



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

on the road to a PhD
Haris Zafeiropoulos

1. Bioinformatics methods for microbial diversity assessment
 - 1.1 pema: a metabarcoding pipeline
 - 1.2 darn: known unknowns in COI amplicon data
2. PREG0: a knowledge-base for organisms - environments - processes associations
3. dingo: a Python library for metabolic flux sampling
 - 3.1 Flux sampling
4. Tristomo swamp: a hybrid amplicon & shotgun metagenomics analysis
5. Publications

eDNA metabarcoding

Marker genes

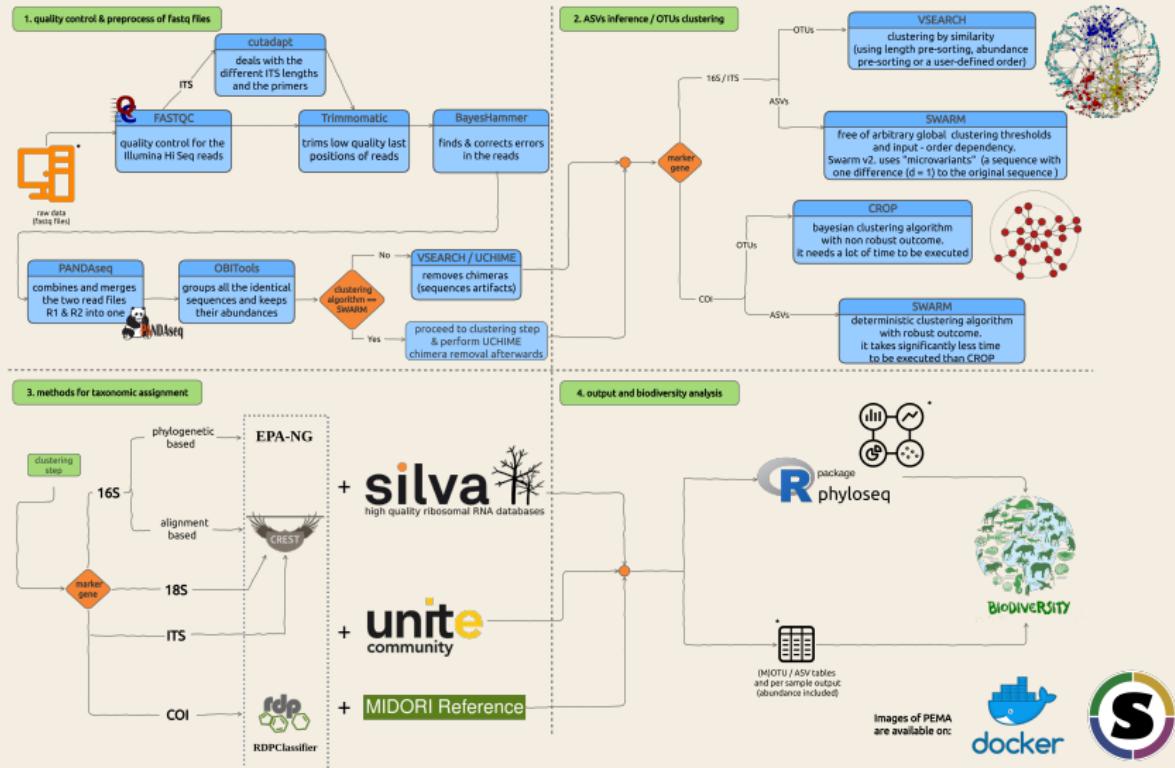
1. **16S rRNA:** Bacteria, Archaea
2. **12S rRNA:** Vertebrates
3. **18S rRNA:** Small eukaryotes, Metazoa
4. **ITS:** Fungi
5. **COI:** Eukaryotes
6. **rbcl:** Plants
7. **dsrb:** Bacteria, Archaea

Bioinformatics analysis steps

1. Sequence pre-processing
2. OTUs clustering / ASVs inference
3. Taxonomic assignment
4. Biodiversity analysis

PEMA architecture

PEMA in a nutshell



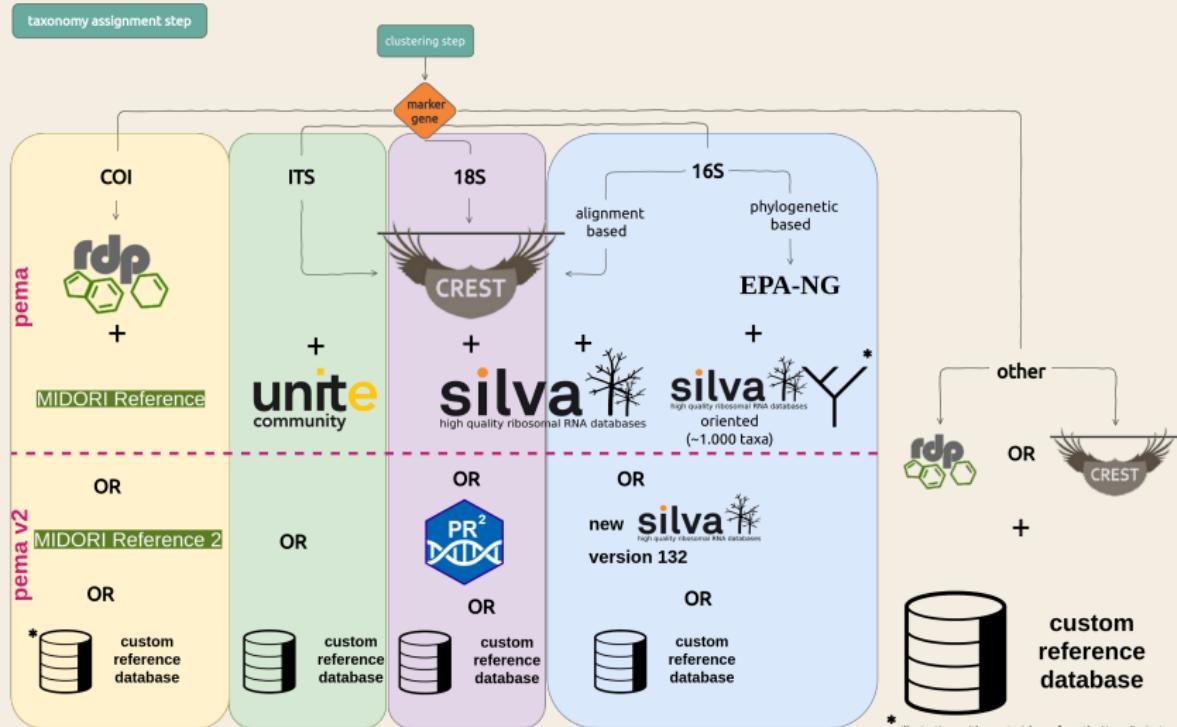
Being a geek just for a bit !

```
for(int i : range(1,  
    in := "in_$i.tx  
    sys date > $in  
  
    out := "out_$i.t  
    task( out <- in  
        sys echo Tas  
    }  
}
```

Containerization

BigDataScript
programming language

PEMA v.2



DARN: investigating known unknown in COI amplicon data

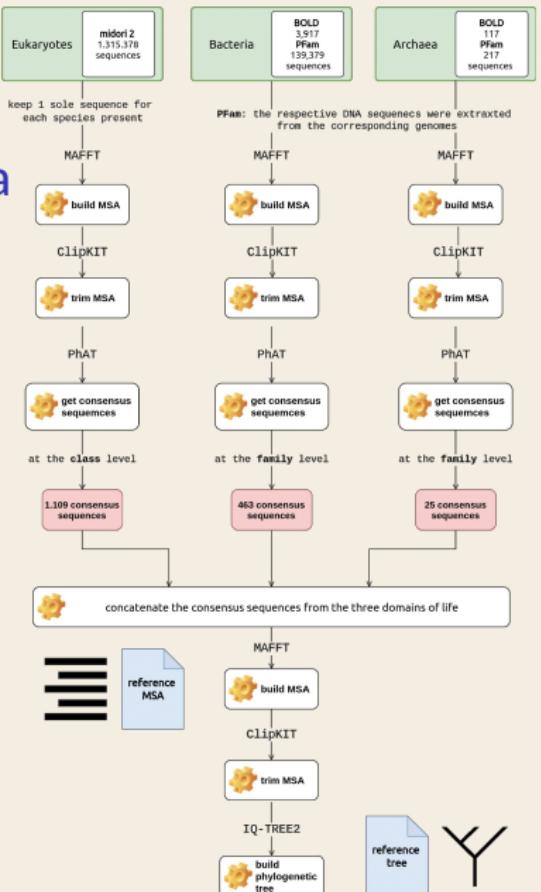


Figure from: Heirendt, Laurent, et al. Nature protocols 14.3 (2019): 639-702.

1. Bioinformatics methods for microbial diversity assessment
 - 1.1 pema: a metabarcoding pipeline
 - 1.2 darn: known unknowns in COI amplicon data
2. PREG0: a knowledge-base for organisms - environments - processes associations
3. dingo: a Python library for metabolic flux sampling
 - 3.1 Flux sampling
4. Tristomo swamp: a hybrid amplicon & shotgun metagenomics analysis
5. Publications

PREGO methodology

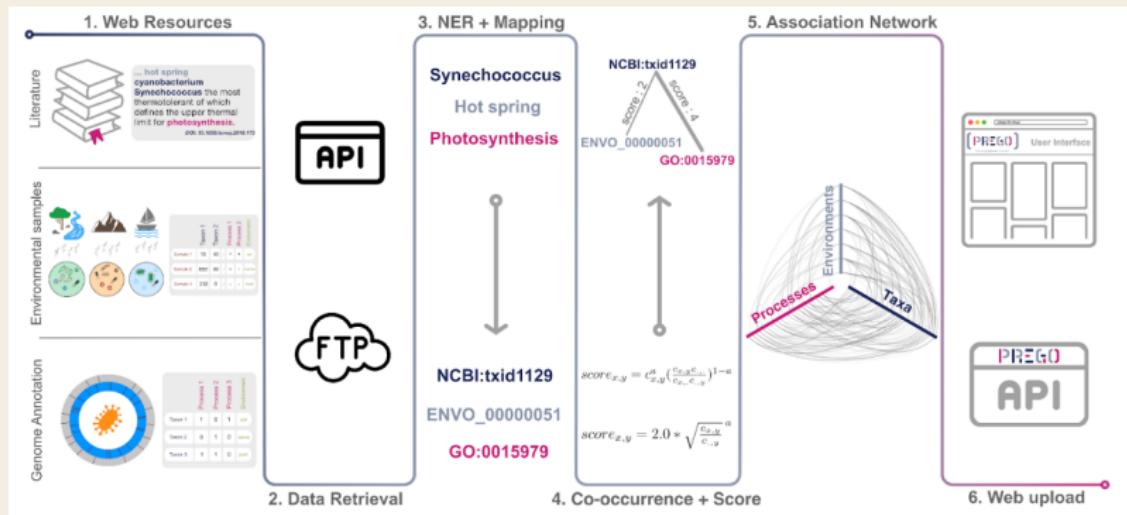


Figure from: Heirendt, Laurent, et al. Nature protocols 14.3 (2019): 639-702.

1. Bioinformatics methods for microbial diversity assessment
 - 1.1 pema: a metabarcoding pipeline
 - 1.2 darn: known unknowns in COI amplicon data
2. PREG0: a knowledge-base for organisms - environments - processes associations
3. dingo: a Python library for metabolic flux sampling
 - 3.1 Flux sampling
4. Tristomo swamp: a hybrid amplicon & shotgun metagenomics analysis
5. Publications

Genome-scale metabolic reconstruction

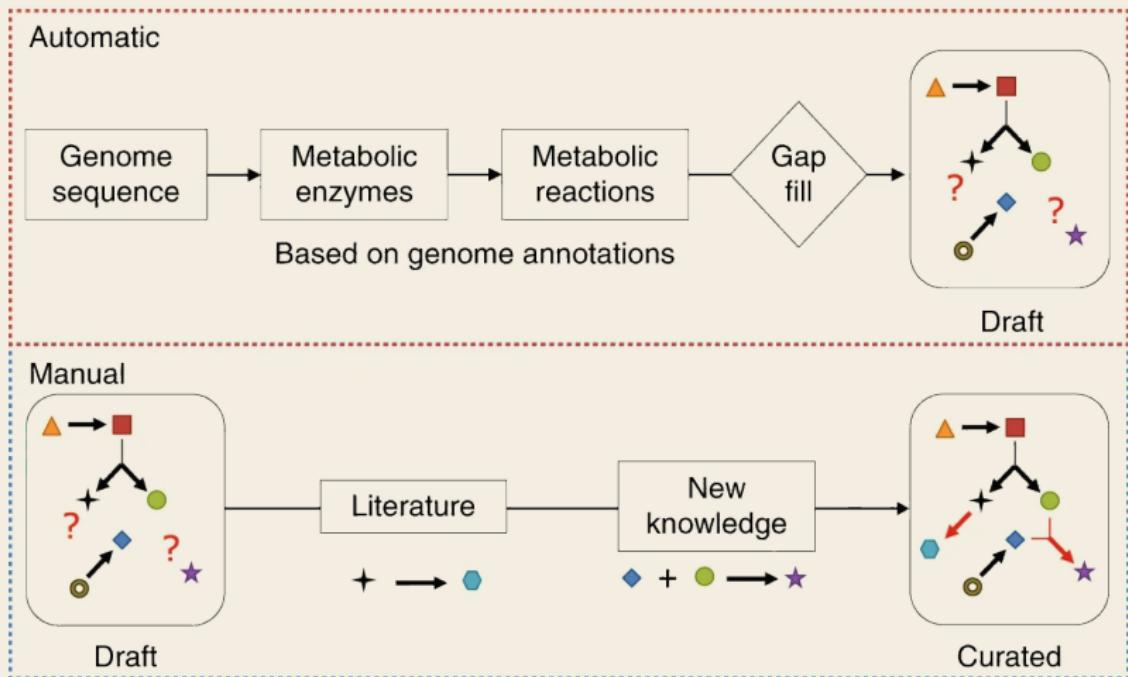


Figure from: Heirendt, Laurent, et al. Nature protocols 14.3 (2019): 639-702.

From a stoichiometric matrix to a constraint-based model

$$\begin{array}{c}
 \text{Metabolites} \\
 \begin{array}{c}
 \textcolor{orange}{\Delta} \\
 \textcolor{brown}{\blacksquare} \\
 \star \\
 \textcolor{olive}{\bullet} \\
 \textcolor{teal}{\lozenge} \\
 \textcolor{brown}{\circlearrowleft} \\
 \textcolor{blue}{\blacklozenge} \\
 \textcolor{purple}{\star}
 \end{array}
 \end{array}
 \begin{array}{c}
 \text{Reactions} \\
 \begin{array}{c}
 R_1 \quad R_2 \quad R_3 \quad R_4 \quad R_5
 \end{array}
 \end{array}
 \begin{array}{c}
 \text{S-matrix} \\
 \left(\begin{array}{ccccc}
 -1 & 0 & 0 & 0 & 0 \\
 1 & -1 & 0 & 0 & 0 \\
 0 & 1 & -1 & 0 & 0 \\
 0 & 1 & 0 & 0 & -1 \\
 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & -1 & 0 \\
 0 & 0 & 0 & 1 & -1 \\
 0 & 0 & 0 & 0 & 1
 \end{array} \right)
 \end{array}
 \times
 \begin{array}{c}
 \text{Flux vector} \\
 \left(\begin{array}{c}
 v_1 \\
 v_2 \\
 v_3 \\
 v_4 \\
 v_5
 \end{array} \right) = \left(\begin{array}{c}
 0 \\
 0 \\
 0 \\
 0 \\
 0
 \end{array} \right)
 \end{array}$$

Flux Balance Analysis

Maximize minimize an
objective function:

$\psi = c_1v_1 + c_2v_2 + \dots + c_5v_5$
such that:

$$S * v = 0$$

and for each reaction i :

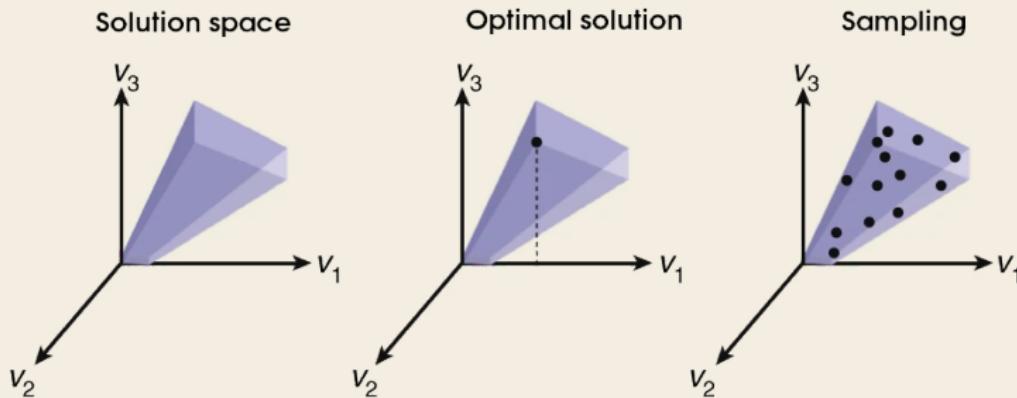
$$lb_j \leq v_j \leq ub_j$$

where lb : lower bound,
 ub : upper bound and

S: the stoichiometric matrix

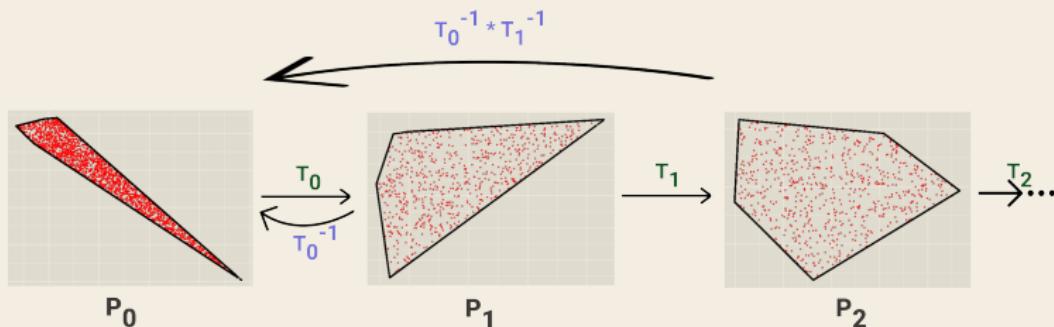
Flux sampling

an alternative approach



- enables the analysis of GEMs without the need of an objective function
- determines the feasible solution spaces for fluxes in a network based on a set of conditions as well as the probability of obtaining a solution

Our Markov Chain Monte Carlo (MCMC) algorithm for flux sampling



Steps of an MMCS phase

- **sampling step:** using a variant of the **Billiard walk**
- **rounding step:** calculate a linear transformation T_i that puts the sample into isotropic position and then apply it on P_i to obtain the polytope of the next phase
- check several statistic tests

Find possible targets against SARS-CoV-2

a flux sampling application

Bioinformatics, 36(26), 2020, i813–i821

doi: 10.1093/bioinformatics/btaa813

ECCB2020

OXFORD

Systems

FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2

Alina Renz^{1,2,*}, Lina Widerspick¹ and Andreas Dräger^{1,2,3,*}

¹Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI) and ²Department of Computer Science, University of Tübingen, Tübingen 72076, Germany and ³German Center for Infection Research (DZIF), partner site Tübingen, Germany

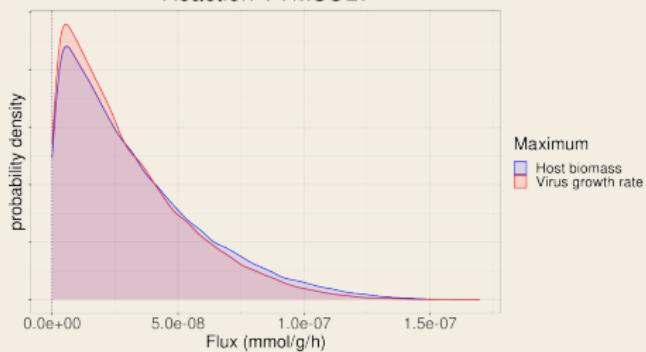
- Renz et al. '20 built the biomass function of Sars-Cov-2 to build a host - virus network
- Using FBA they computed an optimal steady state using
 - (i) human biomass maintenance,
 - (ii) virus growth rate
- They found reaction GK1 as a possible anti-viral target.

Find possible targets against SARS-CoV-2

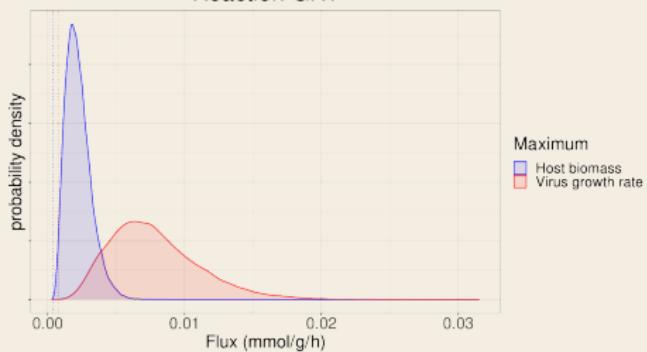
a flux sampling application



Reaction TYMSULT



Reaction GK1

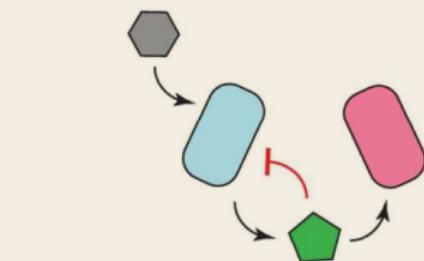


- Check if the flux distribution of a reaction changes.
- Find possible anti-viral targets and study further.

Further applications of metabolic flux sampling



Scott, William T., et al. "Metabolic flux sampling predicts strain-dependent differences related to aroma production among commercial wine yeasts." Microbial cell factories 20.1 (2021): 1-15.



What about microbial interactions ?

dingo: a Python library for flux sampling



<https://github.com/GeomScale/dingo>

how to GColab notebook



1. Bioinformatics methods for microbial diversity assessment
 - 1.1 pema: a metabarcoding pipeline
 - 1.2 darn: known unknowns in COI amplicon data
2. PREG0: a knowledge-base for organisms - environments - processes associations
3. dingo: a Python library for metabolic flux sampling
 - 3.1 Flux sampling
4. Tristomo swamp: a hybrid amplicon & shotgun metagenomics analysis
5. Publications

1. Bioinformatics methods for microbial diversity assessment
 - 1.1 pema: a metabarcoding pipeline
 - 1.2 darn: known unknowns in COI amplicon data
2. PREG0: a knowledge-base for organisms - environments - processes associations
3. dingo: a Python library for metabolic flux sampling
 - 3.1 Flux sampling
4. Tristomo swamp: a hybrid amplicon & shotgun metagenomics analysis
5. Publications

Publications

TeX, L^AT_EX, and Beamer

- [1] Zafeiropoulos, H., Paragkamian, S., Ninidakis, S., Pavlopoulos, G.A., Jensen, L.J. & Pafilis, E. PREGO: a literature- and data-mining resource to associate microorganisms, biological processes, and environment types. - *under submission*
- [2] Zafeiropoulos, H., Gargan, L., Hintikka, S., Pavloudi, C., & Carlsson, J. (2021). The Dark mAtteR iNvestigator (DARN) tool: getting to know the known unknowns in COI amplicon data. *Metabarcoding and Metagenomics*, 5, e69657.
- [3] Chalkis, A., Fisikopoulos, V., Tsigaridas, E., & Zafeiropoulos, H. (2021). Geometric algorithms for sampling the flux space of metabolic networks, *37th International Symposium on Computational Geometry (SoCG 2021)*.
- [4] Zafeiropoulos, H., Gioti, A., Ninidakis, S., Potirakis, A., Paragkamian, S., ... & Pafilis, E. (2021). Os and 1s in marine molecular research: a regional HPC perspective. *GigaScience*, 10(8), giab053.
- [5] Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., ... & 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.

Thank you for your attention
and your patience ;)

GitHub : <https://github.com/hariszaf>

email : haris-zaf@hcmr.gr

Twitter : haris_zaf

web-site : <https://hariszaf.github.io/>



GEOMSCALE