ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

Microbial communities through the lens of data integration, knowledge aggregation 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

Dr Jens Carlsson
University College Dublin
School of Biology and Environmental Science/Earth Institute

Dr Christina Pavloudi
George Washington University, US

Prof Elias Tsigaridas
Sorbonne Université and Paris Université
Inria Paris and IMJ-PRG

Prof Karoline Faust
KU Leuven
Department of Microbiology and Immunology, Rega Institute

Preface

Hello friend.

Haris Zafeiropoulos

Contents

Abstract

The abstract environment contains a more extensive overview of the work. But it should be limited to one page.

Περίληψη

γεια σου φίλε ..

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

Symbols

42	"The Answer to the Ultimate Question of Life, the Universe, and Everything" according to [**?**]
c	Speed of light
E	Energy
m	Mass
π	The number pi

Chapter 1

Introduction

The first contains a general introduction to the work. The goals are defined and the modus operandi is explained.

1.1 Microbes and their functions..

NGS –> breakthrough in what we can see Taxonomy
Extreme environments

1.2 .. make the world go round!

* ecosystem functioning

1.3 Microbial interactions

Evolution

Chapter 2

Microbial diversity: *who*

This chapter will be about finding the taxa present in an environment sample.

First we will discuss a few things about the biodiversity assessment methods in general in terms of a short introduction.

2.1 Metabarcoding..

Then we will describe PEMA [5]

2.2 .. has caveats

And here we will talk about DARN

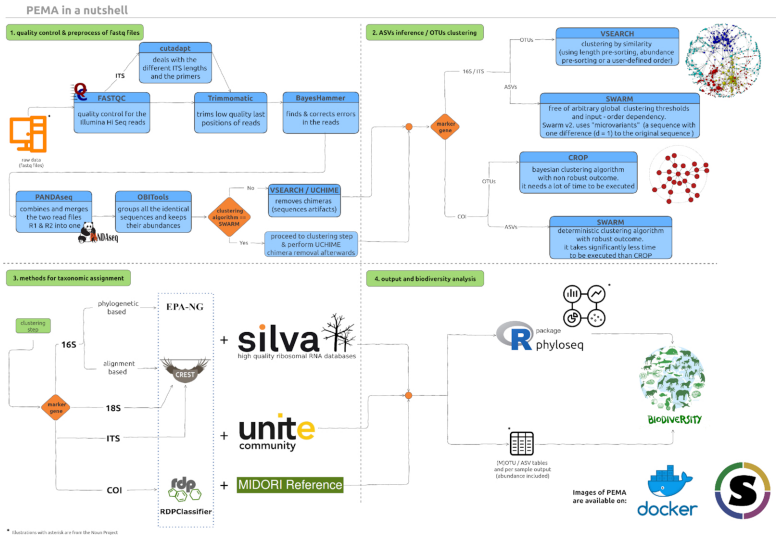

FIGURE 2.1: workflow from publication

Resources	bacteria		archaea	
	# of sequences	# of strains	# of sequences	# of strains
BOLD	3,917	2,267	117	117
PFam-oriented	9,154	4,532	217	115
Total unique entries	11,421	6,798	334	201

TABLE 2.1: Number of sequences and taxonomic species per domain of life and resources. The (#) symbols stands for "number".

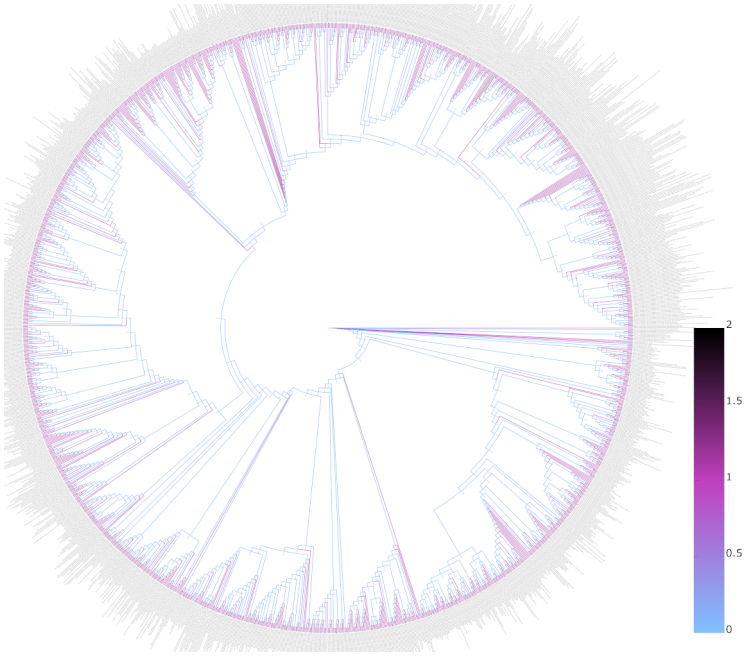

FIGURE 2.2: Placements of the consensus sequences used to build the COI reference phylogenetic tree for the DARN tool, onto the phylogenetic tree (stroke width for the branches of the tree is 5). The color coding represents the placements per branch, with a range from zero (blue) to a maximum of 2 (blue). The 1 leaf – 1 placement relationship, as well as the maximum of 2 placements in the color coding bar, indicate the proper placement of each consensus sequence to its corresponding branch.

2.3 What about metagenomics?

2.3.1 Afoulo-iky

2.3.2 EOSC Life project

And at this point we ll mention our work and findings (if any) in the framework of the EOSC Life project.

Chapter 3

Ecosystem functioning: the *what*, the *where*

3.1 Data integration, Knowledge aggregation

PREGO will be described here

Chapter 4

Microbial interactions: the *why*

4.1 Metabolism as the corner stone

The relationship between genotype and phenotype is fundamental to biology. Many levels of control are introduced when moving from one to the other. Systems biology aims at deciphering "the strategy" both at the cell and at higher levels of organization, in case of multicell species, that enables organisms to produce orderly adaptive behavior in the face of widely varying genetic and environmental conditions ([4]); the term "strategy" is used as per [2]. Systems biology approaches aim at interpreting how a system's properties emerge; from the cell to the community level.

4.1.1 Genome-scale metabolic network reconstruction

[1]

4.2 The flux sampling approach

From Price et al. [3] : "Pairwise correlation coefficients can be calculated between all reaction fluxes based on uniform random sampling. Perfectly correlated reactions (R2 = 1) operate as functional modules within a biochemical network, whereas uncorrelated reactions (R2 0) operate independently of each other. The degree of independence between reactions is an important consideration when choosing a set of fluxes to measure that will best determine the operating state of a biochemical network"

Write something from Polanyi

4.3 The 'dingo' Python library

our approach

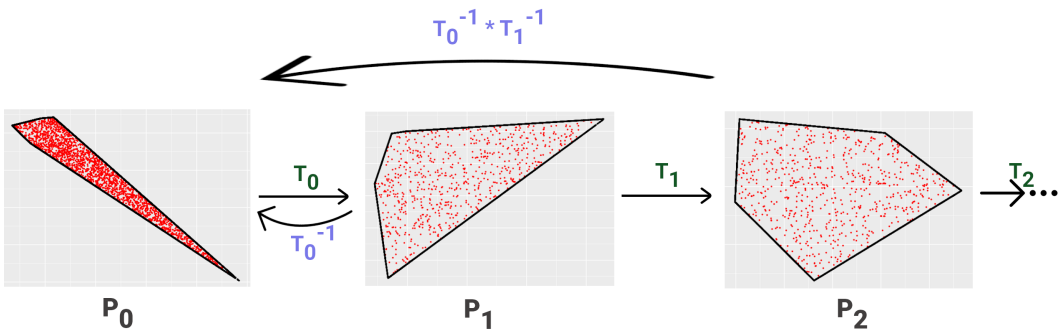

FIGURE 4.1: Our MMCS algorithm and its first phases

Chapter 5

Diving into (the dirt of) a swamp

5.1 A metagenome study..

5.2 ..to see how they can live there!

FIGURE 5.1: The KU Leuven logo.

gnats	gram	$13.65
	each	.01
gnu	stuffed	92.50
emu		33.33
armadillo	frozen	8.99

TABLE 5.1: A table with the wrong layout.

Chapter 6

Not the sky, but the computing resources is now the limit

6.1 HPC solutions

HPC paper

6.2 white paper of Elixir microbiome community

Infrustuctures could be of use

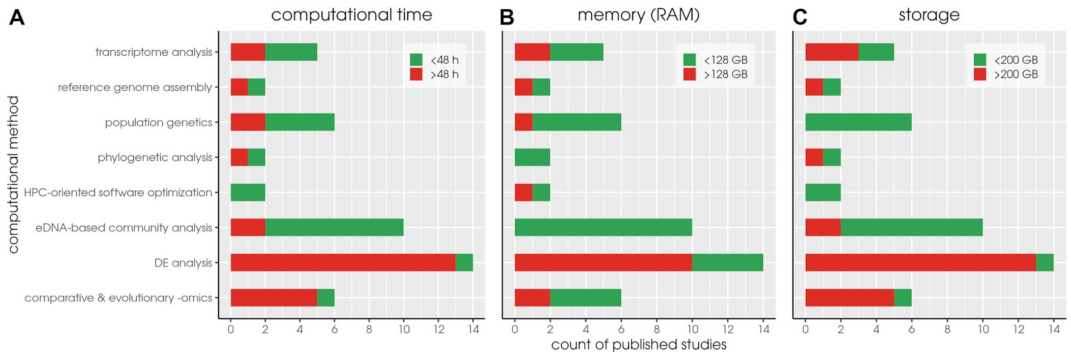

FIGURE 6.1: Red bars denote published research with high resource requirements of the various computational methods employed at the IMBBC HPC facility due to (a) long computational times (>48 h), (b) high memory requirements (>128 GB), or (c) high storage requirements (>200 GB). For instance, no eDNA-based community analyses performed at Zorba thus far have required a large amounts of memory.

Chapter 7

Conclusion

The final chapter contains the overall conclusion. It also contains suggestions for future work and industrial applications.

Appendices

Bibliography

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

[2] Polanyi, M. (1968). Life's irreducible structure: Live mechanisms and information in dna are boundary conditions with a sequence of boundaries above them. *Science*, 160(3834):1308–1312.

[3] Price, N. D., Reed, J. L., and Palsson, B. Ø. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11):886–897.

[4] Strohman, R. (2002). Maneuvering in the complex path from genotype to phenotype. *Science*, 296(5568):701–703.

[5] Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., Pavloudi, C., and Pafilis, E. (2020). Pema: a flexible pipeline for environmental dna metabarcoding analysis of the 16s/18s ribosomal rna, its, and coi marker genes. *GigaScience*, 9(3):giaa022.

PhD disseration

Student: Haris Zafeiropoulos

Titen: Microbial communities through the lens of data integration, knowledge aggregation 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

:

: