

Google Summer of Code 2021  
CODING PROJECT PROPOSAL  
Sampling on the flux space of metabolic networks to extract key  
processes linking multiple

**Name:** Haris Zafeiropoulos  
**Affiliation:** Department of Biology, University of Crete  
**Program:** PhD candidate, Second participation in GSoC  
**Mentors:** Elias Tsigaridas, Apostolos Chalkis, Zafeirakis Zafeirakopoulos  
**email:** haris.zafr@gmail.com  
**GitHub:** <https://github.com/hariszaf>  
**Address:** Valestra 99, Heraklion, Crete, Greece, 71202  
**Phone:** +30 694 909 3089

April 10, 2021

## Contents

<b>1</b>	<b>Synopsis</b>	<b>1</b>
<b>2</b>	<b>The Project</b>	<b>1</b>
<b>3</b>	<b>Biographical information</b>	<b>1</b>
3.1	Education . . . . .	1
3.2	Publications . . . . .	2
3.3	Programming . . . . .	2
<b>4</b>	<b>Personal motivation</b>	<b>2</b>

## 1 Synopsis

## 2 The Project

## 3 Biographical information

### 3.1 Education

- PhD candidate at University of Crete (2018 - ongoing). Dissertation on: “Merging NGS data, knowledge aggregation and data integration techniques, along with ecological network analysis (ENA): an attempt to decipher microbial community ecology and ecosystem functioning by taking advantage of the hypothesis-generating method”.

- MSc in Bioinformatics at the University of Crete (2016 - 2018). Thesis: “eDNA metabarcoding for biodiversity assessment: Algorithm design and bioinformatics analysis pipeline implementation”
- BSc in Biology at the National and Kapodistrian University of Athens (2010 - 2016). Thesis: “Morphology, morphometry and anatomy of species of the genus *Pseudamnicola* in Greece”

## 3.2 Publications

- Zafeiropoulos, Haris, Anastasia Gioti et al. 0s and 1s in marine molecular research: a regional HPC perspective (**under review in GigaScience journal**)
- Zafeiropoulos, Haris, Anastasia Gioti et al. (2021, April 5). The IMBBC HPC facility: history, configuration, usage statistics and related activities (Version 1.0.0). Zenodo. DOI: 10.5281/zenodo.4665308
- Polymenakou, Paraskevi N., et al. "The Santorini Volcanic Complex as a Valuable Source of Enzymes for Bioenergy." *Energies* 14.5 (2021): 1414. DOI: 10.3390/en14051414
- Chalkis, Apostolos, et al. "Geometric algorithms for sampling the flux space of metabolic networks." arXiv preprint arXiv:2012.05503 (2020) <https://hal.inria.fr/hal-03047049v2>
- Zafeiropoulos, Haris, et al. "PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes." *GigaScience* 9.3 (2020): g1aa022.. DOI:10.1093/gigascience/g1aa022

## 3.3 Programming

- Author of the PEMA workflow
- Very experienced in programming with scripting programming languages (Python, BigDataScript, R).
- Basic knowledge of C++
- Very experienced in container-based technologies (Docker, Singularity)
- Experienced in web

## 4 Personal motivation

Bioinformatics play a great part in almost every aspect of modern biology. That is why after graduating Biology school (BSc), I focused on computer science and Bioinformatics (MSc). I am now quite experienced with a series of scripting programming languages ( Python, BigDataScript, R) and with container-based technologies (Docker, Singularity). PEMA [3], a pipeline for the analysis of metabarcoding data, was my first coding project; PEMA has now been published and selected from LifeWatch - ERIC a European Research Infrastructure, to support . PREGO and DARN are ongoing bioinformatics projects in the framework of my PhD.

My research interests focus on ecology and ecosystem functioning at the microbial dimension. As Systems Biology approaches can benefit the most this field, I have spent the last 2 years working on knowledge aggregation and data integration techniques as well as networks analysis are employed.

The last year, I have been working with the GeomScale group on the **volestipy** project and sampling on the flux space of metabolic networks [2]. This work was accepted in the proceedings of the 37th Symposium on Computational Geometry.

This GSoC project would allow me to continue working with in the framework of the GeomScale project whilst building the proposed application would benefit me the most, especially as it will allow me to extend the scopus of my PhD dissertation.

Metabolic interactions [1] is now the next big thing in systems biology.

In such a study, the changing phenotype of an ecosystem, as derived by the combination of phenotypes of all the ecosystem's individual entities, under different scenarios of constraint values could be predicted. The impact of such studies would be more than significant, especially on a time when nature suffers the most, mostly due to anthropogenic activities. Indicatively possible datasets such a study could be conducted on: "Impact of exogenous nitrogen on the cyanobacterial abundance and community in oil-contaminated sediment: A microcosm study" from Wang et al. [19] (peer-reviewed, no public data, email request) "Discovery of functional gene markers of bacteria for monitoring hydrocarbon pollution in the marine environment - a metatranscriptomics approach" from Knapik et al. [20] (not peer-reviewed, data not published yet) In their study [21], Gossart et al. describe the research framework of such studies we suggest to apply the aforementioned methods .

Coding has been a great part of my everyday routine through the recent years. The different needs of the different types of analysis in biology force you to deal with a great range of computing tasks, such as choosing the proper programming language for each issue, working with High Performance Computing environments, finding ways to make your code easy-to-use, easy-to-distribute and flexible at the same time. It would be both a profit and an enjoyment for me to be part of the Google Summer of Code and upgrade my so-far programming skills, particularly as volesti functions have been developed in C++, meaning that I will have to deal with new personal coding challenges.

## References

- [1] Jingyi Cai, Tianwei Tan, and SH Joshua Chan. "Predicting Nash equilibria for microbial metabolic interactions". In: *Bioinformatics* (2020).
- [2] Apostolos Chalkis et al. "Geometric algorithms for sampling the flux space of metabolic networks". In: *arXiv preprint arXiv:2012.05503* (2020).
- [3] Haris Zafeiropoulos et al. "PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes". In: *GigaScience* 9.3 (2020), giaa022.