ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

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

Haris Zafeiropoulos

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Science (PhD) in Biology

Promotors:

Prof. Emmanouil Ladoukakis
Dr Evangelos Pafilis
Dr Christoforos Nikolaou

Academic year 2021 – 2022

Members of the examination committee
&
reading committee

Prof. Emmanouil Ladoukakis
Univeristy of Crete
Biology Department

Dr Evangelos Pafilis
Hellenic Centre for Marine Research
Institute of Marine Biology, Biotechnology and Aquaculture

Dr Christoforos Nikolaou
Biomedical Sciences Research Center "Alexander Fleming"
Institute of Bioinnovation

Please let me know if you would like to suggest someone in particular! Here are some thoughts of mine for the rest of the committe, feel free to share yours!
I am considering of asking **Dr Jens Carlsson** who is familiar with my work on COI as well as **Prof Faust** that I will join her lab for a few months (for an EMBO short term fellowsip)
Also, **Prof. Elias Tsigaridas** that we have worked together on flux sampling.
Finally, in case that it is ok to have non permanent researchers in the committe, **Dr. Christina Pavloudi** with whom I have worked all these years.

Preface

Haris Zafeiropoulos

Contents

Abstract

Περίληψη

Και στα ελληνικά

List of Figures and Tables

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

NGS	Next Generation Sequencing
HPC	High Performance Computing
MCMC	Markov Chain Monte Carlo
MMCS	Multiphase Monte Carlo Sampling
PREGO	PRocess Environment OrGanism
PEMA	Pipeline for Environmental DNA Metabarcoding Analysis
DARN	Dark mAtteR iNvestigator

Chapter 1

Introduction

1.1 Microbial communities: structure & function

1.1.1 The role of microbial communities in biogeochemical cycles

1.1.2 HTS approaches to study the microbiome

1.2 Microbial interactions: the way to unravel the microbiome

1.2.1 Methods for microbial interactions inference

1.2.2 Competition and mutualism: a dialectic relationship

1.3 The hypersaline Tristomo swamp: a case study of an extreme environment

1.4 Aims and objectives

Chapter 2

Software development to establish quality HTS-oriented bioinformatics methods for microbial diversity assessment

2.1 Environmental DNA metabarcoding: challenges and caveats

2.2 PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes

Publication relative to this chapter: [1].

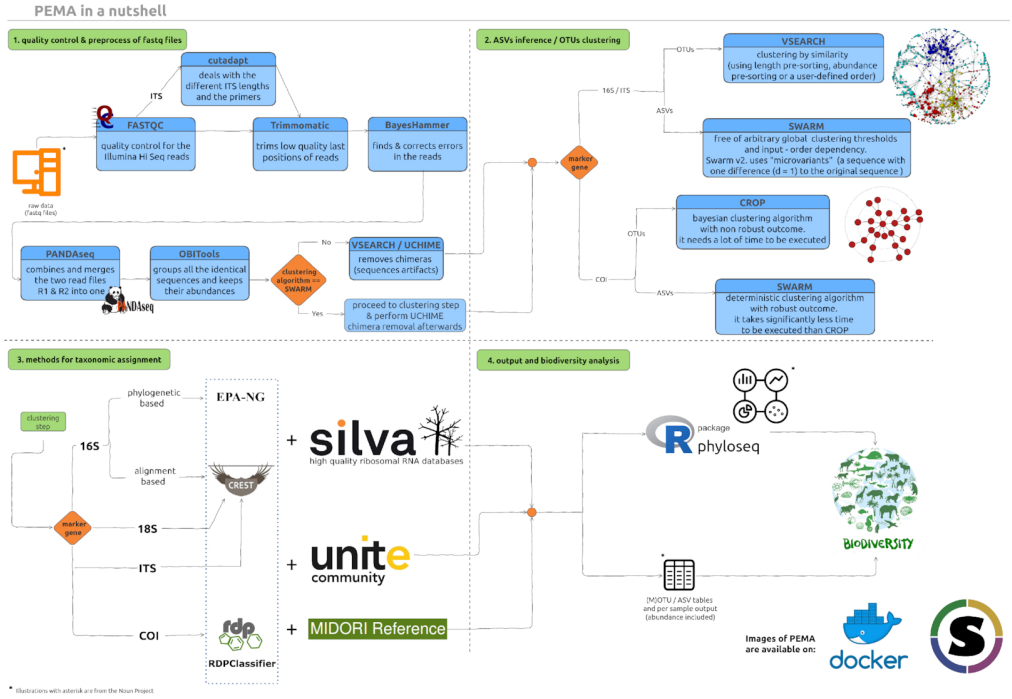

FIGURE 2.1: The PEMA workflow: figure from publication

2.3 The Dark mAtteR iNvestigator (DARN) tool: getting to know the known unknowns in COI amplicon data

Publication relative to this chapter: [2]

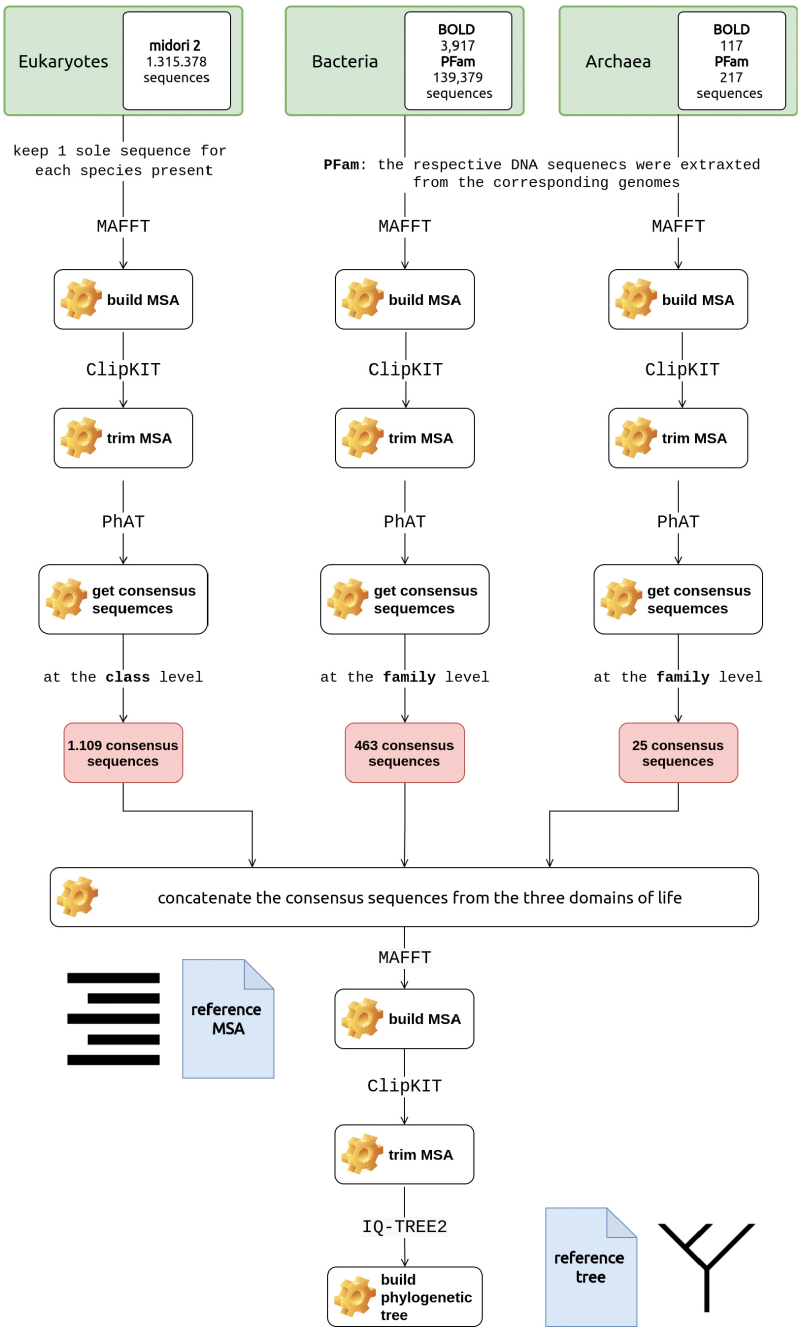

FIGURE 2.2: DARN methodology: figure in the publication

2.4 A workflow for marine Genomic Observatories data analysis

Chapter 3

Software development to build a knowledge-base at the systems biology level

3.1 Metadata: a key issue for robust metanalyses

3.2 Ontologies & databases: the corner stone of mordern biology

3.3 PREGO: a literature- and data-mining resource to associate microorganisms, biological processes, and environment types

Publication relative to this chapter: under submission

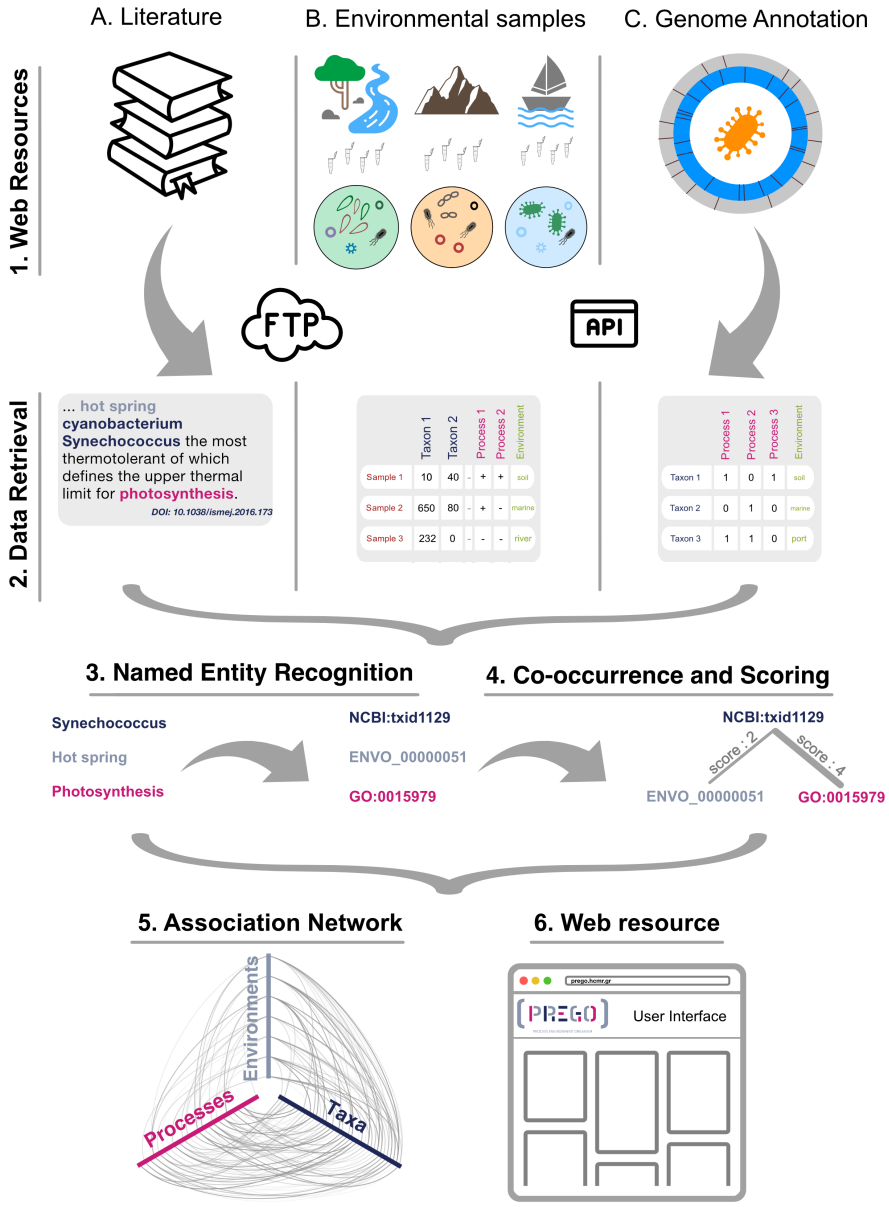

FIGURE 3.1: PREGO methodology: figure in the publication under submission

Chapter 4

Software development to establish metabolic flux sampling approaches at the community level

4.1 Genome-scale metabolic model analysis

4.2 A New MCMC Algorithm for Sampling the Flux Space of Metabolic Networks

Publication relative to this chapter: [3]

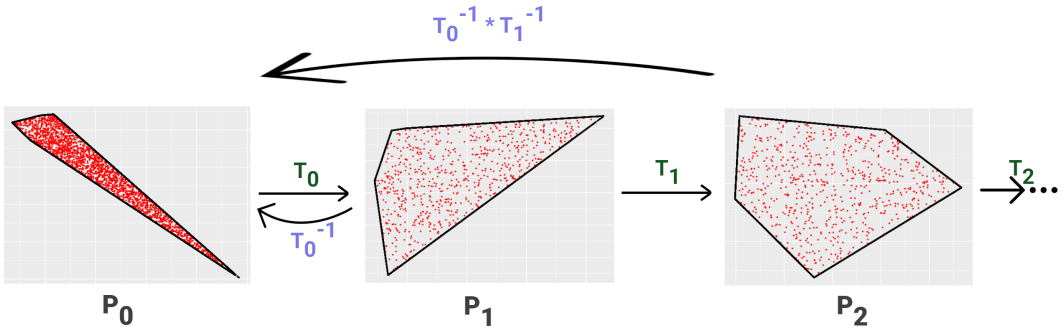

FIGURE 4.1: Our MMCS algorithm and its first phases. Figure published on SoCG21

4.3 Flux sampling at the community level

Chapter 5

Microbial interactions inference in communities of a hypersaline swamp elucidate mechanisms governing taxonomic & functional profiles

Publication relative to this chapter: ongoing work, to be submitted before phd defense, probably not accepted by then though.

5.1 Amplicon & shotgun metagenomic analysis

darn and PEMA will be used at this point, among other software

5.2 Inferring microbial interactions

PREGO and dingo will be used to this end

Chapter 6

0s and 1s in marine molecular research

Publication relative to this chapter: [4]

6.1 Computing resources: a prerequisite & a limitation in modern microbial ecology

6.2 High Performance Computing and Cloudification: scaling up bioinformatics analysis

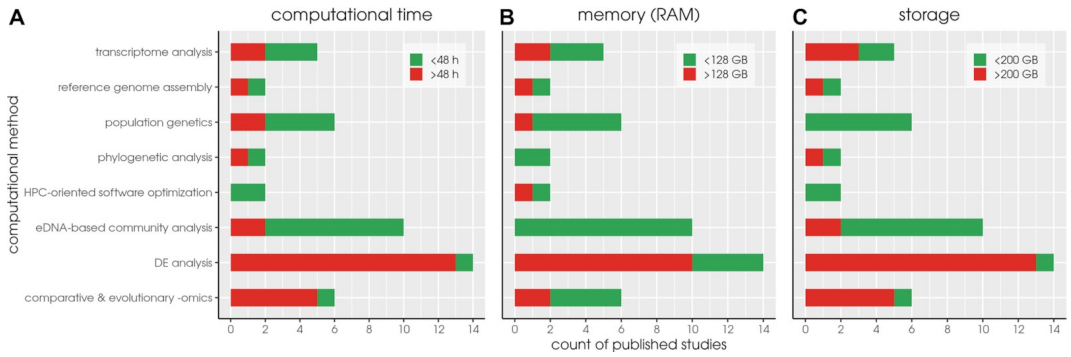

FIGURE 6.1: Computing requirements of the published studies performed on the IMBBC HPC facility over the last decade. Figure from publication.

Chapter 7

Conclusions

Appendices

Bibliography

[1] H. Zafeiropoulos, H. Q. Viet, K. Vasileiadou, A. Potirakis, C. Arvanitidis, P. Topalis, C. Pavloudi, and E. Pafilis, "Pema: a flexible pipeline for environmental dna metabarcoding analysis of the 16s/18s ribosomal rna, its, and coi marker genes," *GigaScience*, vol. 9, no. 3, p. giaa022, 2020.

[2] H. Zafeiropoulos, L. Gargan, S. Hintikka, C. Pavloudi, and J. Carlsson, "The dark matter investigator (darn) tool: getting to know the known unknowns in coi amplicon data," *Metabarcoding and Metagenomics*, vol. 5, p. e69657, 2021.

[3] A. Chalkis, V. Fisikopoulos, E. Tsigaridas, and H. Zafeiropoulos, "Geometric Algorithms for Sampling the Flux Space of Metabolic Networks," in *37th International Symposium on Computational Geometry (SoCG 2021)* (K. Buchin and E. Colin de Verdière, eds.), vol. 189 of *Leibniz International Proceedings in Informatics (LIPIcs)*, (Dagstuhl, Germany), pp. 21:1–21:16, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.

[4] H. Zafeiropoulos, A. Gioti, S. Ninidakis, A. Potirakis, S. Paragkamian, N. Angelova, A. Antoniou, T. Danis, E. Kaitetzidou, P. Kasapidis, *et al.*, "0s and 1s in marine molecular research: a regional hpc perspective," *GigaScience*, vol. 10, no. 8, p. giab053, 2021.

PhD disseration

Student: Haris Zafeiropoulos

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

UDC: 621.3

Korte inhoud:

Hier komt een heel bondig abstract van hooguit 500 woorden. LaTeX commando's mogen hier gebruikt worden. Blanco lijnen (of het commando \par) zijn wel niet toegelaten!

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Science (PhD) in Biology

Promotors: Prof. Emmanouil Ladoukakis
Dr Evangelos Pafilis
Dr Christoforos Nikolaou

:

: