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

on the road to a PhD

Haris Zafeiropoulos

eDNA metabarcoding

Marker genes

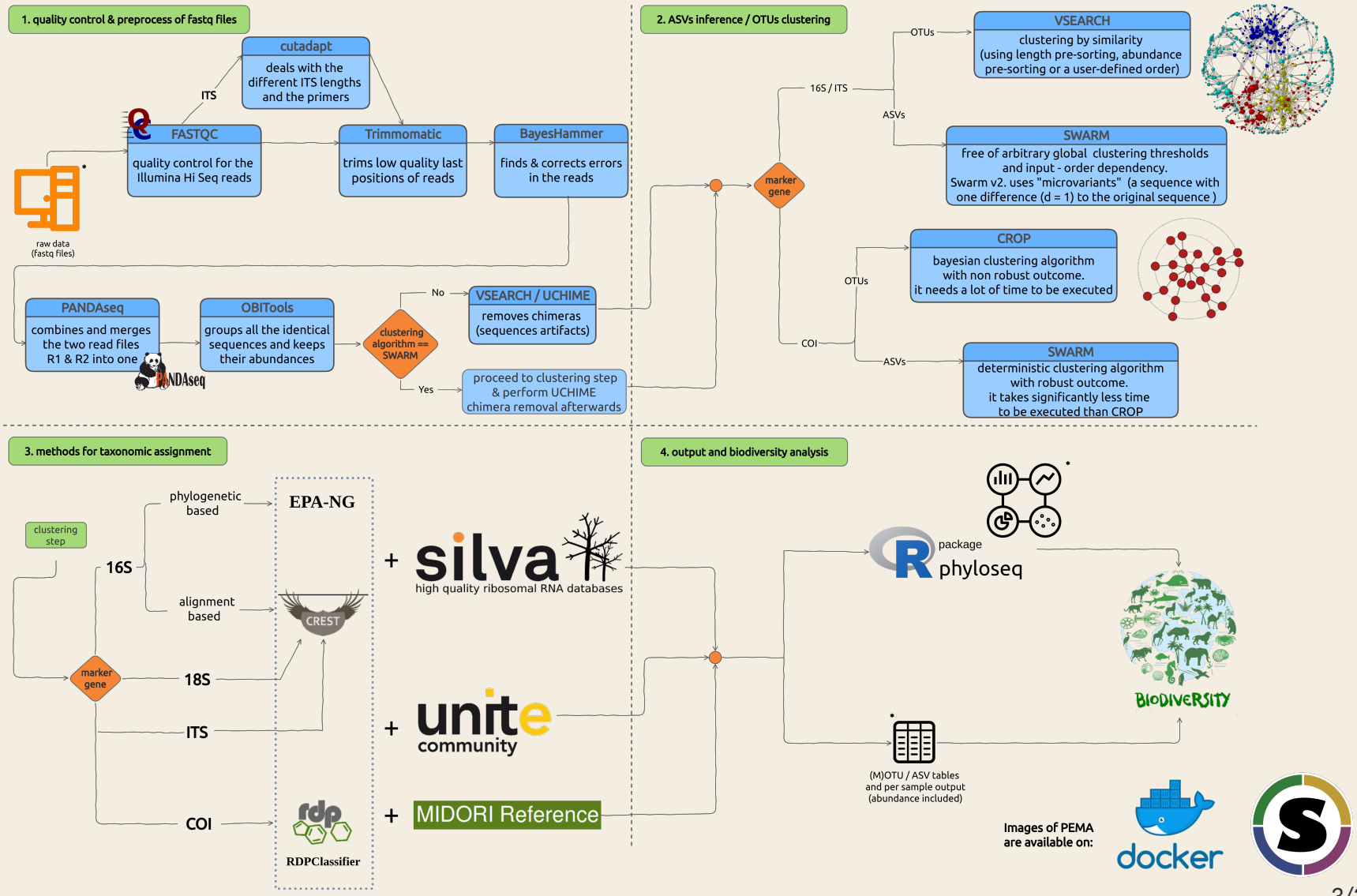
1. **16S rRNA:** Bacteria, Archaea
2. **12S rRNA:** Vertebrates
3. **18S rRNA:** Small eukaryotes, Metazoa
4. **ITS:** Fungi
5. **COI:** Eukaryotes
6. ***rbcl*:** Plants
7. ***dsrb*:** Bacteria, Archaea
8. ...

Bioinformatics analysis steps

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

PEMA architecture

PEMA in a nutshell



* Illustrations with asterisk are from the Noun Project

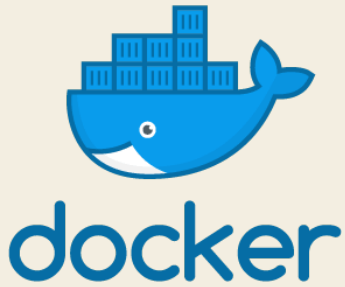
Being a geek just for a bit !

```
for(int i : range(1,
    in  := "in_$i.tx
    sys date > $in

    out := "out_$i.t
    task( out <- in
        sys echo Tas
    }
}
```

`BigDataScript`

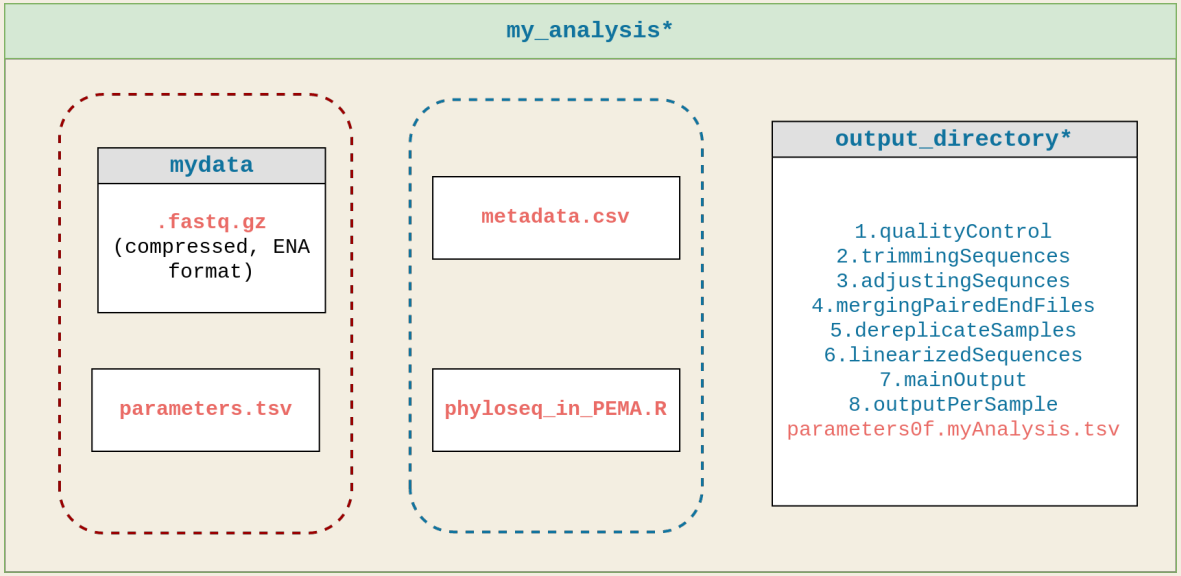
programming language



Containerization

Mount your I/O

you give something - you take something

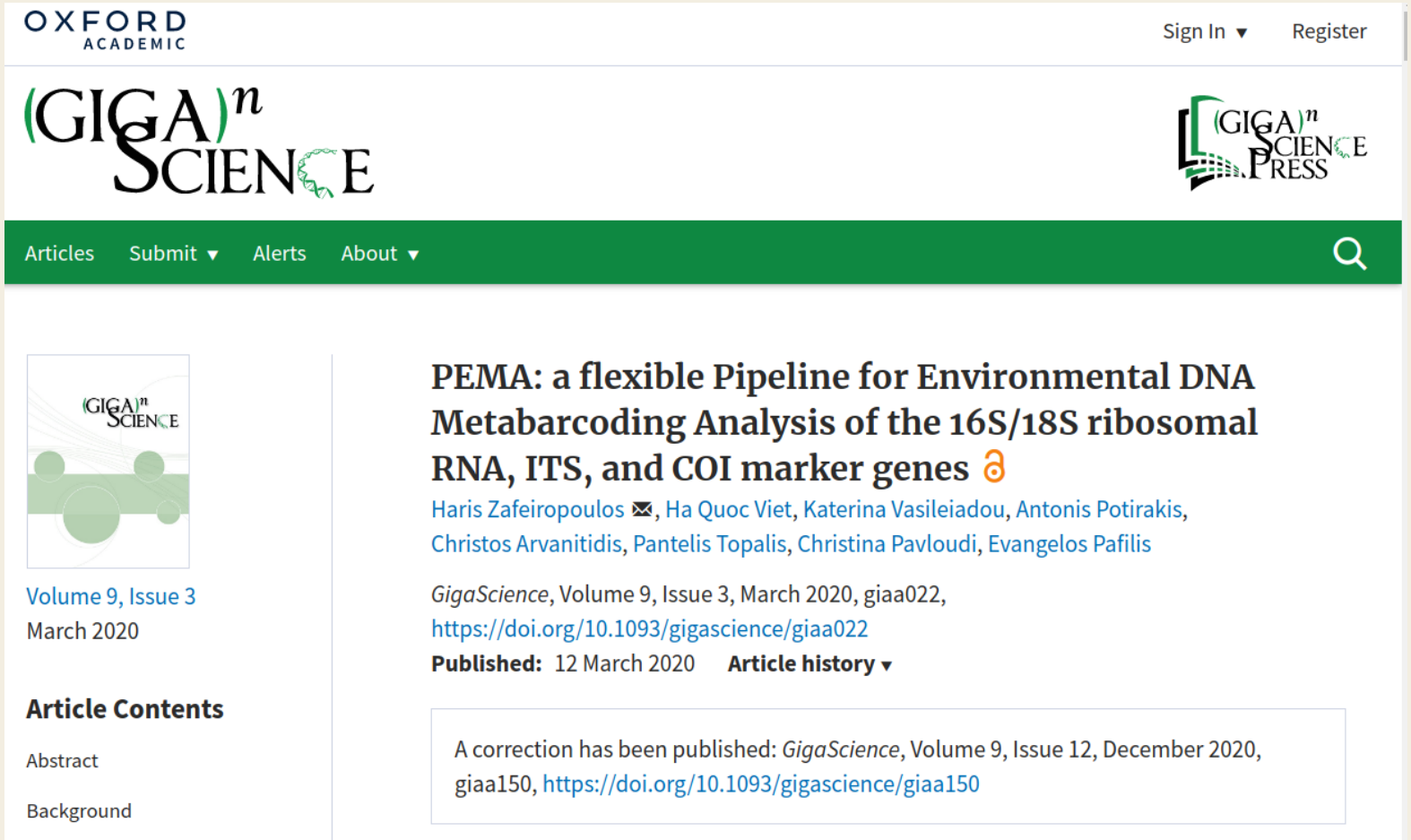


	Sample 1	Sample 2	Sample 3	Sample 4
Taxon 1	1	0	1	2
Taxon 2	0	1	0	2
Taxon 3	1	1	0	4

text file
directory
***user can edit the name**
(the rest **need** or will have the exact names shown)

- - - - - mandatory input files
- - - - - optional input files

1st version published in 2020



OXFORD ACADEMIC

Sign In ▾ Register

(GIGA)n SCIENCE

(GIGA)n SCIENCE PRESS

Articles Submit ▾ Alerts About ▾

Volume 9, Issue 3
March 2020

Article Contents

Abstract

Background

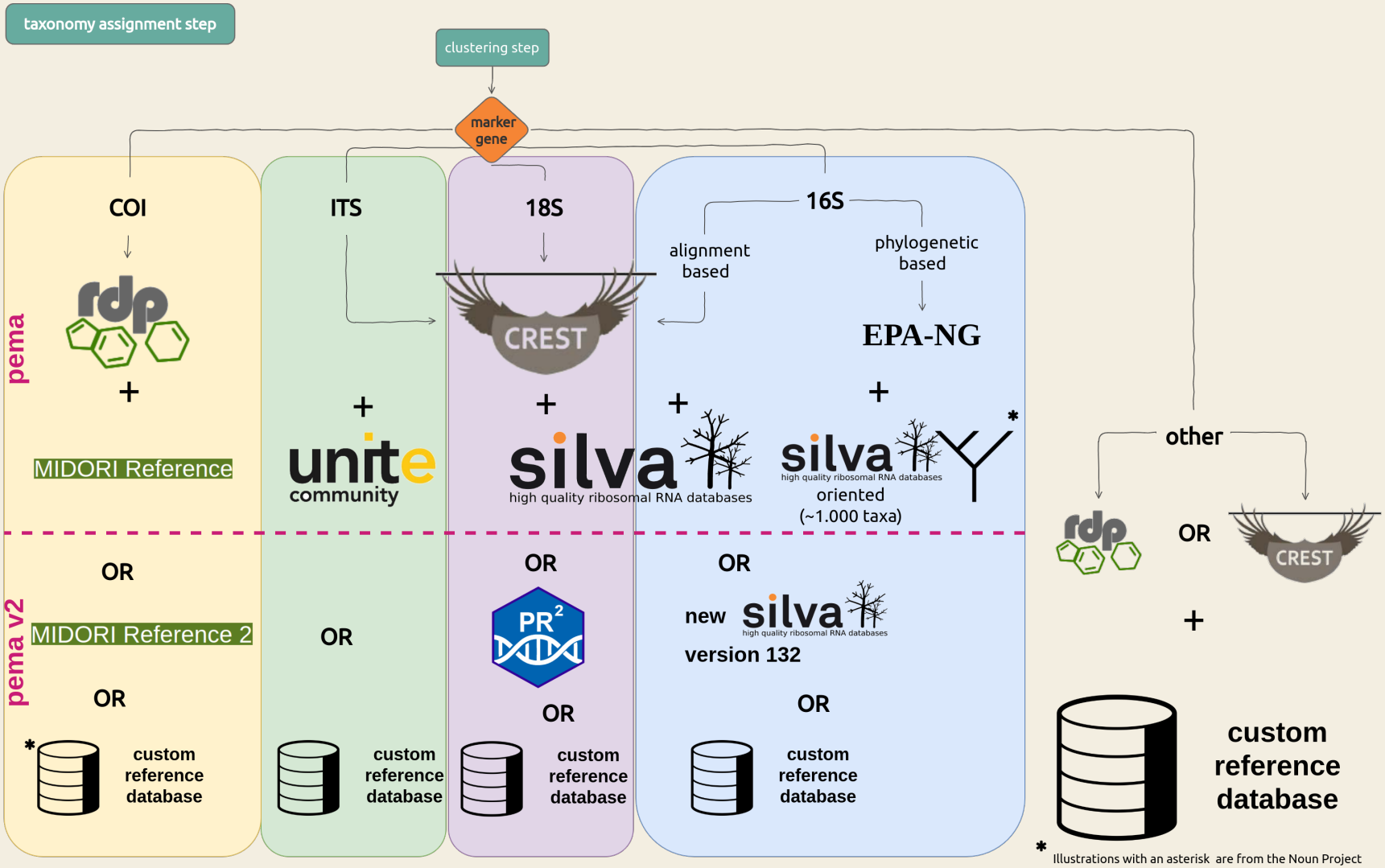
PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes

Haris Zafeiropoulos, Ha Quoc Viet, Katerina Vasileiadou, Antonis Potirakis, Christos Arvanitidis, Pantelis Topalis, Christina Pavloudi, Evangelos Pafilis

GigaScience, Volume 9, Issue 3, March 2020, giaa022,
https://doi.org/10.1093/gigascience/giaa022
Published: 12 March 2020 **Article history** ▾

A correction has been published: *GigaScience*, Volume 9, Issue 12, December 2020, giaa150, https://doi.org/10.1093/gigascience/giaa150

PEMA v.2


taxonomy assignment step
clustering step
marker gene
COI
ITS
18S
16S
alignment based
phylogenetic based
other
pema
rdp
+
MIDORI Reference
unite community
CREST
silva
high quality ribosomal RNA databases
EPA-NG
silva
high quality ribosomal RNA databases
oriented
(~1.000 taxa)
OR
pema v2
MIDORI Reference 2
PR2
new silva
high quality ribosomal RNA databases
version 132
custom reference database
custom reference database
custom reference database
custom reference database
custom reference database
Illustrations with an asterisk are from the Noun Project

Latest PEMA version

addressing the challenges of the community

ASSEMBLE
ASSOCIATION OF EUROPEAN MARINE BIOLOGICAL LABORATORIES EXPANDED

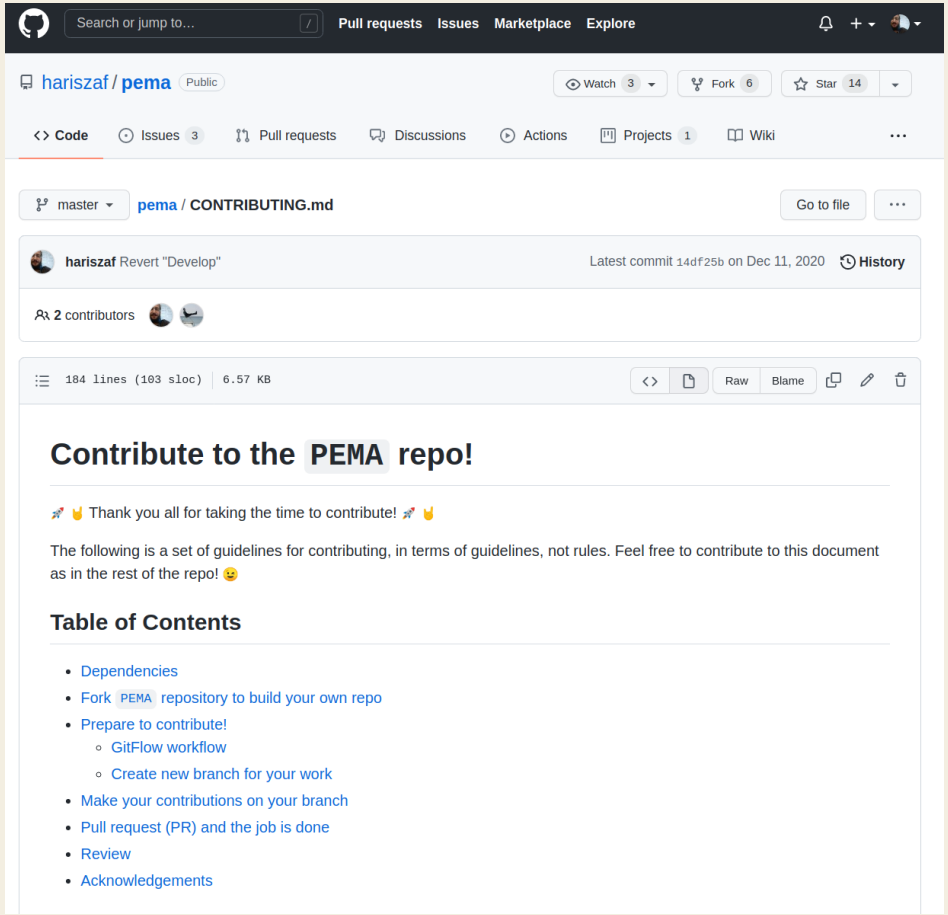
MBON
Marine Biodiversity Observation Network

`pema:v.2.1.4` includes:

1. analysis of 12S rRNA data now supported (12S Vertebrate Classifier v2.0.0-ref database)
2. PR2 as an alternative reference database for the case of 18S rRNA
3. the `ncbi-taxonomist` tool was added to return the NCBI Taxonomy Id of the taxonomies found

open source is cool

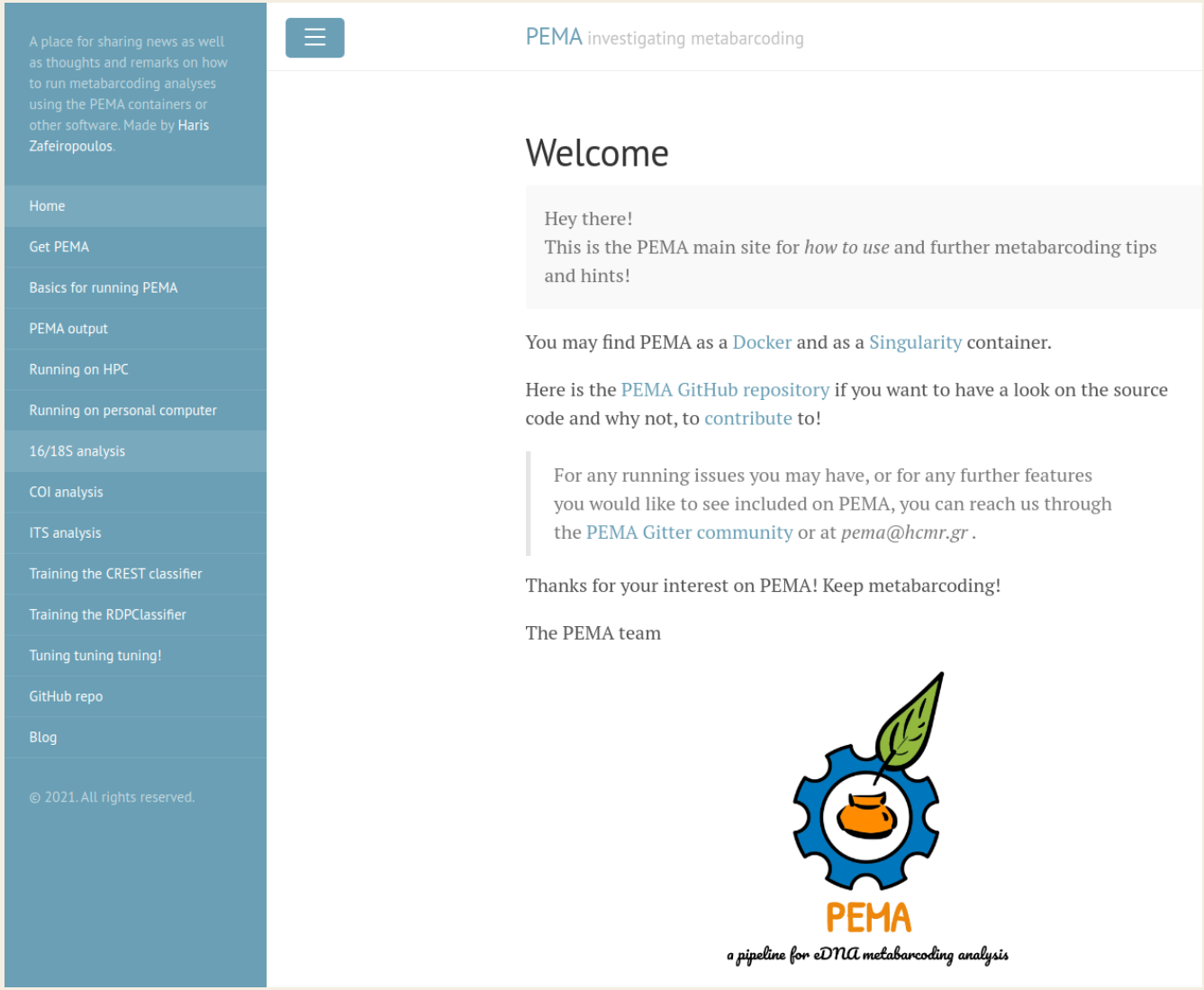
interacting with the community



PEMA aims at building a community to discuss challenges on metabarcoding come up with solutions and why not develop some of them!

How to and further documentation

at pema.hcmr.gr



DARN: investigating known unknown in COI amplicon data

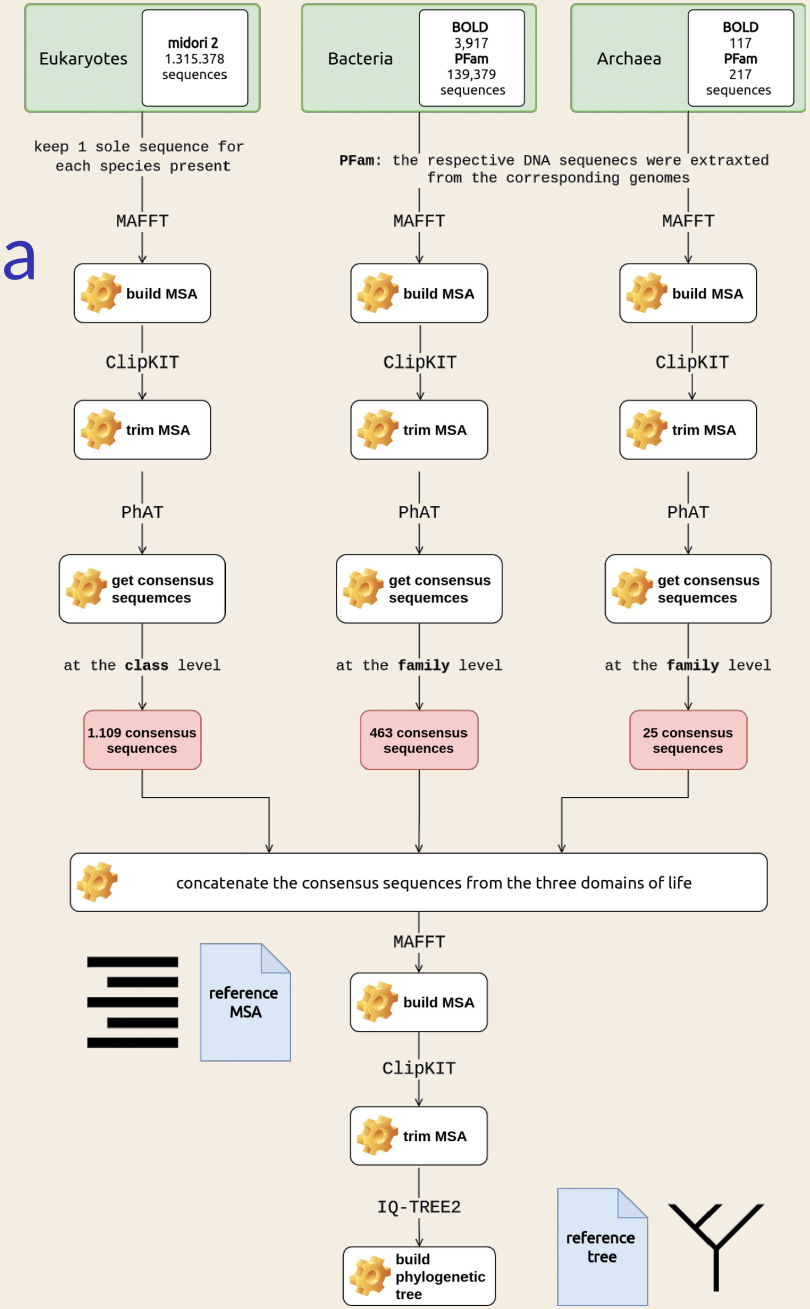


Figure from: Heirendt, Laurent, et al. Nature protocols 14.3 (2019): 639-702.

PREGO methodology

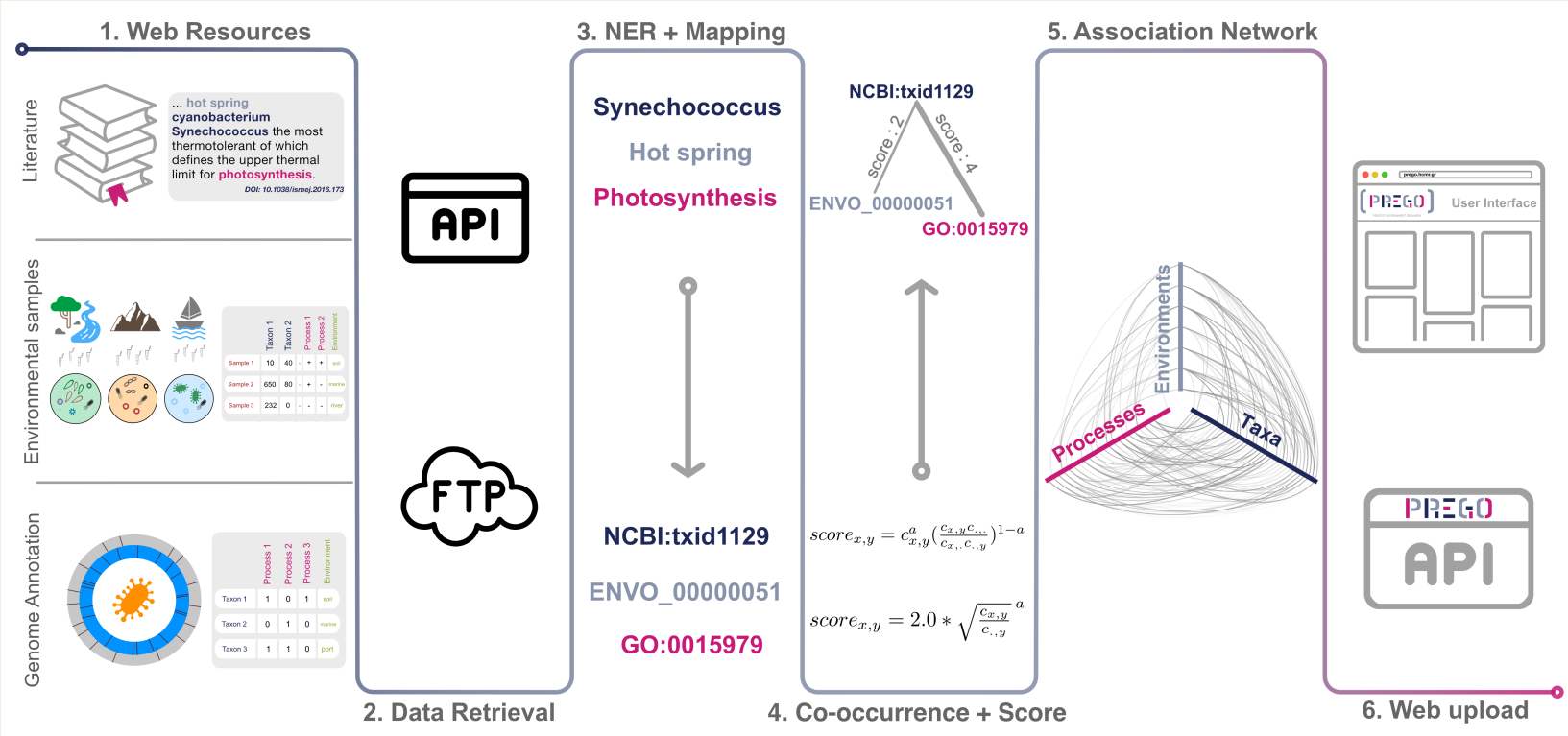


Figure from: Heirendt, Laurent, et al. Nature protocols 14.3 (2019): 639-702.

Genome-scale metabolic reconstruction

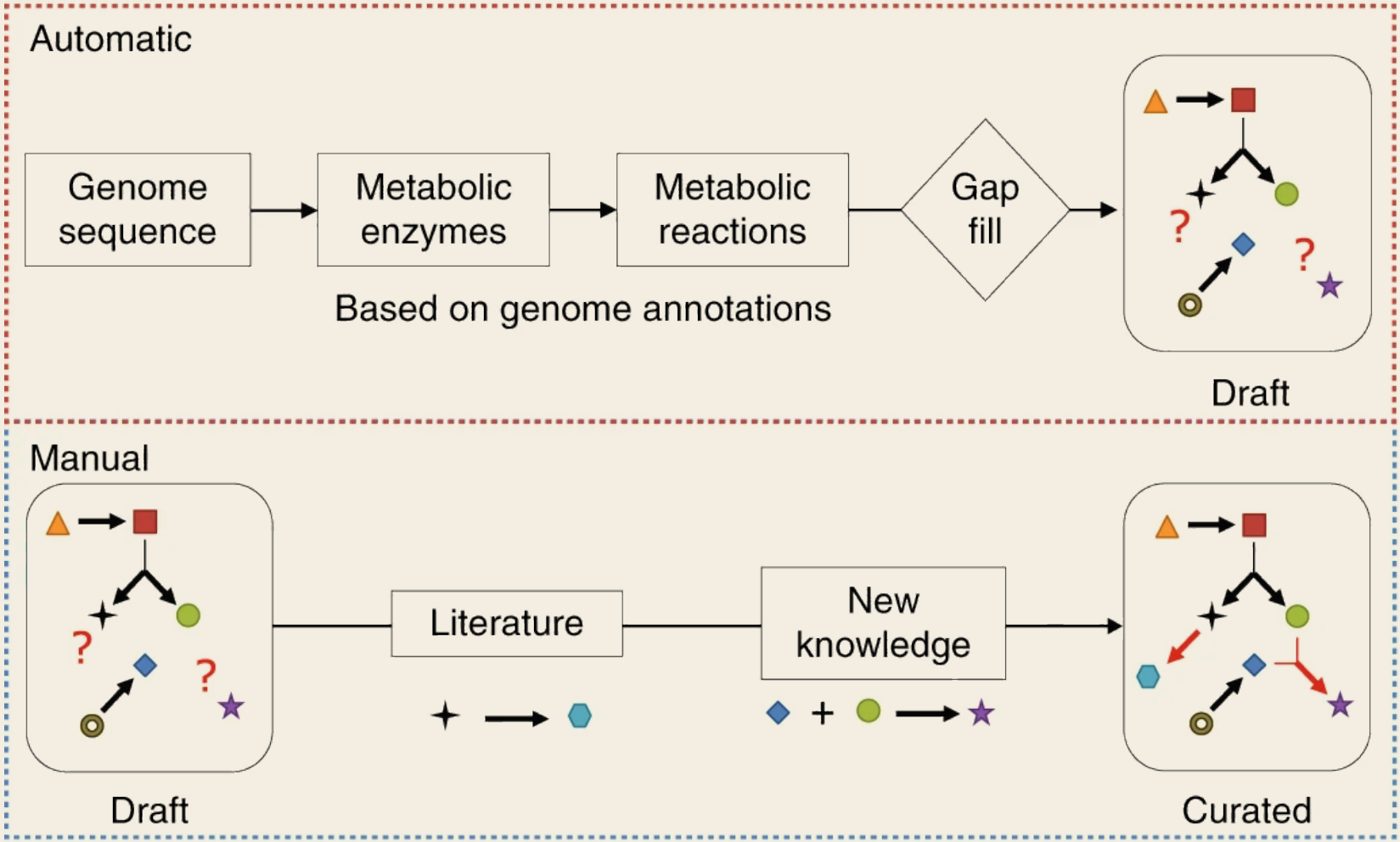
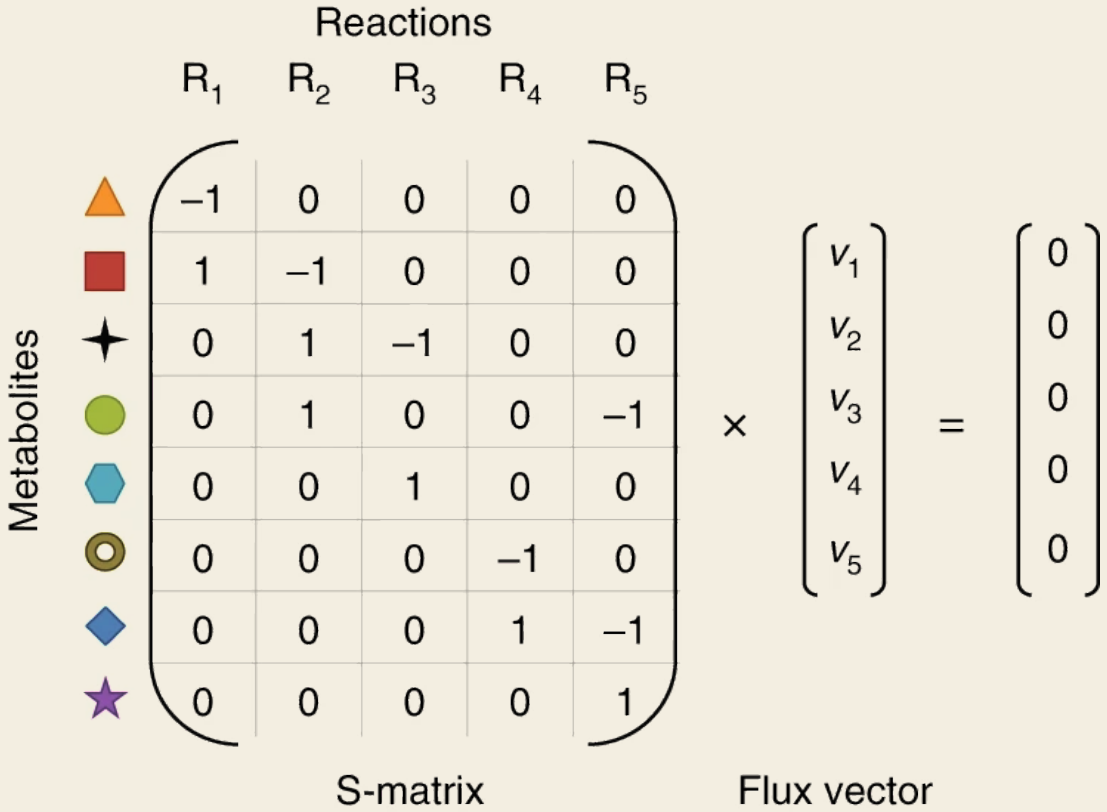


Figure from: Heirendt, Laurent, et al. Nature protocols 14.3 (2019): 639-702.

From a stoichiometric matrix

to a constraint-based model



Flux Balance Analysis

Maximize minimize an objective function:

$$\psi = c_1 v_1 + c_2 v_2 + .. + c_5 v_5$$

such that:

$$S * v = O$$

and for each reaction i:

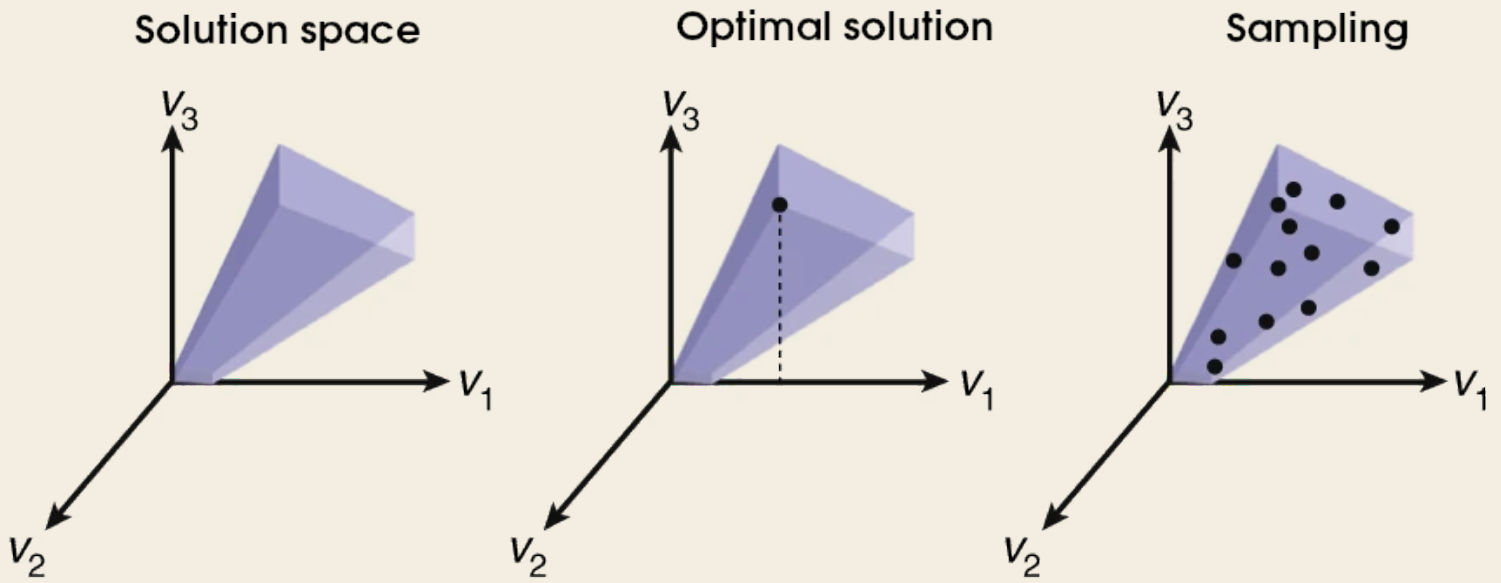
$$lb_i <= v_i <= ub_i$$

where lb: lower bound, ub: upper bound and S: the stoichiometric matrix

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

Flux sampling

an alternative approach

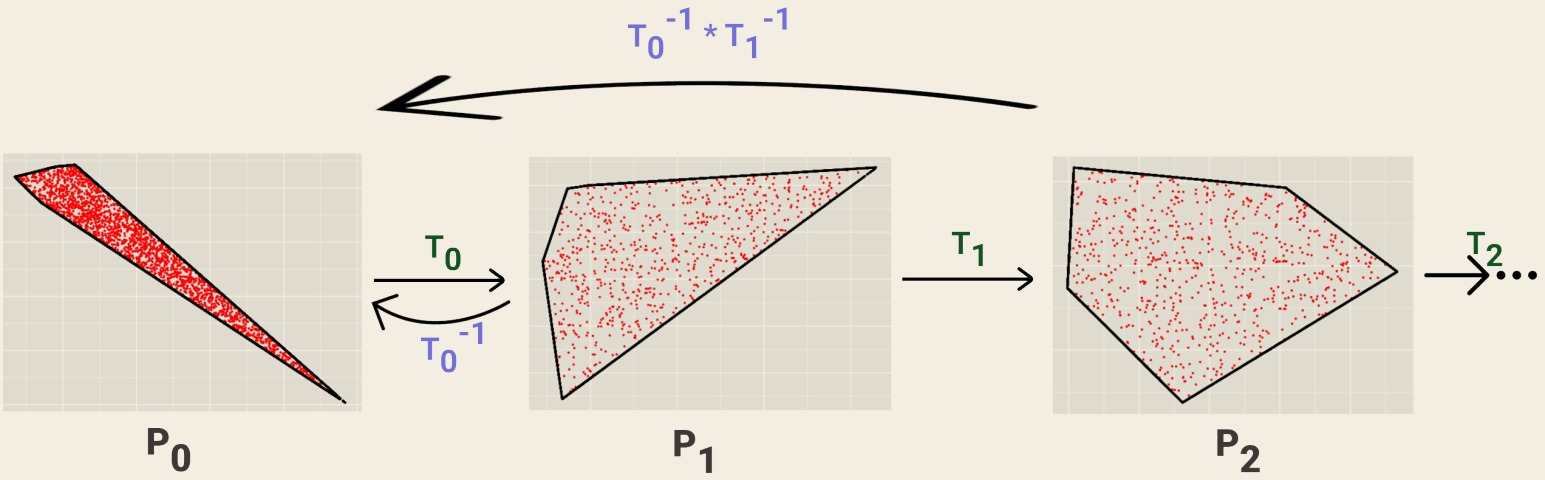


- enables the analysis of GEMs without the need of an objective function
- determines the feasible solution spaces for fluxes in a network based on a set of conditions as well as the probability of obtaining a solution

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

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

Chalkis, Fisikopoulos, Tsigaridas and Zafeiropoulos "Geometric Algorithms for Sampling the Flux Space of Metabolic Networks", SoCG 2021, DOI: 10.4230/LIPIcs.SoCG.2021.21

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

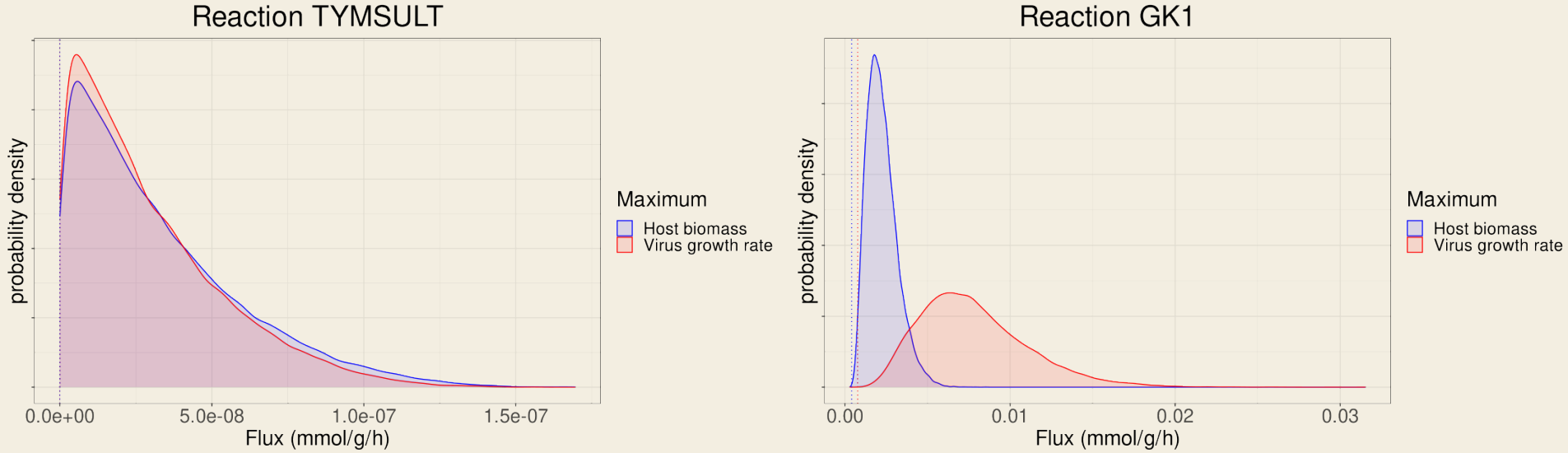
Alina Renz[1,2,*], Lina Widerspick[1] and Andreas Dräger[1,2,3,*]

[1]Computational Systems Biology of Infections and Antimicrobial-Resistant Pathogens, Institute for Bioinformatics and Medical Informatics (IBMI) and [2]Department of Computer Science, University of Tübingen, Tübingen 72076, Germany and [3]German 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

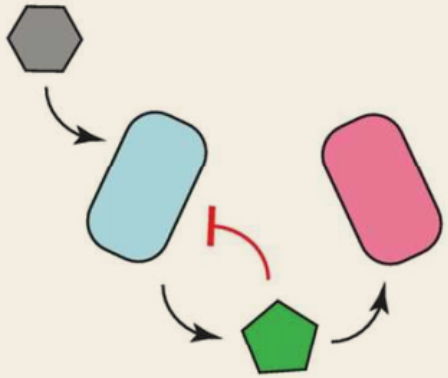


- Check if the flux distribution of a reaction changes.
- Find possible anti-viral targets and study further.

Further applications

of metabolic flux sampling

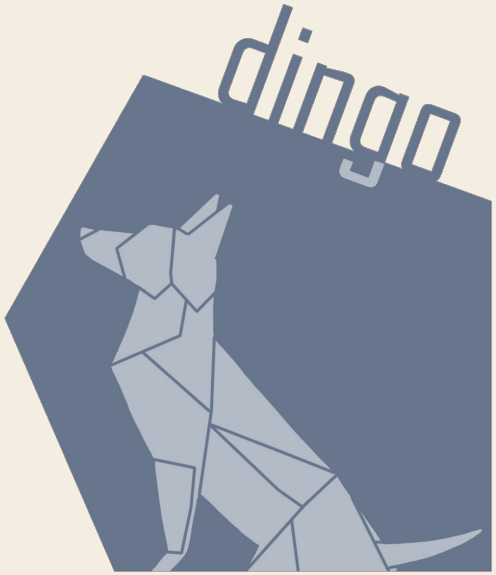
Scott, William T., et al. "Metabolic flux sampling predicts strain-dependent differences related to aroma production among commercial wine yeasts." Microbial cell factories 20.1 (2021): 1-15.



What about microbial interactions ?

`dingo`: a Python library

for flux sampling



`how to` GCollab notebook

https://github.com/GeomScale/dingo

Publications

TEX, LATEX, 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). 0s 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