

Microbial communities through the lens of high throughput sequencing, data integration and metabolic networks analysis

developing computational approaches to better understand microbial assemblages

Haris Zafeiropoulos

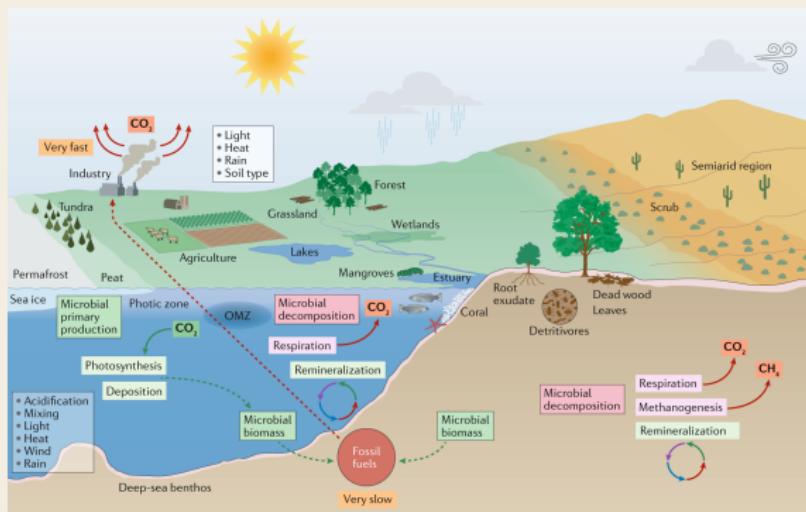
PhD candidate



- 1. Introduction**
- 2. Software development to support HTS-oriented methods for microbial diversity assessment**
 - 2.1 PEMA: a Pipeline for Environmental DNA Metabarcoding Analysis**

Microbial ecology & biogeochemical cycles

a corner-stone for life on earth



- composition
 - functions
 - interactions
-
- power biogeochemical cycling

Figure from: Nature Reviews Microbiology 17.9 (2019): 569-586.

Main questions regarding a microbial community for a deeper understanding of such assemblages



Community
structure
who

everyone is everywhere



Functional
potential
what

zero-sum game

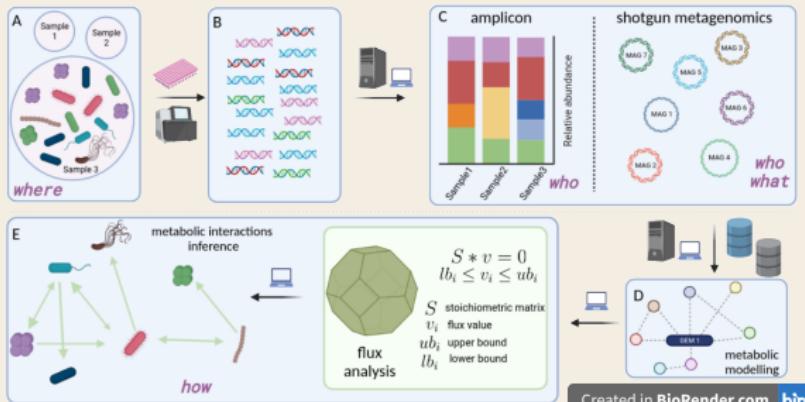


Microbial
interactions
why / how

the entangled bank

Reverse ecology

transforming ecology into a high-throughput field



community ecology studies with **no *a priori*** assumptions about the organisms under consideration, by exploiting **HTS** and **systems biology** approaches

From raw reads to community analysis

not a straight-forward way

- biology-oriented issues
- technology-oriented issues
- computing requirements
- data and metadata accessibility



Aims and objectives

- to enhance the analysis of microbiome data by building algorithms and software that address limitations and on-going computational challenges
- to exploit state-of-the-art methods to identify taxa and functions that play a key part in microbial community assemblages in hypersaline sediments

1. Introduction
2. Software development to support HTS-oriented methods for microbial diversity assessment
 - 2.1 PEMA: a Pipeline for Environmental DNA Metabarcoding Analysis

eDNA metabarcoding for biodiversity assessment

Marker genes

1. **16S rRNA**: Bacteria, Archaea
 2. **12S rRNA**: Vertebrates
 3. **18S rRNA**: Small eukaryotes, Metazoa
 4. **ITS**: Fungi
 5. **COI**: Eukaryotes
 6. **rbcl**: Plants
 7. **dsrb**: Bacteria, Archaea
 8. ...

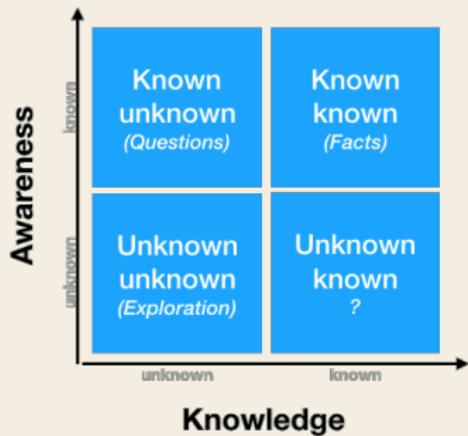
Methodology



- Sampling
 - Extraction
 - Bioinformatics
 - Biodiversity analysis

Bioinformatics challenges

for the analysis and the interpretation of amplicon data



PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis

Aim of the study and contribution



To build an open source pipeline that bundles state-of-the-art bioinformatics tools for all necessary steps of amplicon analysis and aims to address:

- one-stop-shop for several marker genes & approaches
- easy-to-set & easy-to-use
- scalable
- flexible
- reproducible

Results: PEMA features

an overview

PEMA in a nutshell

