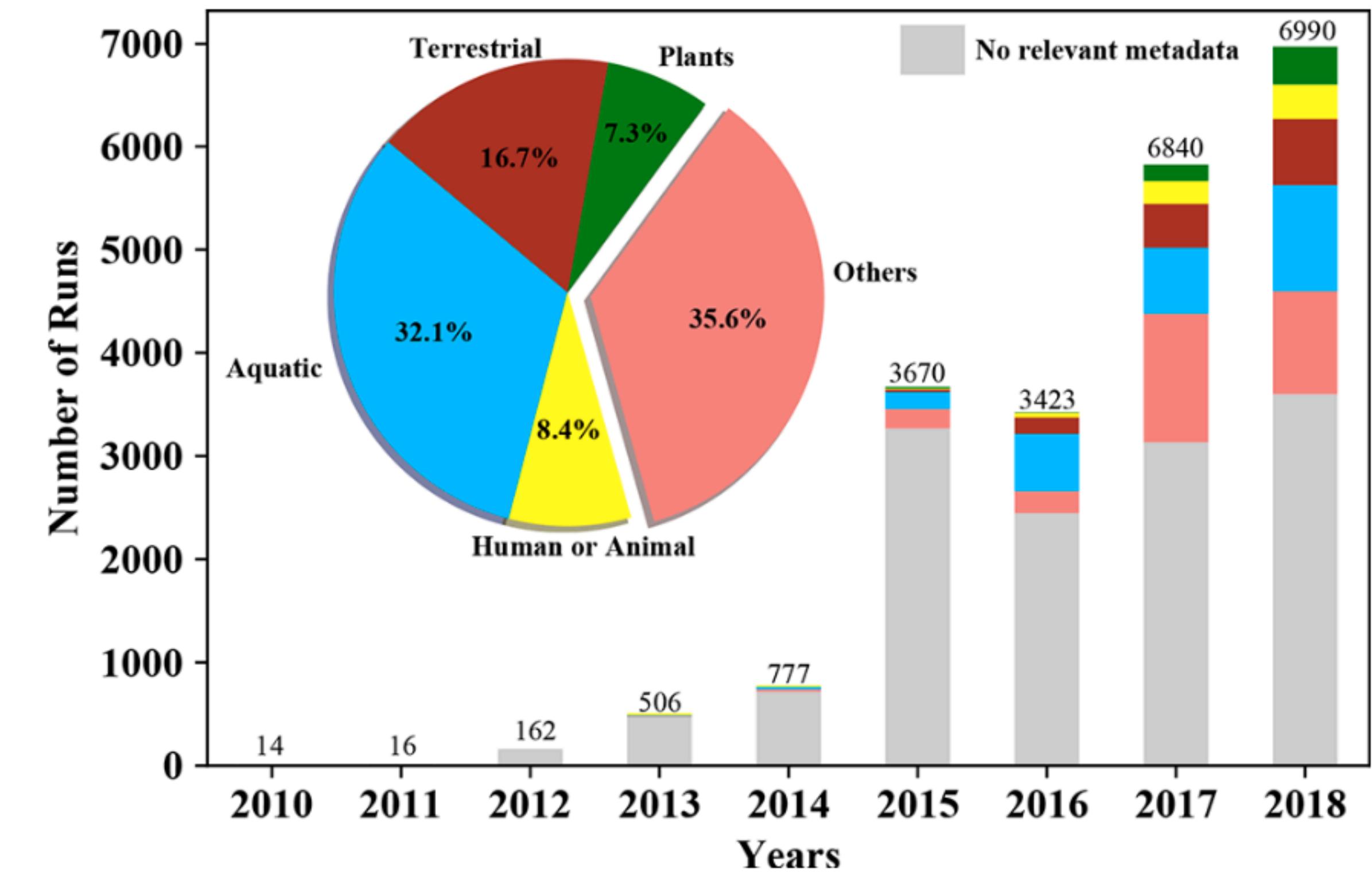
(Modified from: Lee, 2019)

Metatranscriptomics

Is typically used to:

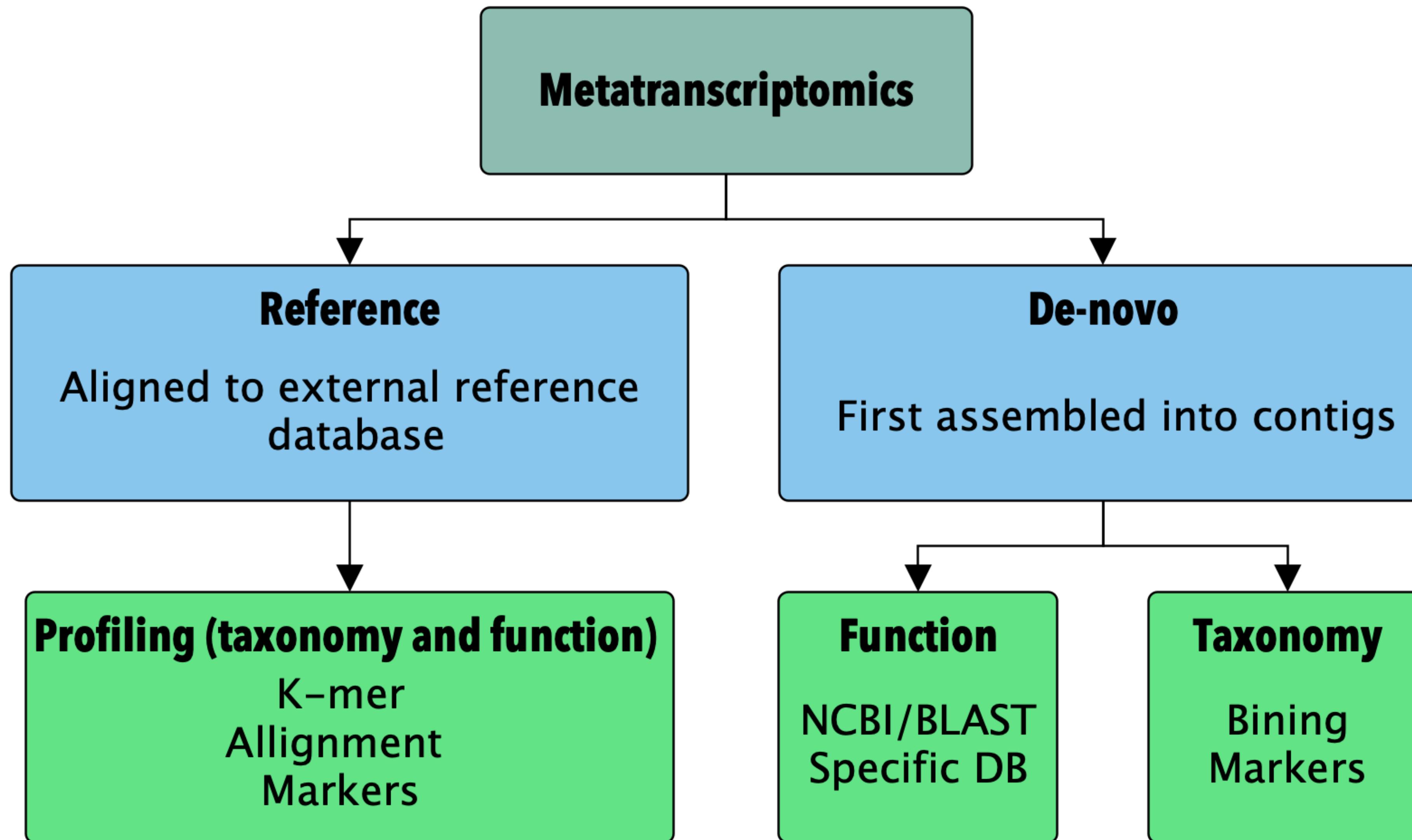
- Identify
- Quantify
- Compare

Functional responses of microbial communities in natural habitats or in relation to environmental or physiological impacts.



(Shakya *et al.* 2019)

- Host
- Organism of interest



Reference

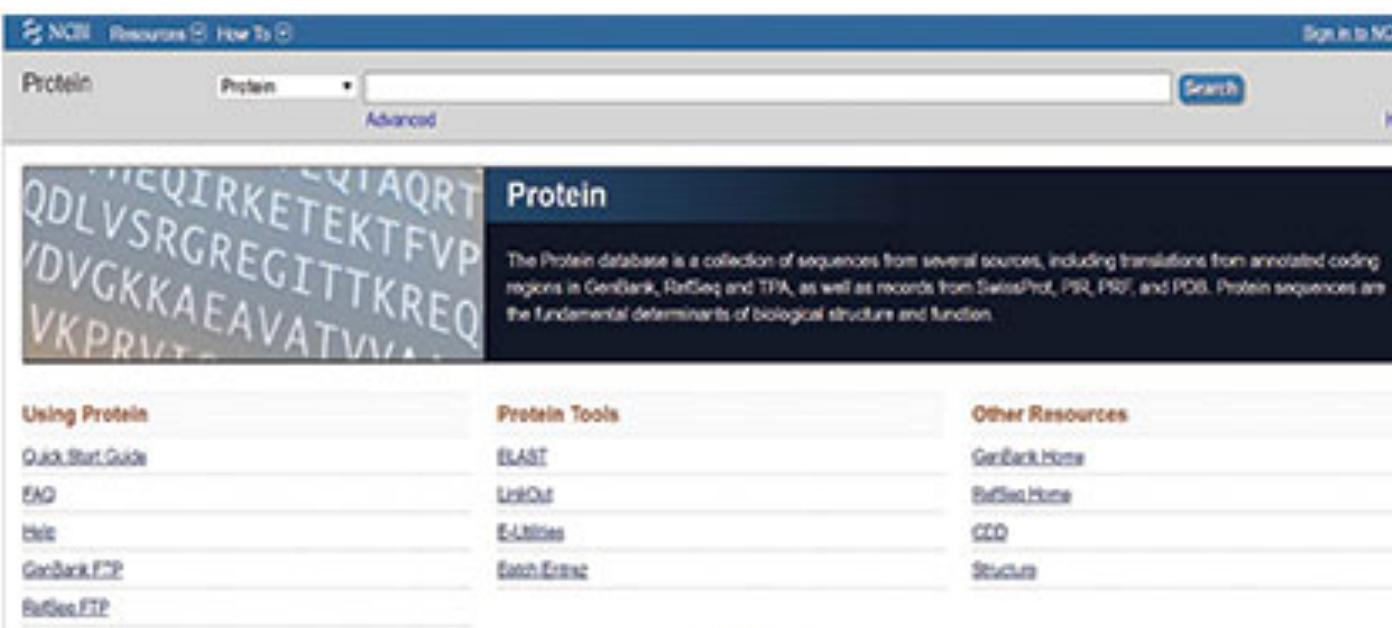
Unassembled mRNA reads compared against reference databases to assign taxonomy or annotate genes.

Taxonomic or functional assignment depends on homology between the single read and the reference.

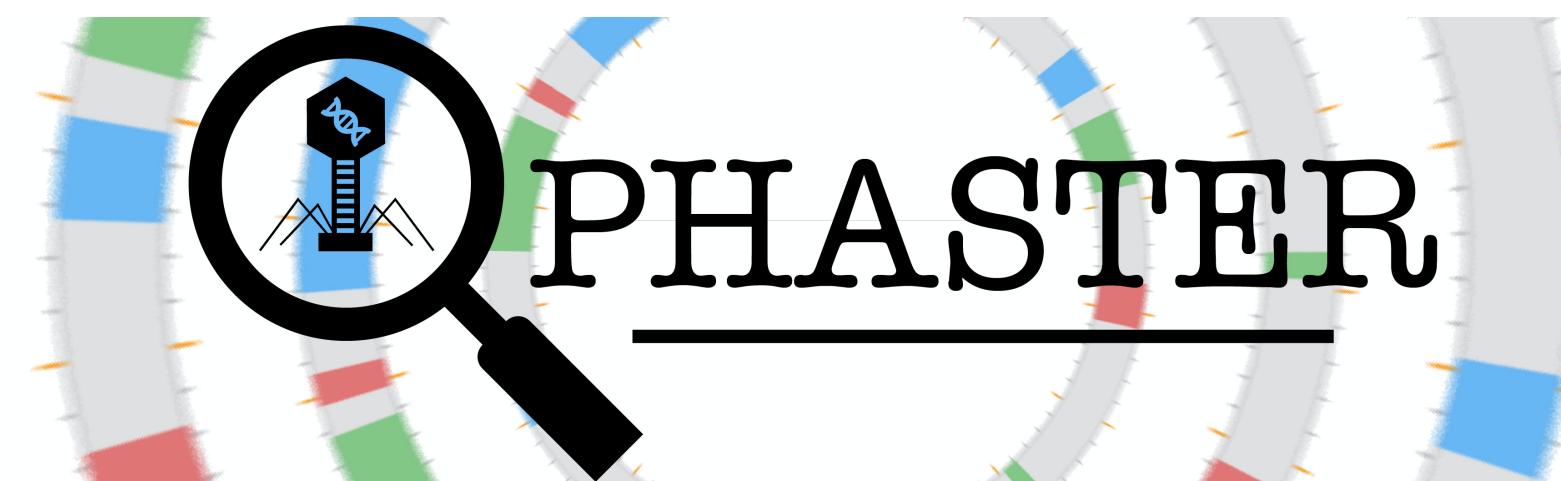
DATABASE CHOICE IS CRUCIAL

(Knight *et al.* 2018)

Databases



MetaHIT

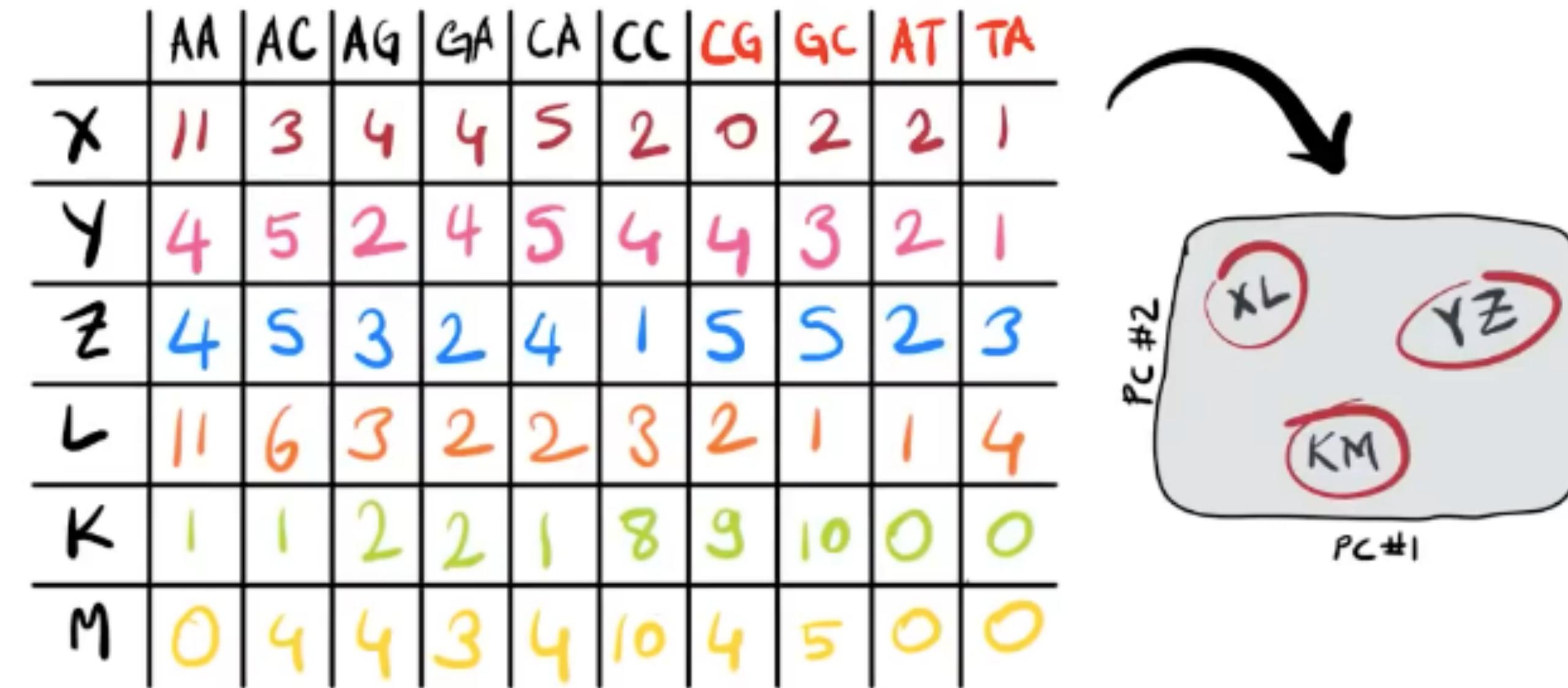


GREENGENES



Reference

***k-mer*:** Assigning taxonomy to short DNA/RNA fragments of length k .



Genes from Nine Genomes Are Separated into Their Organisms in the Dinucleotide Composition Space

Hiroshi NAKASHIMA,^{1,*} Motonori OTA,² Ken NISHIKAWA,² and Tatsuo OOI³

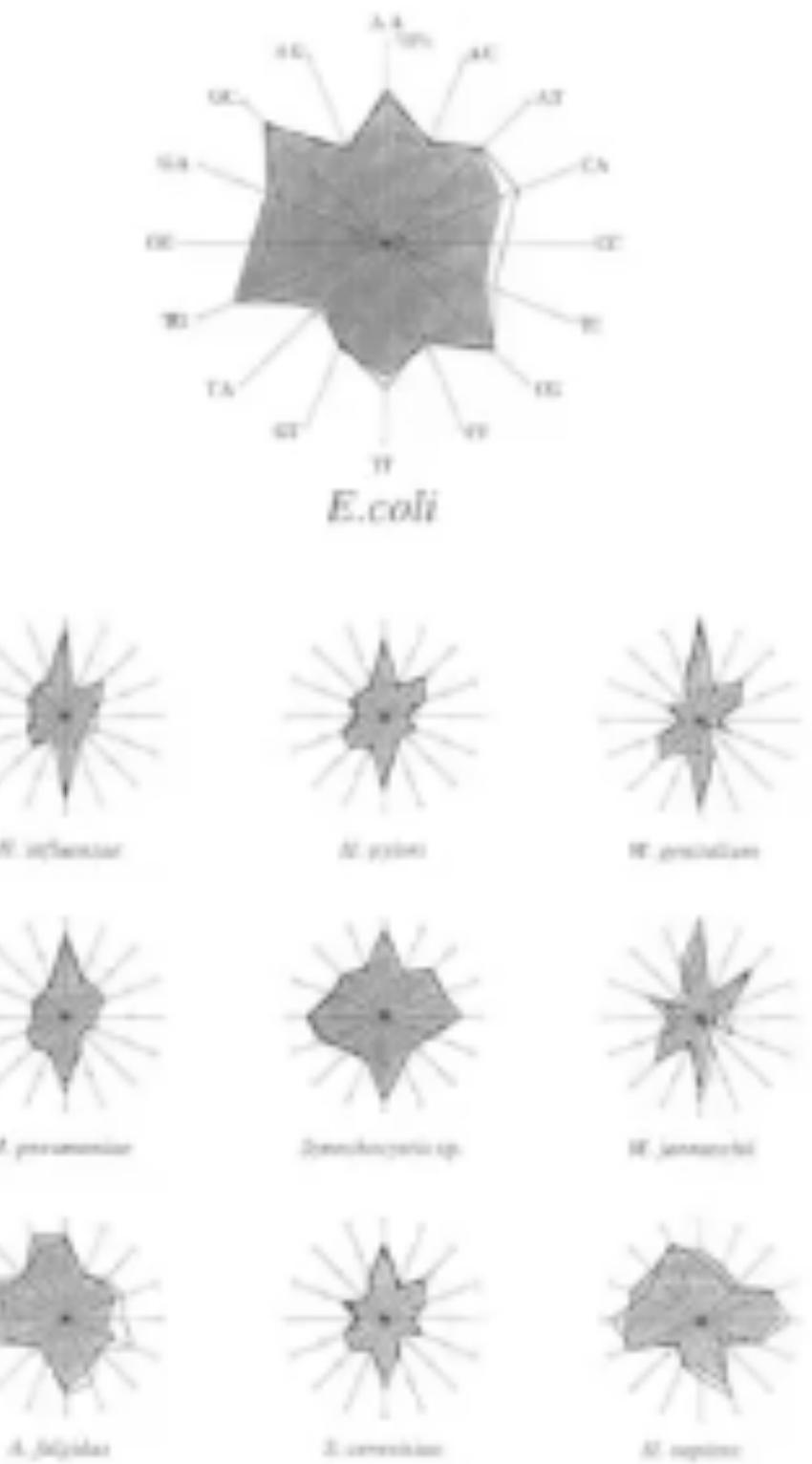
School of Health Sciences, Faculty of Medicine, Kanazawa University, 5-11-80 Kodatsuno, Kanazawa 920-0942, Japan,¹ Center for Information Biology, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan,² and Kyoto Women's University, Kitahiyoshi-cho 35, Higashiyama-ku, Kyoto 605, Japan³

(Received 2 September 1998)

Abstract

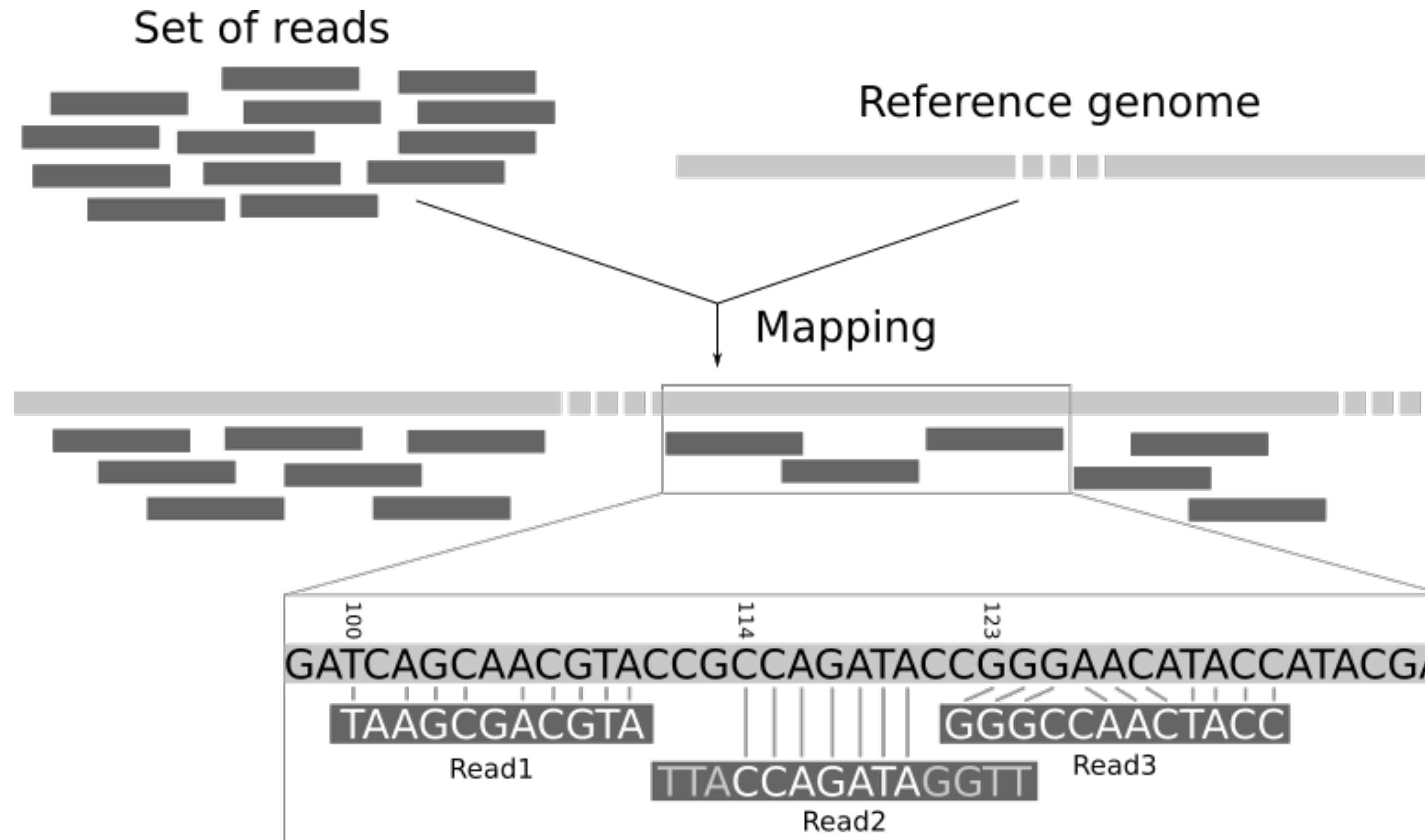
A set of 16 kinds of dinucleotide compositions was used to analyze the protein-encoding nucleotide sequences in nine complete genomes: *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Synechocystis* sp., *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, and *Saccharomyces cerevisiae*. The dinucleotide composition was significantly different between the organisms. The distribution of genes from an organism was clustered around its center in the dinucleotide composition space. The genes from closely related organisms such as Gram-negative bacteria, mycoplasma species and eukaryotes showed some overlap in the space. The genes from nine complete genomes together with those from human were discriminated into respective clusters with 80% accuracy using the dinucleotide composition alone. The composition data estimated from a whole genome was close to that obtained from genes, indicating that the characteristic feature of dinucleotides holds not only for protein coding regions but also noncoding regions. When a dendrogram was constructed from the disposition of the clusters in the dinucleotide space, it resembled the real phylogenetic tree. Thus, the distinct feature observed in the dinucleotide composition may reflect the phylogenetic relationship of organisms.

Key words: separation of genes; dinucleotide frequency; phylogenetic tree



Reference

Alignment: To all sequences or taxonomically unique markers.



(Wolff *et al.* 2020)

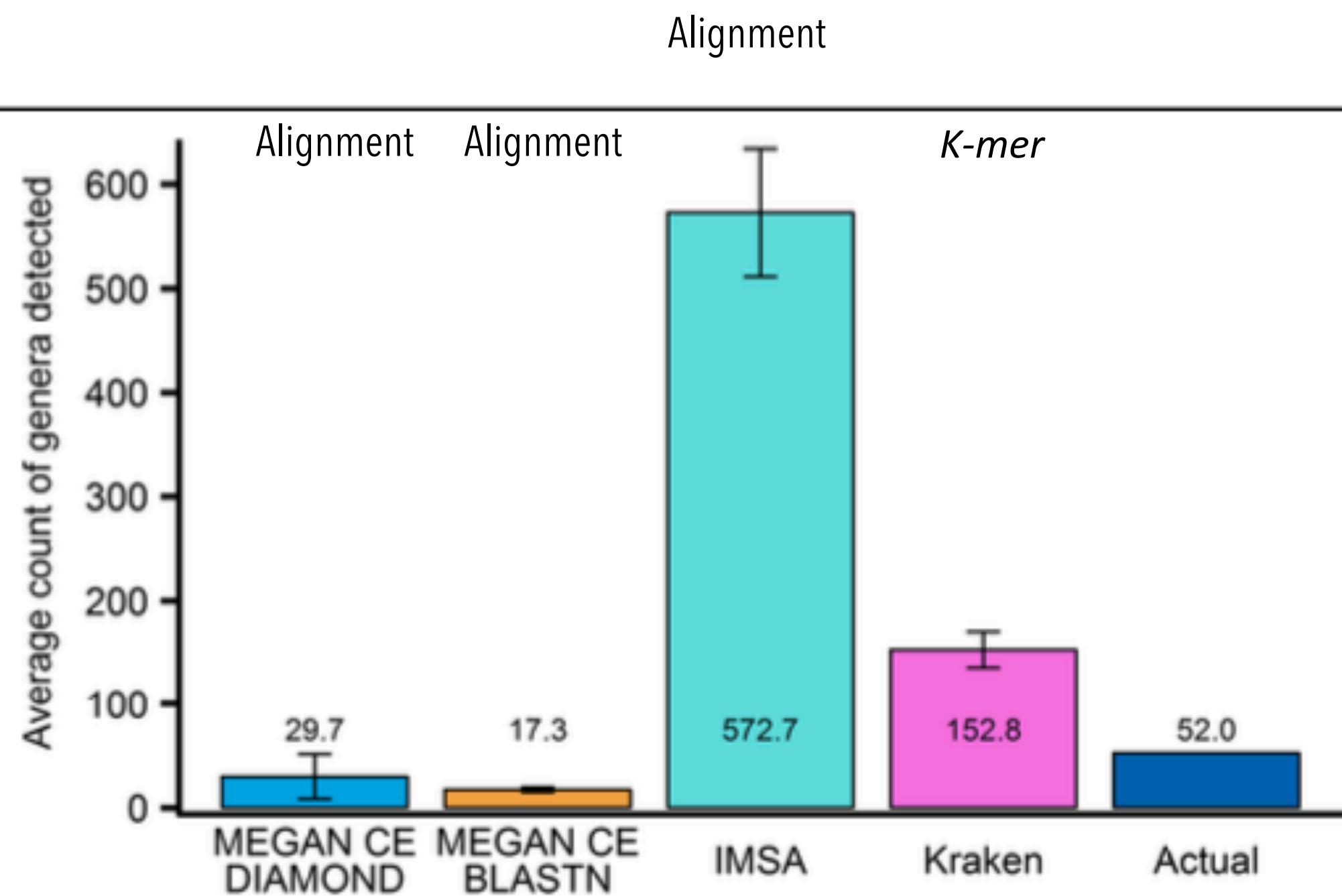


Fig. 1 Comparison of the selected metataxonomics workflows on detection of genera within a set of simulated datasets (Table 1). IMSA and Kraken identify too many taxa. Both versions of MEGAN CE find too few taxa, most likely due to the weighted LCA that filters out noise, which also filters out weak signal of organisms present

Table 7 Summary of Comparison of Various Tools on ASF data sample

Method	Total genera detected	False positives	True positives	Correct next relative ^a
IMSA+A	19	2	6	11
MEGAN CE DIAMOND	13	6	5	2
MEGAN CE BLASTN	15	9	5	1
Kraken	72	55	6	11

^aThe number of genera representing organisms closely related to the ASF bacteria without sequenced genomes

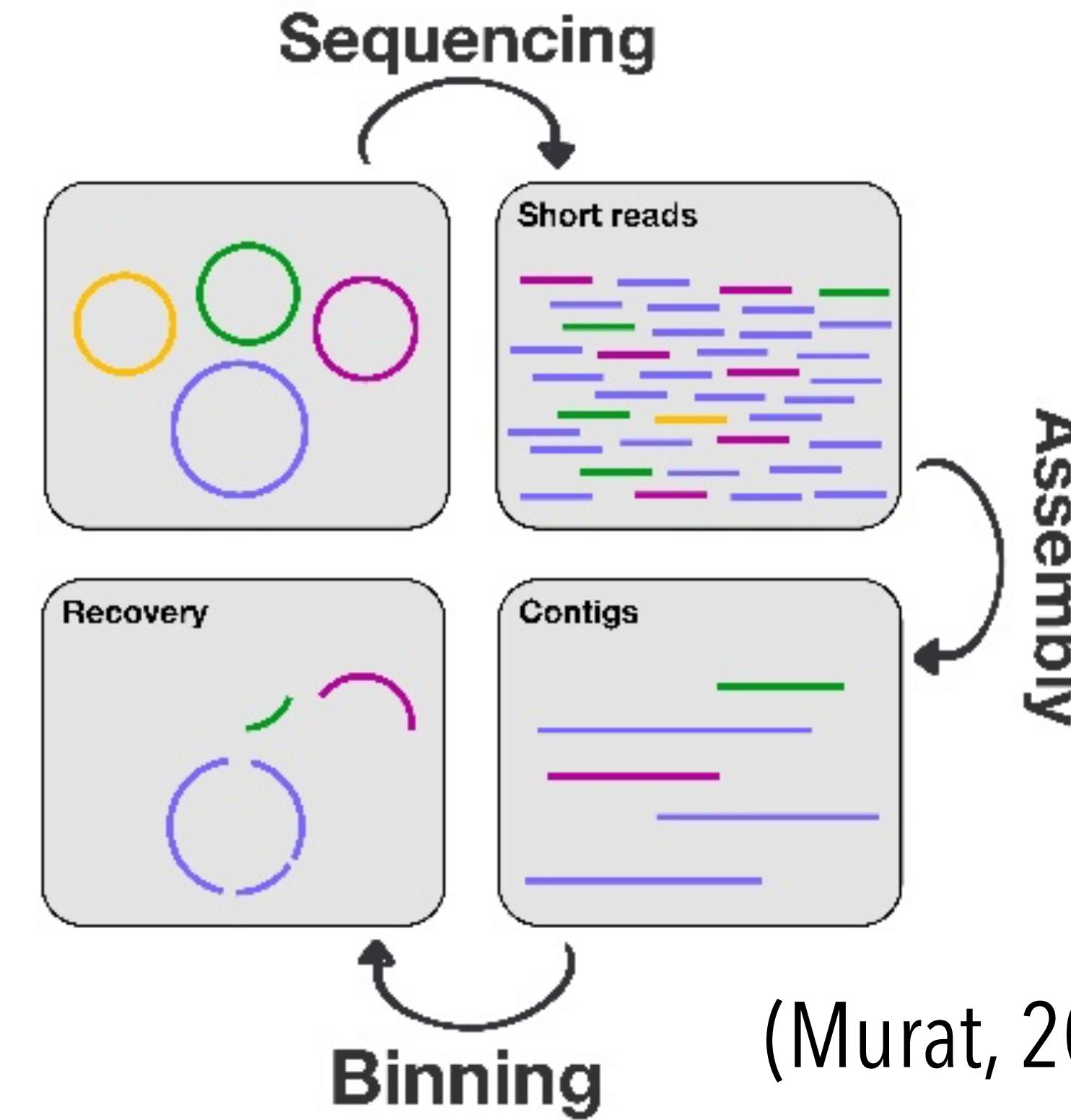
Cox *et al.* 2017

De-novo

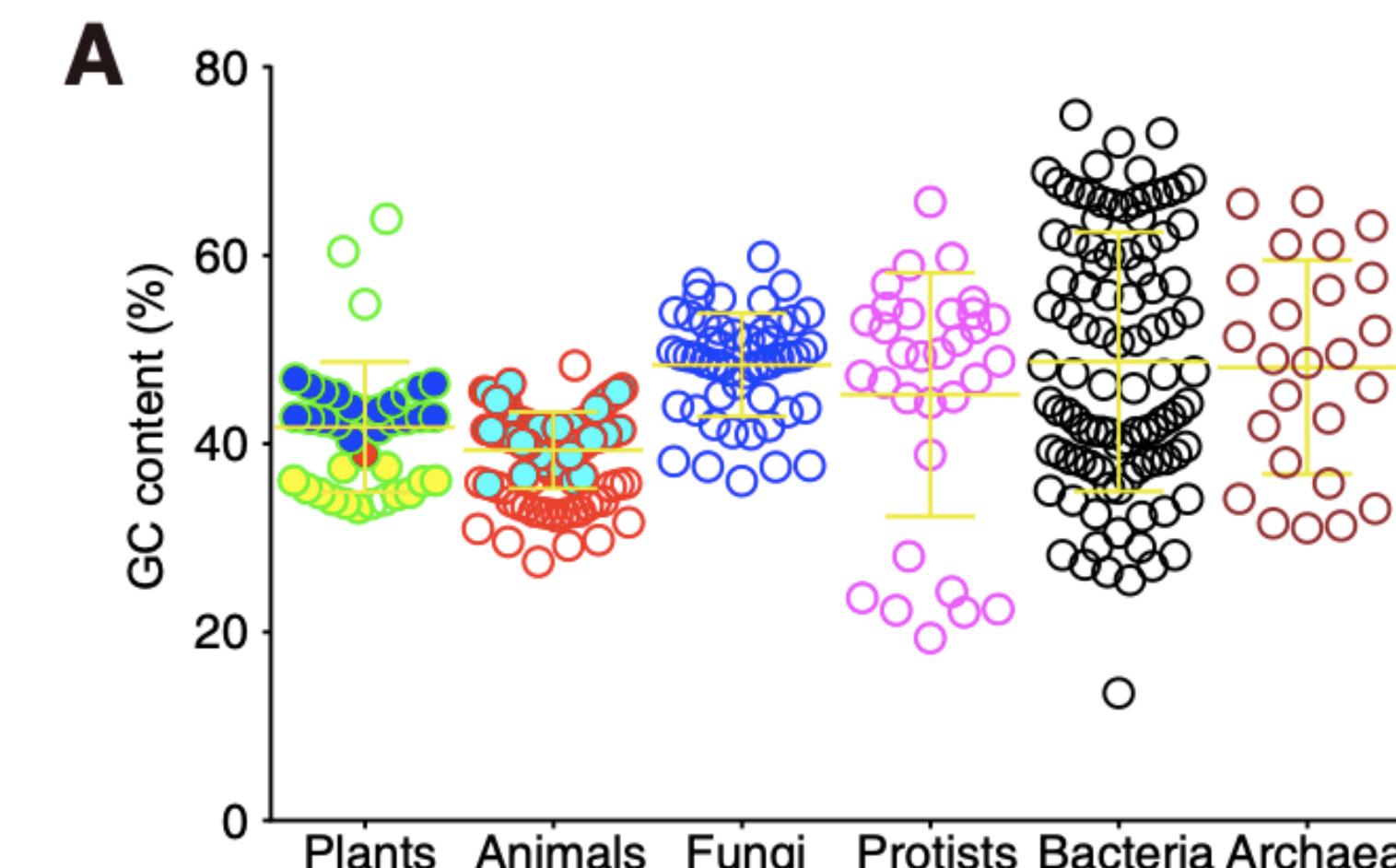
Assembly to short reads into longer sequences (**contigs**)

To assemble partial to full genomes of individual microorganism, contigs are sorted (**binned**) into putative transcriptomes which evaluate:

- Nucleotide composition
- Abundance patterns

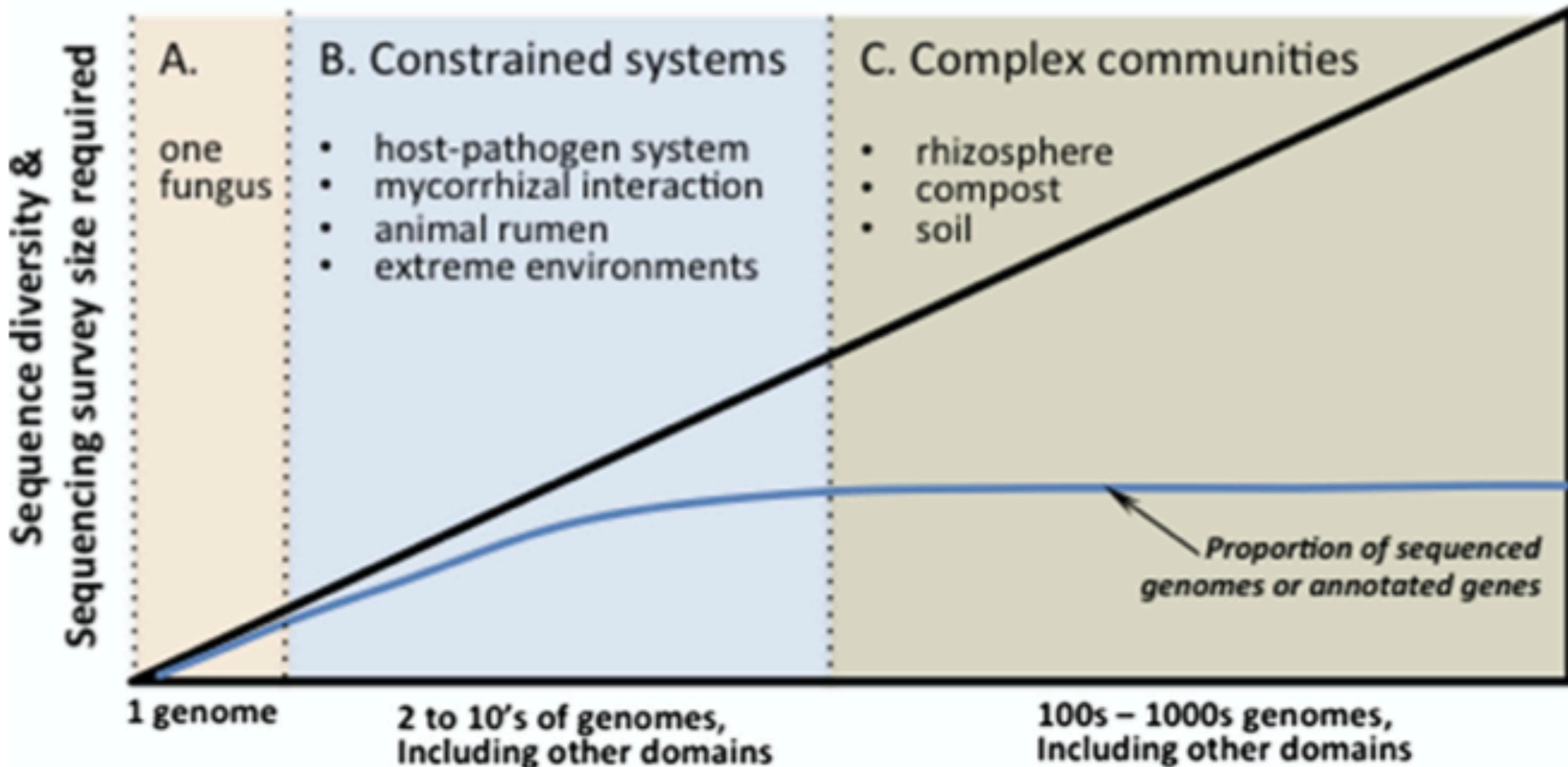


(Murat, 2020)



(Guo, 2017)

COMMUNITY COMPLEXITY



METATRANSCRIPTOME ANALYSES AND RESEARCH TOPICS

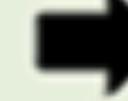
Single genome

- Map reads to genome
- Cell metabolism & regulation



de novo assembly

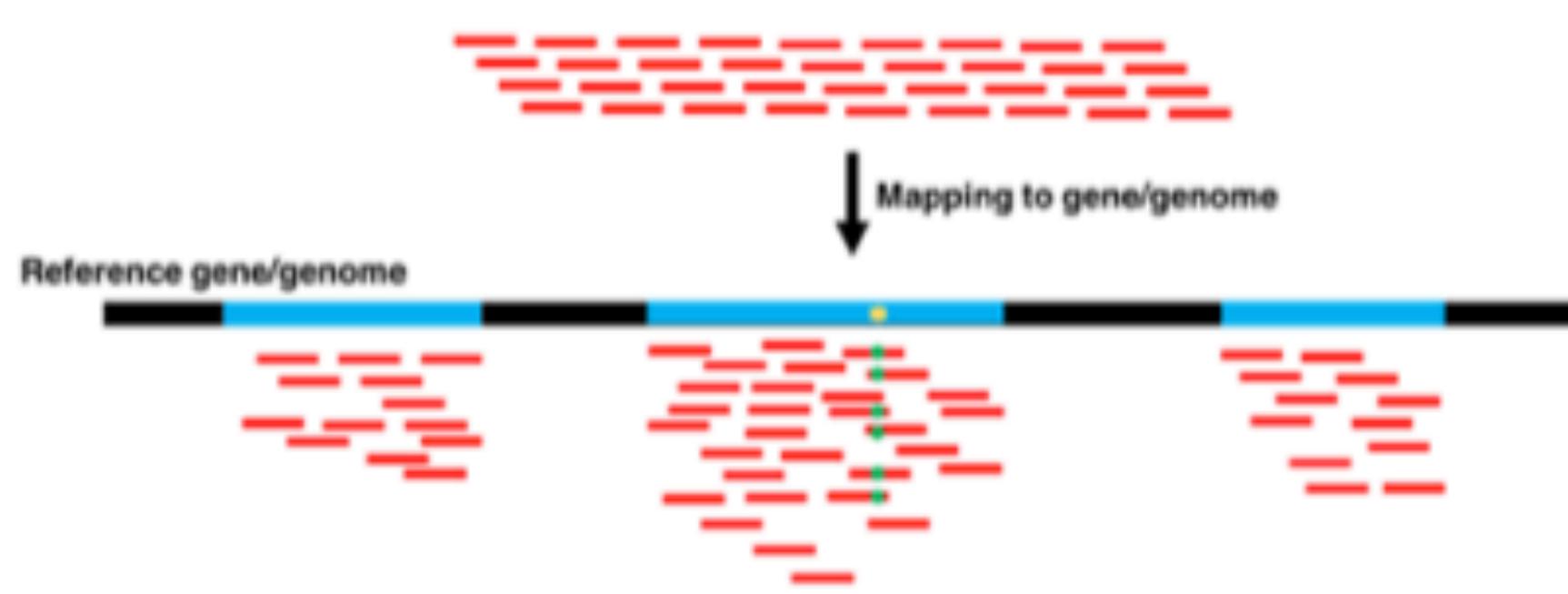
- Few or no sequenced genomes
- Target genes & pathways
- Species interactions & signaling



Read-based surveys

- Few or no sequenced genomes
- Target genes & pathways where databases exist
- Metabolic networks

Computational methods for metatranscriptome analysis



Read

- Much faster and scalable
- Detect known genes at low abundance
- Many reads are left uncharacterized
- Cannot detect novel genes



De-novo

- No reference genomes required
- Can detect genes from novel strains or species
- Assembly errors (chimeras, gene similarity)
- Difficulty reconstructing genes with low coverage

<http://ecoevo.unit.oist.jp/lab/wp-content/uploads/2013/08/GenomeAssembly.png>
https://galaxyproject.github.io/training-material/topics/proteomics/images/variant_calling.png

Figure 1: The two algorithmic approaches used in metatranscriptomics pipelines with their advantages and disadvantages.

TABLE 1 | A list of metatranscriptomics pipelines and their capabilities.

		Read based					Assembly based		
		MetaTrans	COMAN	FMAP	SAMSA2	HUMAnN2	SqueezeMeta	IMP	MOSCA
Preprocessing	QC	✓	✓	✓	✓	✗	✓	✓	✓
	Removes host reads	✗	✗	✓	✗	✗	✗	✓	✗
	Removes rRNA	✓	✓	✗	✓	✗	✓	✓	✓
<i>de novo</i> Assembly		✗	✗	✗	✗	✗	✓	✓	✓
Binning		✗	✗	✗	✗	✗	✓	✓	✗
Taxonomic	Reads	✓	✓	✗	✓	✓	✗	✗	✗
Profiling	Contigs	✗	✗	✗	✗	✗	✓	✓	✓
Functional	Reads	✓	✓	✓	✓	✓	✗	✗	✗
Annotation	Contigs	✗	✗	✗	✗	✗	✓	✓	✓
Pathway Analysis		✓	✓	✓	✗	✓	✓	✓	✗
Requires Metagenomes		✗	✗	✗	✗	✗	✗	✓	✗
Summary Report		✗	✗	✗	✗	✗	✗	✓	✗
Web Interface		✗	✓	✗	✗	✗	✗	✗	✗
Multiple Sample Comparisons		✓	✓	✓	✓	✓	✓	✗	✓
Differential Expression		✓	✓	✓	✓	✗	✗	✗	✓
Docker		✗	✗	✗	✗	✓	✗	✓	✓
Conda		✗	✗	✗	✗	✓	✗	✓	✗
Long Read Support		✗	✗	✗	✗	✗	✓	✗	✗
Public Code Repository		✓	✗	✓	✓	✓	✓	✓	✓

Metatranscriptomic approach for microbiome characterization and host gene expression evaluation for “Hoja de malvón” disease in *Vitis vinifera* cv. Malbec

CURRENT STATUS: POSTED



Marcos Paolinelli
CONICET Mendoza

✉ paolinellimarc@gmail.com *Corresponding Author*
ORCID: <https://orcid.org/0000-0003-0098-0646>

Georgina Escoriaza
Instituto Nacional de Tecnología Agropecuaria

Cecilia Cesari
Instituto Nacional de Tecnología Agropecuaria

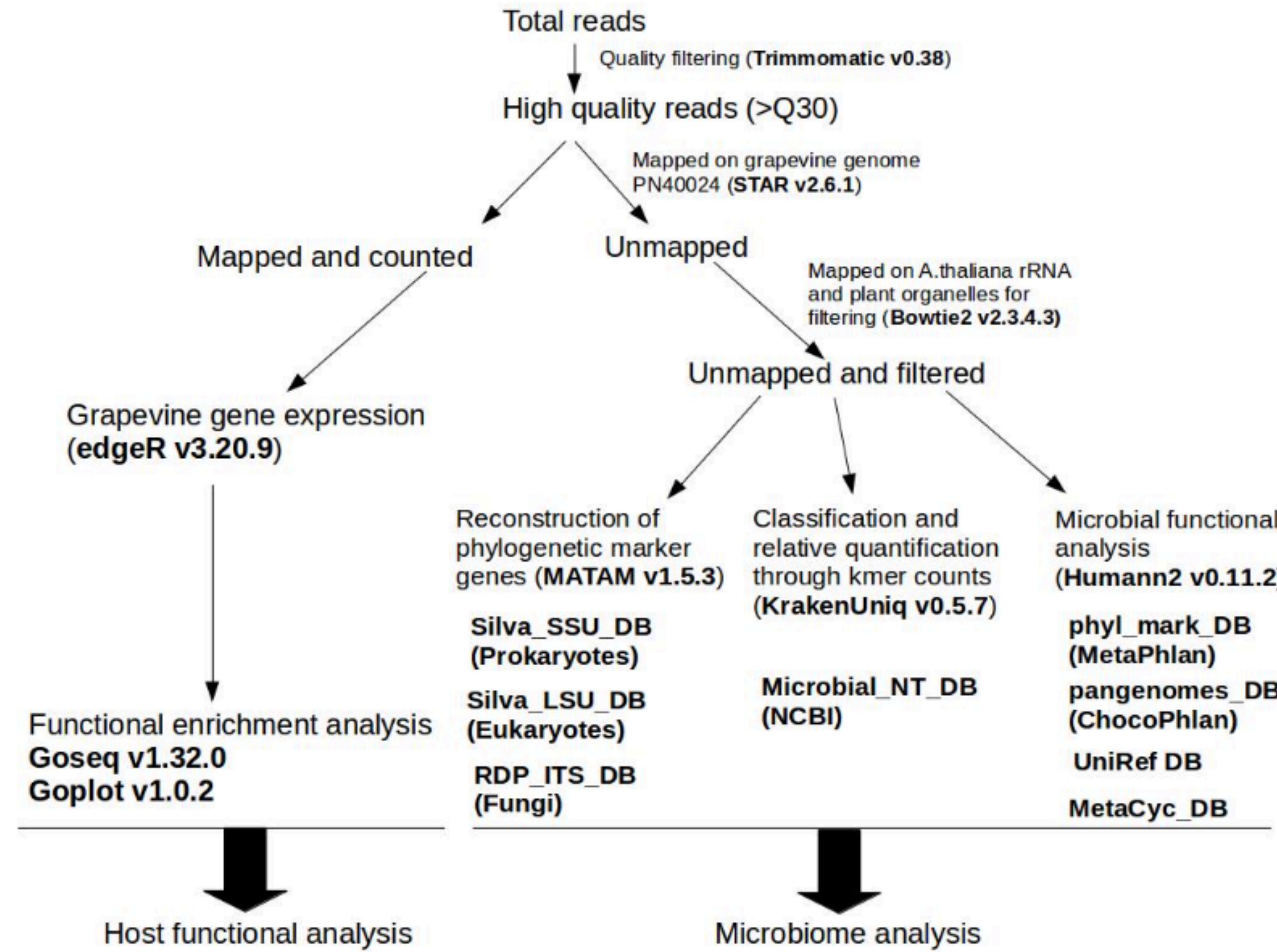
Sandra García-Lampasona
Instituto Nacional de Tecnología Agropecuaria

Rufina Hernández-Martínez
Centro de Investigación Científica y de Educación Superior de Ensenada

Grapevine Trunk Disease



Schematic summary of the analysis



(Paolinelli *et al.* *in press*)

Fungal Species classification and relative abundance according to the method used: MATAM vs. KrakenUniq

Fungal Genus Identified	MATAM-based percentage respect to the total abundance of Fungi	KrakenUniq-based percentage respect to the total abundance of Fungi
<i>Exophiala</i> sp.	7-15	0.4-0.5
<i>Morchella</i> sp.	3-5	0.05-0.09
<i>Cryptococcus</i> sp.	0-0.9	0.3 in all samples
<i>Rhodotorula</i> sp.	0-0.05	0.2 in all samples
<i>Leveillula</i> sp.	6-10	0-0.05
<i>Cladosporium</i> sp.	3-8	0.4-0.5
<i>Davidiella</i> sp.	3-9	0.01
<i>Fusarium</i> sp.	4-7	1
<i>Harzia</i> sp.	0-0.1	0.2
<i>Nigrospora</i> sp.	2-3	0.06-0.09
<i>Pestalotiopsis</i>	0.5-10	0.1-0.2
<i>Morinia</i> sp.	0-0.2	0.1-0.2
<i>Paraconiothyrium</i> sp.	0-0.07	0.4-0.6
<i>Periconia</i> sp. (<i>Massarina</i> sp. in MATAM)	15-32	0.07-0.1
<i>Coniothyrium</i> sp.	0-0.04	Not found
<i>Alternaria</i> sp.	23-43	1

References

- Anwar MZ, Lanzen A, Bang-Andreasen T, Jacobsen CS.** 2019. To assemble or not to resemble-A validated Comparative Metatranscriptomics Workflow (CoMW). *GigaScience* **8**: 1–10.
- Cox JW, Ballweg RA, Taft DH, Velayutham P, Haslam DB, Porollo A.** 2017. A fast and robust protocol for metataxonomic analysis using RNAseq data. *Microbiome* **5**: 1–13.
- Nakashima H, Ota M, Nishikawa K, Ooi, T.** 1998. Genes from nine genomes are separated into their organisms in the dinucleotide composition space. *DNA Research* **5**: 251–259.
- Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelle J, Gonzalez A, Kosciolak T, McCall LI, McDonald D, et al.** 2018. Best practices for analysing microbiomes. *Nature Reviews Microbiology* **16**: 410–422.
- Kuske CR, Hesse CN, Challacombe JF, Cullen D, Herr JR, Mueller RC, Tsang A, Vilgalys R.** 2015. Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecology* **14**: 133–137.
- Marcelino VR, Irinyi L, Eden JS, Meyer W, Holmes EC, Sorrell TC.** 2019. Metatranscriptomics as a tool to identify fungal species and subspecies in mixed communities – A proof of concept under laboratory conditions. *IMA Fungus* **10**: 1–10.
- Maric J, Šikic M.** 2019. Approaches to metagenomic classification and assembly. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*: 348–356.
- Paolinelli M, Escoriaza G, Cesari C, Garcia-Lampasona S, Hernandez-Martinez R.** *in press*. Metatranscriptomic approach for microbiome characterization and host gene expression evaluation for “Hoja de malvón” disease in *Vitis vinifera* cv. Malbec. : 1–41.
- Shakya M, Lo CC, Chain PSG.** 2019. Advances and challenges in metatranscriptomic analysis. *Frontiers in Genetics* **10**: 1–10.
- Wolff J, Batut B, Rasche H.** 2020. Mapping (Galaxy Training Materials). /training-material/topics/sequence-analysis/tutorials/mapping/tutorial.html Online; accessed Wed Oct 28 2020.

Metagenomics introduction

Murat E. 2020. Microbial ‘Omics for beginners. Week 3: Genome-resolved metagenomics: key concepts in reconstructing genomes from metagenomes. <https://www.youtube.com/watch?v=RjNdHGK4ruo&t=3079s> Online Wed Oct 28 2020.