



Metagenomics

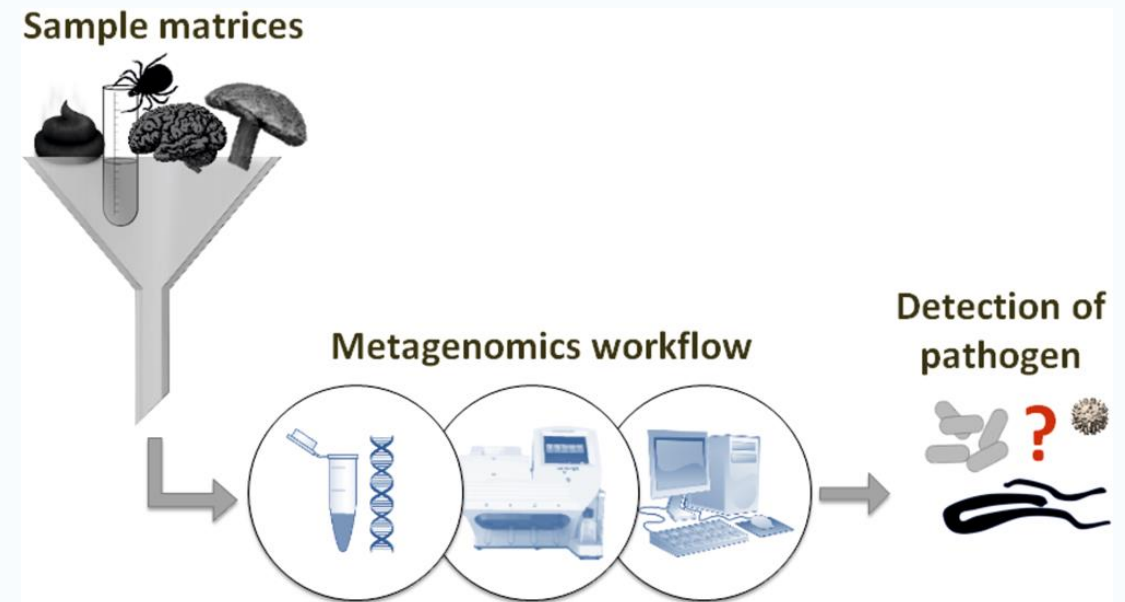
RGGS CG2
Lina Raubold

November 14th 2024

Overview

Definition – NIH:

Metagenomics is the study of the structure and function of entire nucleotide sequences isolated and analyzed from all the organisms (typically microbes) in a bulk sample. Metagenomics is often used to study a specific community of microorganisms, such as those residing on human skin, in the soil or in a water sample.



Wylezich et al., 2018 *Scientific Reports*

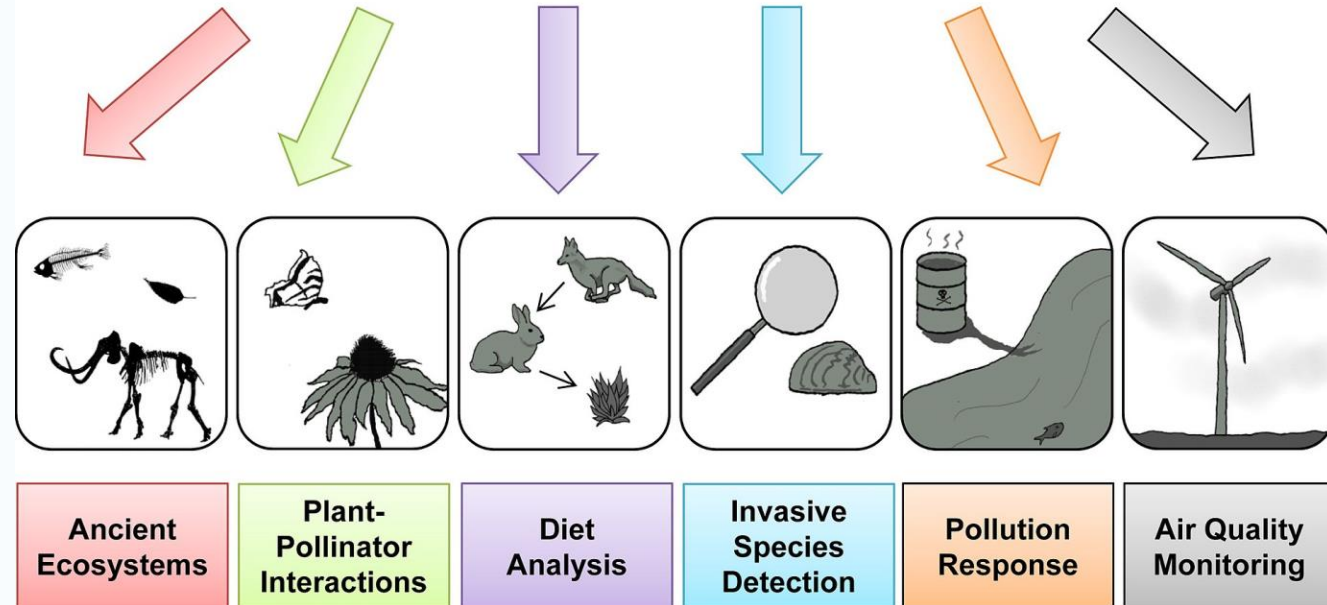
Metagenomics vs. Metabarcoding

- Omics approach to eDNA

Metagenomics:

- Untargeted
- Higher taxonomic resolution
- Functional annotation possible
- More costly
- Computationally more intensive

eDNA Metabarcoding Applications



Metagenomics vs. Metabarcoding taxonomic resolution

Metagenomics:

- Untargeted
- Higher taxonomic resolution



© Carel van der Colff
DiveInn Cape Town

iNaturalist

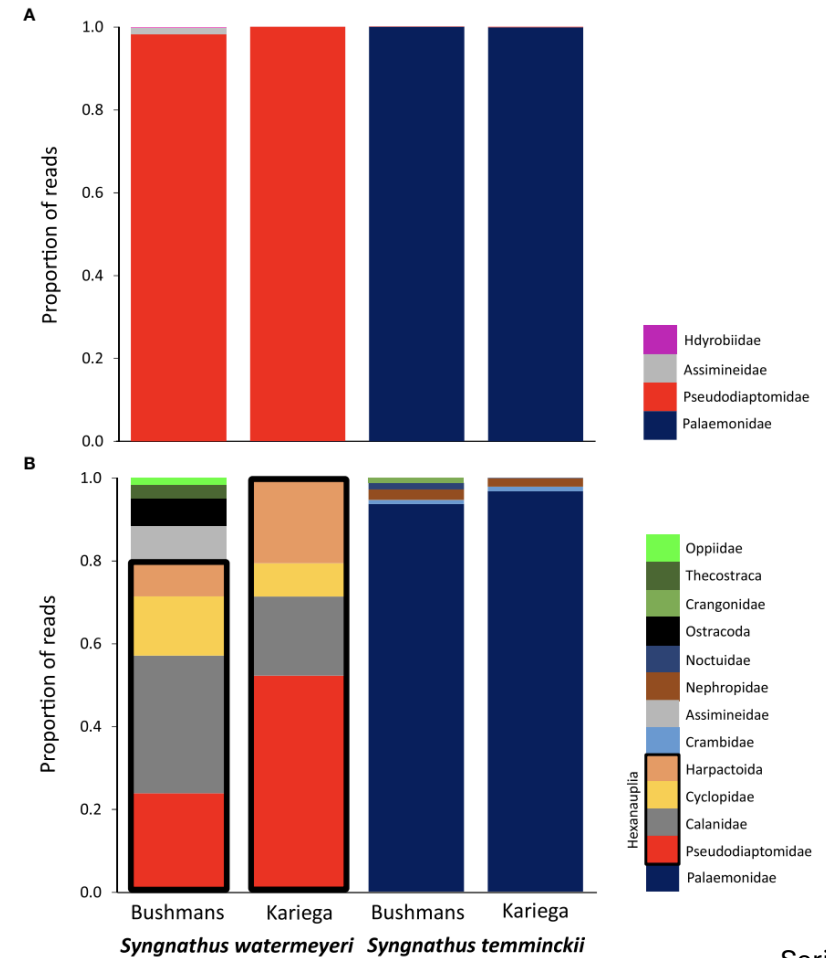
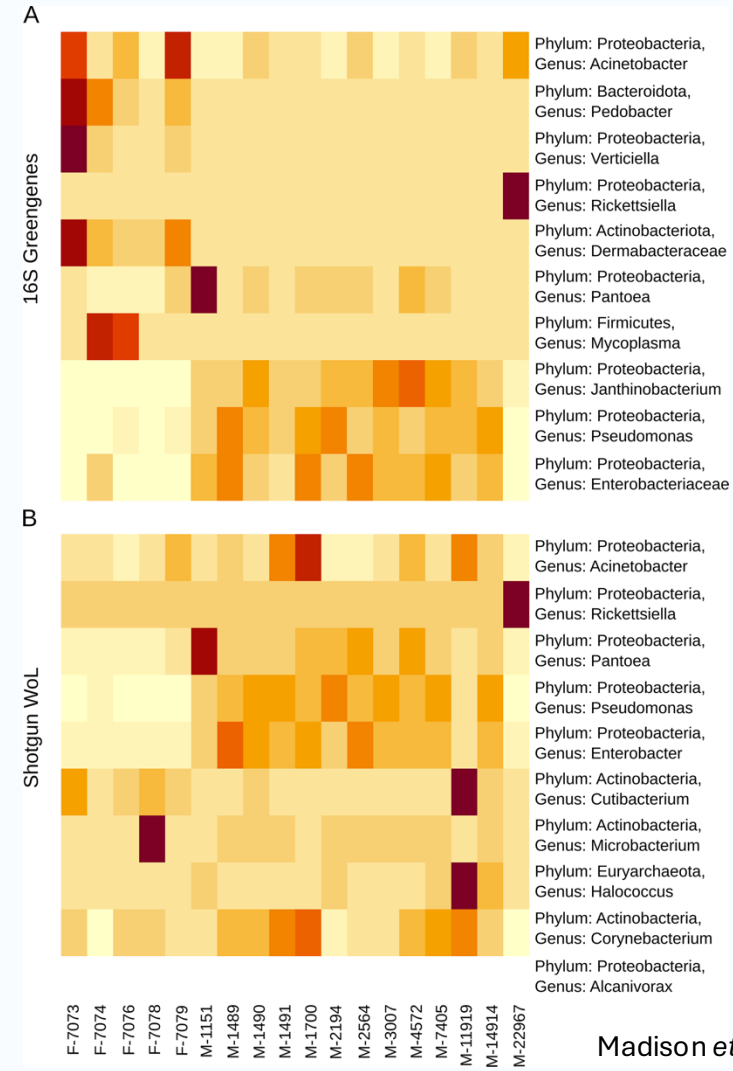
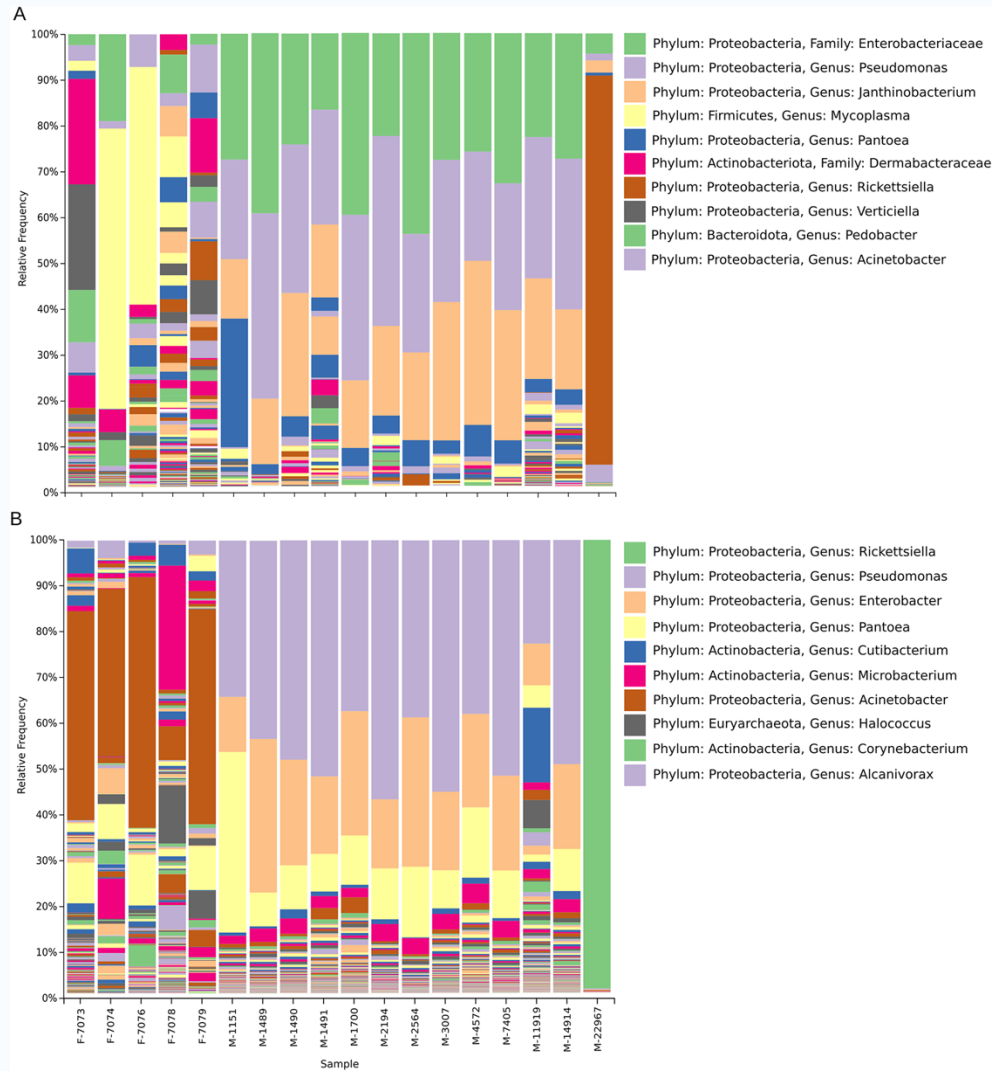


FIGURE 1

A comparison of the proportion of reads from putative prey families found in the feces of the two pipefish species; (A) metabarcoding data; (B) metagenomic data. In each case, species with a number of reads <1% of the total number of reads per fecal samples were excluded. The Hexanauplia (copepod) families found in the metagenomic data of *S. watermeyereri* are grouped together by black boxes.

Serite et al., 2023

Metagenomics vs. Metabarcoding taxonomic resolution & database bias



A detailed microscopic image showing various microorganisms, including bacteria, fungi, and possibly larger eukaryotic cells, in a light blue and white color palette. The organisms are scattered across the top half of the slide, with some showing internal structures and others appearing as simple shapes.

Applications

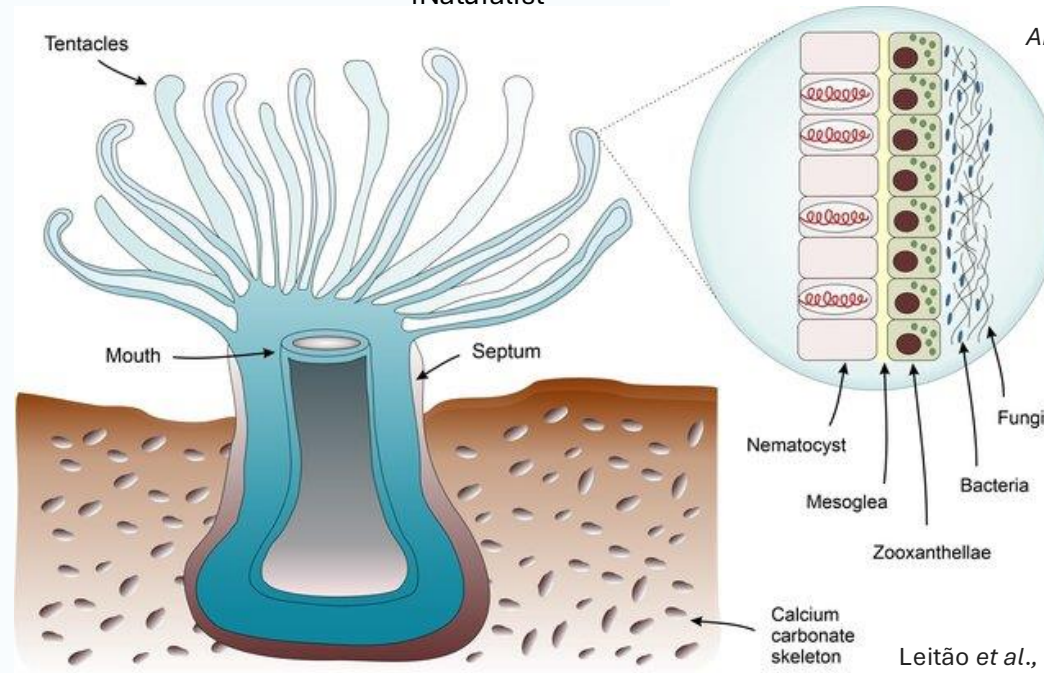
- Microbiology
 - analysis of microbial communities that can't be cultured
- Environmental monitoring
 - Which organisms are present in a particular environment?
 - How does the microbial community of an environment react to e.g., pollution?
- Medical research
 - microbiome analysis
 - pathogen detection
 - How does the microbiome influence different medical conditions?
- Agriculture
 - Are our soils healthy?
 - Which microbes help our agricultural plants grow?

Applications in Zoology

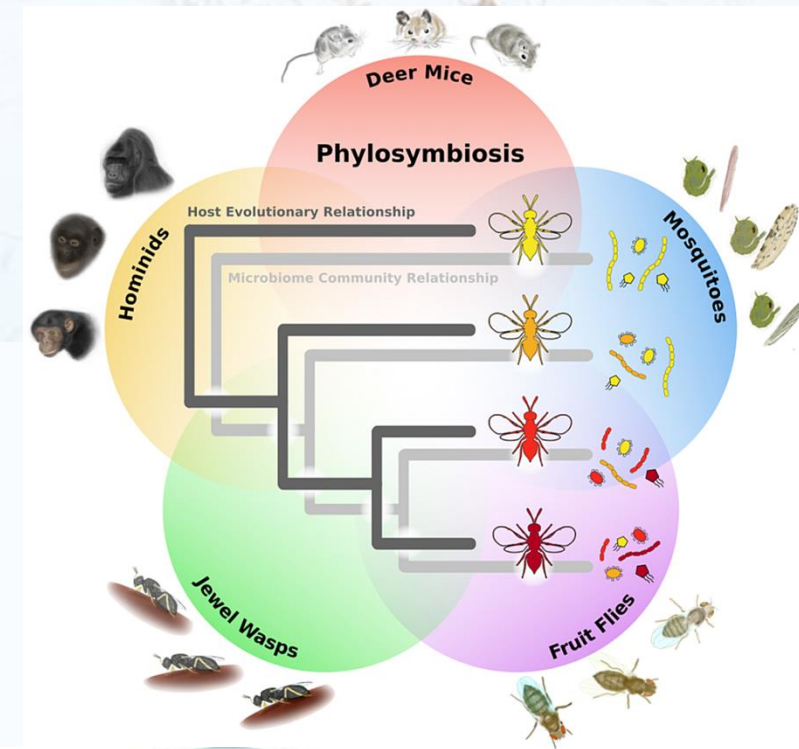
- Lineage-specific microbiomes
- Integrative zoology: holobionts
- Symbiotic relationships
- Trophic interactions
- Gut microbiome adaptations
- Wildlife monitoring
- Paleo-metagenomics



© Carel van der Colff
DiveInn Cape Town
iNaturalist



Leitão et al., 2020



Andrew Brooks in Richardson, 2017



Current limitations

- **Reference databases**

limited reference datasets for genomic data

- **Associated costs**

Shotgun metagenomic seq: \$139/sample

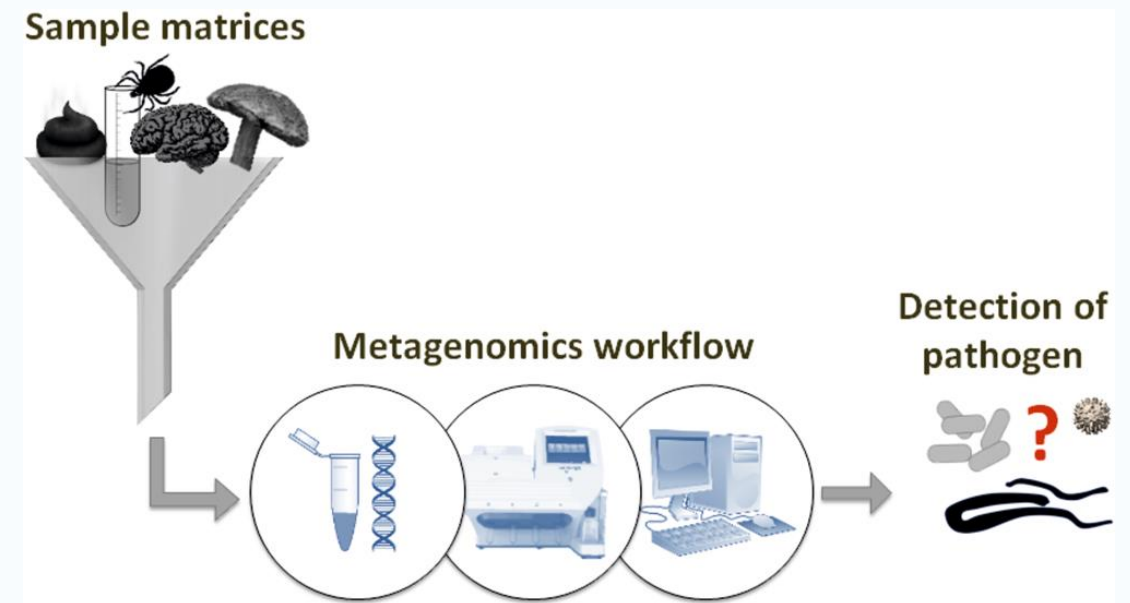
Amplicon sequencing: \$10 (library prep) + \$8 per 0.1M reads

- **Computational intensity**

larger datasets more complex to handle

Requirements – What do you need?

- Samples
- DNA extractions
e.g. Qiagen PowerSoil Pro
- Funding
- Computational infrastructure
- Analysis tools
 - Kraken2 + Bracken
 - Centrifuge
 - MetaPhiAn
 - Kaiju



Wylezich et al., 2018 Scientific Reports

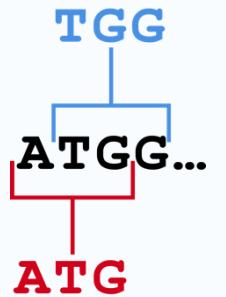
Kraken2 and Bracken

Kraken (taxonomic assignment)

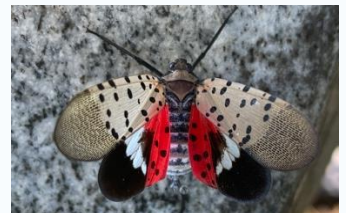
- Fast and memory efficient
- K-mer based approach
- K-mers in a query sequence mapped to lowest common ancestor (LCA) of all genomes that contain this k-mer
 - Unique hit -> taxonomic assignment (*L. meliae*/ *L. delicatula*)
 - Multiple hits -> taxonomic assignment to LCA (*Lycorma*)

Bracken

- Estimates relative abundances of species or genera based on estimation data from kraken



Lycorma meliae
iNaturalist



Lycorma delicatula
iNaturalist

Analysis workflow with Kraken2 and Bracken

- 1) Download Kraken2 database
Optional: Customize the database by adding specific sequences
- 2) Build Kraken2 and Bracken dbs
- 3) Run Kraken
- 4) Run Bracken
- 5) Calculate alpha-Div
- 6) Calculate beta-Div
- 7) Visualize in Pavian

