



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 2020 – 2021

**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

Preface	i
Abstract	iii
Περίληψη	iv
List of Figures and Tables	v
List of Abbreviations and Symbols	vi
1 Introduction	1
1.1 Microbes and their functions..	1
1.2 .. make the world go round!	1
1.3 Microbial interactions	1
2 Microbial diversity: <i>who</i>	3
2.1 Metabarcoding..	3
2.2 .. has caveats	3
2.3 What about metagenomics?	3
3 Ecosystem functioning: the <i>what</i>, the <i>where</i>	5
3.1 Data integration, Knowledge aggregation	5
4 Microbial interactions: the <i>why</i>	7
4.1 Metabolism as the corner stone	7
4.2 The flux sampling approach	7
4.3 The 'dingo' Python library	7
5 Diving into (the dirt of) a swamp	9
5.1 A metagenome study..	9
5.2 ..to see how they can live there!	9
6 Not the sky, but the computing resources is now the limit	11
6.1 HPC solutions	11
6.2 white paper of Elixir microbiome community	11
7 Conclusion	13
Bibliography	17

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

2.1	workflow from publication	3
2.2	COI - reference phylogenetic tree	3
5.1	The KU Leuven logo.	9

List of Tables

5.1	A table with the wrong layout.	9
-----	--	---

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

2.2 .. has caveats

And here we will talk about DARN

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.

FIGURE 2.1: workflow from publication

FIGURE 2.2: COI - reference phylogenetic tree

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 ([3]); the term "strategy" is used as per [1]. 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

4.2 The flux sampling approach

From Price et al. [2] : "Pairwise correlation coefficients can be calculated between all reaction fluxes based on uniform random sampling. Perfectly correlated reactions ($R^2 = 1$) operate as functional modules within a biochemical network, whereas uncorrelated reactions ($R^2 = 0$) 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"

Polanyi

4.3 The 'dingo' Python library

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

Chapter 7

Conclusion

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

Appendices

Bibliography

- [1] Michael Polanyi. 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, 1968.
- [2] Nathan D Price, Jennifer L Reed, and Bernhard Ø Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11):886–897, 2004.
- [3] Richard Strohman. Maneuvering in the complex path from genotype to phenotype. *Science*, 296(5568):701–703, 2002.

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. ~~TEX~~ commando's mogen hier gebruikt worden. Blanco lijnen (of het commando `\par`) zijn wel niet toegelaten!

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

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

Promoters: Prof. Emmanouil Ladoukakis

Dr Evangelos Pafilis

Dr Christoforos Nikolaou

:
: