

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

developing computational approaches to better understand microbial assemblages

Haris Zafeiropoulos
post-doctoral researcher

KU LEUVEN

My story starts at a small but quite a nice place

Ancient Olympia



Great times with a lot of basketball and even more *hows* and *whys*

.. moves to Athens

Biology Department, National & Kapodestrian University



Thesis on the morphology,
morphometry and anatomy of
species of the genus
Pseudamnicola in Greece

.. and then to a "horrible" place called Crete

Medical School (MSc), Biology department (PhD) of UoC,
Hellenic Centre for Marine Research



MSc thesis on eDNA metabarcoding for biodiversity assessment:
Algorithm design and bioinformatics analysis pipeline implementation

to find myself in Leuven
for my first post-doc @ Rega Institute



Post-doc in network and metabolic modelling in the framework of the
3D'omics project.

Main questions regarding a microbial community for a deeper understanding of such assemblages



community
structure
who

taxa, abundance



ecosystem
type
where

habitats



functional
potential
what

processes

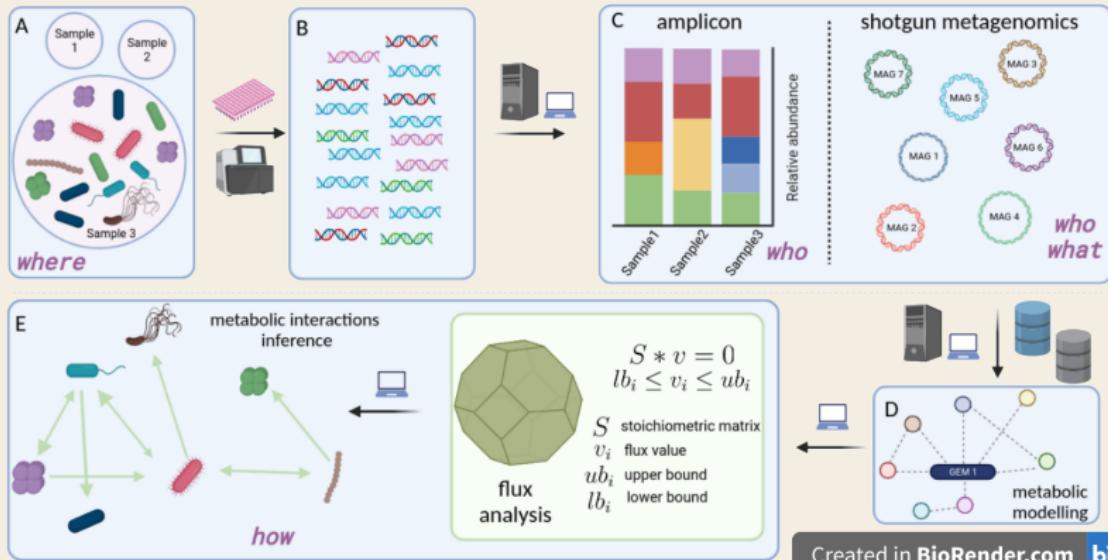


underlying
mechanisms
how

interactions, fluxes

Reverse ecology

transforming ecology into a high-throughput field



Created in BioRender.com

Software development to support the analysis of amplicon and shotgun metagenomics data

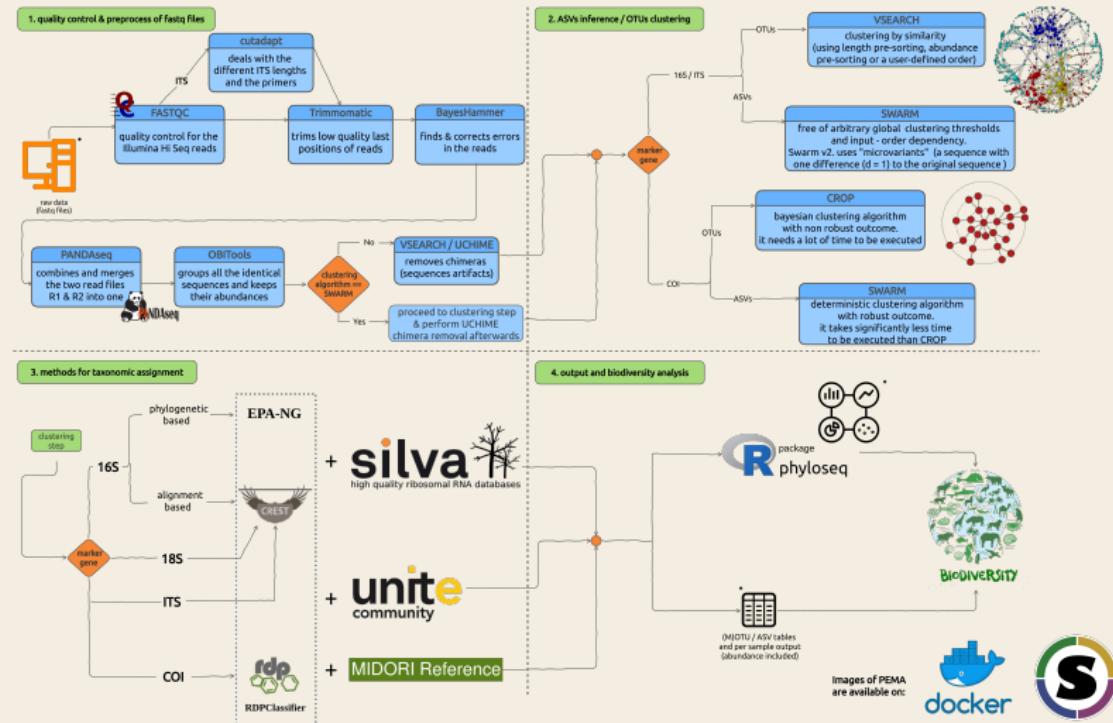
- PEMA
- DARN
- metaGOflow



PEMA features

an overview

PEMA in a nutshell





Moving at the large scale

PEMA @ infrastructures

The screenshot shows the Tesseract VRE Development Portal interface. On the left is a sidebar with navigation links: Personal space, Alaudus Altissima mapping, ARMS, Biotope, Cetacean functional trophodynamics, Metagenomics, and Tools. The main area is titled "Run a ARMS workflow" and contains a "Workflow overview" section with a diagram of the workflow steps: Import TSV (Param file, Import file), PEMA (csv / tsv, Ondefine), WalMS (Ondefine), and WRMS (Tab dataset). Below this are sections for "Workflow description", "ARMS parameters", "Create workflow", and "Workflow created". At the bottom left is a "Save theme" button.

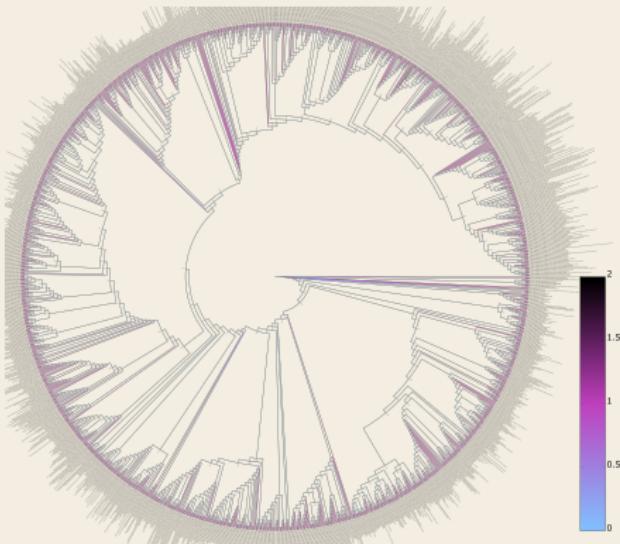
LifeWatch ERIC:
www.lifewatch.eu/
Tesseract VRE Development Portal:
www.lifewatch.dev/dashboard

1. web - interface make analysis even easier
2. researchers without access to HPC/clouds are now able to run big scale analyses
3. combine with other tools

Elixir Greece:
<https://elixir-greece.org/>
Hypatia cloud infrastructure:
<https://hypatia.athenarc.gr/>



Reference phylogenetic tree of the COI consensus sequences retrieved



The consensus sequences have been placed in their corresponding taxonomic branches, proving the tree valid.

EMO BON

a network of genomics observatories sampling periodically



Based on **MGNify**
along with the EMBL-EBI
group, we developed
a workflow to address
the needs of the
EMO-BON community.

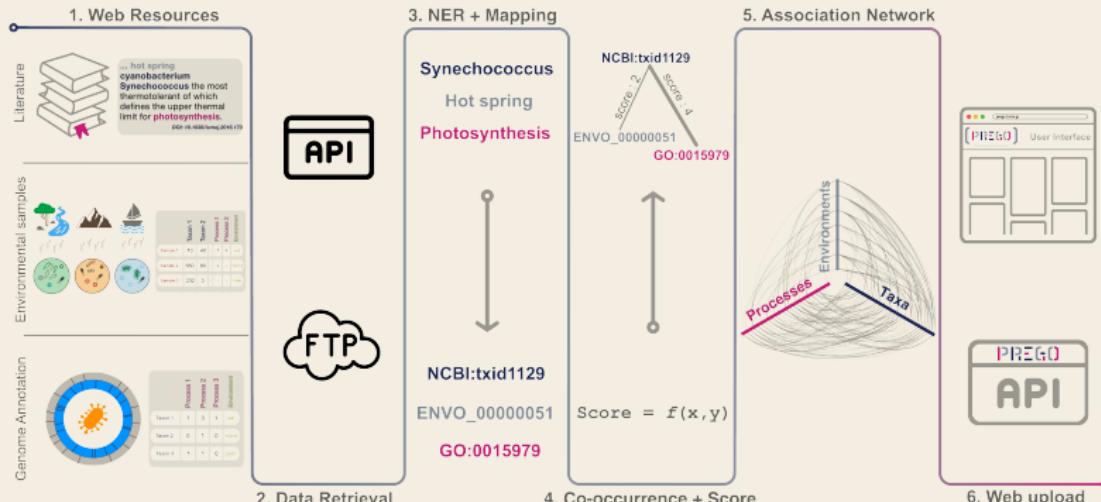
GitHub repository
ReadTheDocs

Data integration to enhance hypothesis generation approaches

- PREGO
- microbetag

PREGO methodology

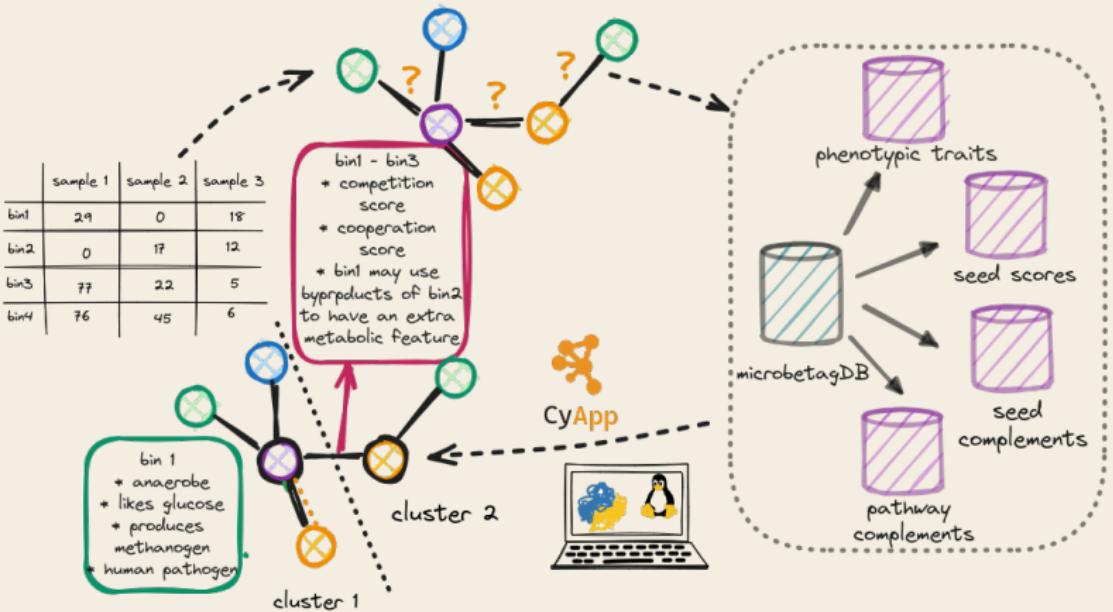
3 channels of information - 1 framework





Annotating co-occurrence networks

microbetag software platform



Microbial interactions' inference using metabolic modelling

- dingo

dingo



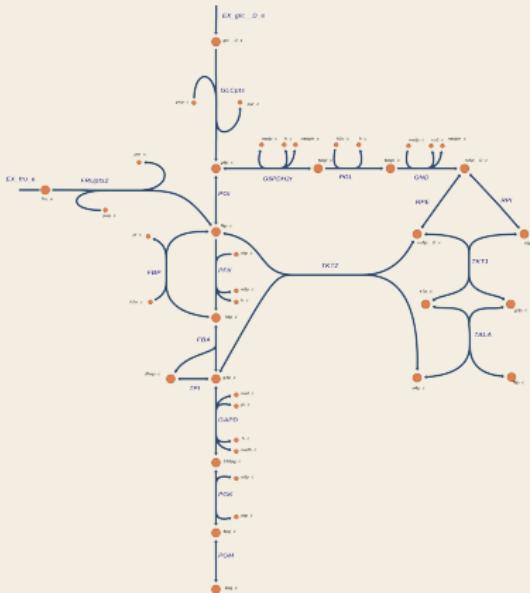
support flux sampling at high dimensional polytopes such as those of complex organisms and multispecies and/or host-microbe communities

Definitions

- *flux*: the rate of turnover of molecules through a reaction
- *polytope*: a bounded polyhedron
- *flux sampling*: calculation of a sufficiently large number of uniformly distributed points in the polytope derived from a metabolic model



Metabolic modelling and the biomass function



Metabolic models allow us to move from a metabolic map to mathematical structures the study of which may provide fundamental biological insight.

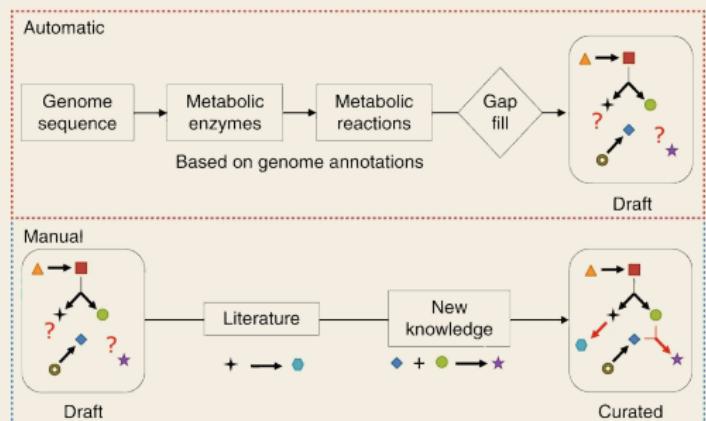


Figure from: Heirendt et al. Nature protocols 14.3 (2019): 639-702.



From a stoichiometric matrix to a constraint-based model

Reactions

	R ₁	R ₂	R ₃	R ₄	R ₅
Metabolites	-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

S-matrix

Flux vector

In a **steady state**
the production rate
of each metabolite
equals its consumption rate.

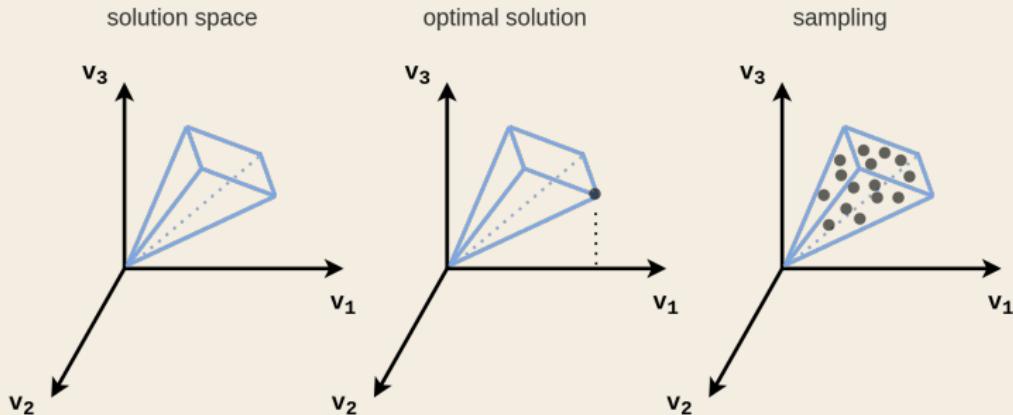
The **flux vector** is a vector with
the value of
each reaction flux
in a certain steady
state.

The steady state assumption
is ensured by
the **zero-vector**.



Flux sampling

an alternative approach



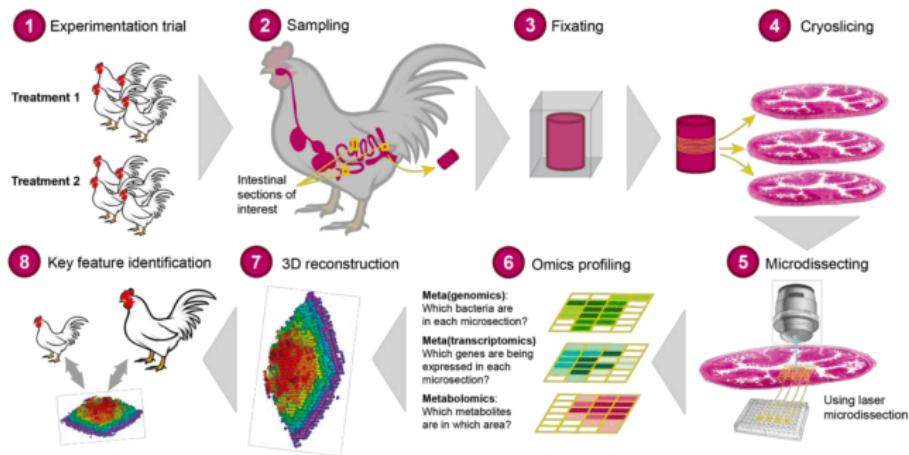
- enables the analysis of GEMs without the need of an objective function
- determines the feasible solution spaces for fluxes in a network based on a set of conditions as well as the probability of obtaining a solution

The 3D' omics approach

- Building a 3D representation of the chicken and the swine gut using cryoslicing, microdissection and multi-omics

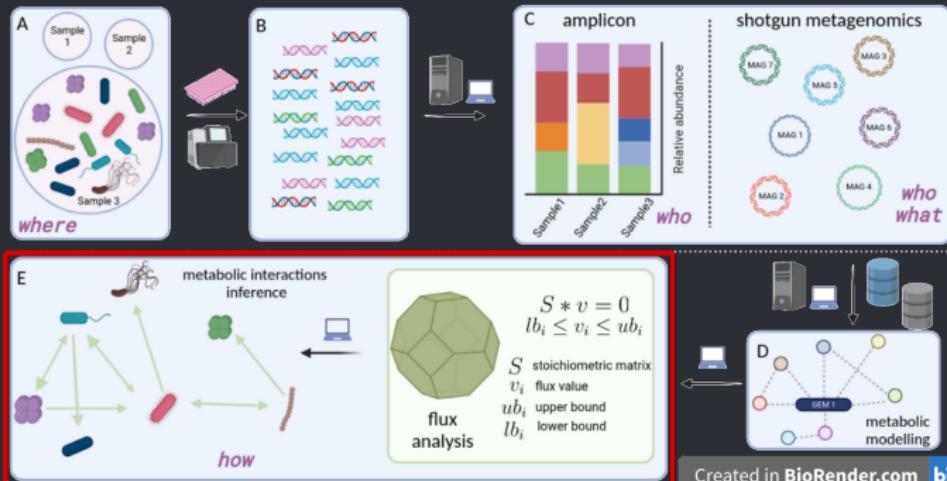
Network and metabolic modelling

my task



- [1] Chalkis, A., Fisikopoulos, V., Tsigaridas, E. & **Zafeiropoulos, H.** dingo: a Python library for metabolic networks analysis (*minor revision in Bioinformatics Advances*)
- [2] Boer, M. D., Melkonian, C., **Zafeiropoulos, H.**, Haas, A. F., Garza, D., & Dutillh, B. E. Improving genome-scale metabolic models of incomplete genomes with deep learning. bioRxiv, 2023-07. (*under review in iScience*)
- [3] Finn, Robert D., et al. Establishing the ELIXIR Microbiome Community. **F1000Research** 13 (2024): ELIXIR-50.
- [4] Joseph, C., **Zafeiropoulos, H.**, Bernaerts, K., & Faust, K. Predicting microbial interactions with approaches based on flux balance analysis: an evaluation. **BMC bioinformatics** 25.1 (2024): 36.
- [5] Garza, D.R., Gonze, D., **Zafeiropoulos, H.**, Liu, B. & Faust, K. Metabolic models of human gut microbiota: advances and challenges **Cell Systems** 14.2 (2023): 109-121.
- [6] **Zafeiropoulos, H.** et al. metaGoflow: a workflow for the analysis of marine Genomic Observatories shotgun metagenomics data **GigaScience** 12 (2023): giad078.
- [7] Pavloudi, C. & **Zafeiropoulos, H.** Deciphering the community structure and the functional potential of a hypersaline marsh microbial mat community **FEMS Microbiology Ecology** 98.12 (2022): fiac141.
- [8] Paragkamian, S., ..., **Zafeiropoulos, H.**, ..., & Gerovasileiou, V. Automating the curation process of historical literature on marine biodiversity using text mining: the DECO workflow **Frontiers in Marine Science** 9 (2022): 940844.
- [9] **Zafeiropoulos, H.**, Paragkamian et al. (2022). PREGO: a literature- and data-mining resource to associate microorganisms, biological processes, and environment types. **Microorganisms** 10(2), 293.
- [10] **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.
- [11] Polymenakou, P.N., Nomikou, P., **Zafeiropoulos, H.**, ..., Kyrpides, N.C., Kotoulas, G. & Magoulas, A. (2021). The santorini volcanic complex as a valuable source of enzymes for bioenergy. **Energies**, 14(5), p.1414.
- [12] 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)**.
- [13] **Zafeiropoulos, H.**, Gioti, A. et al. Os and 1s in marine molecular research: a regional HPC perspective. **GigaScience**, 10(8), (2021) giab053.
- [14] **Zafeiropoulos, H.** 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) giaa022.

Research perspective



A (bit) more holistic framework

"a combination of quantitative high - throughput experiments and predictive metabolic models can elucidate the genotype - phenotype map of microbial metabolic strategies" - Bajic and Sanchez (2020)

My research questions focus on the role of auxotrophies in the different evolutionary strategies.

Acknowledgments

presented research was funded by..



Thank you for your time!

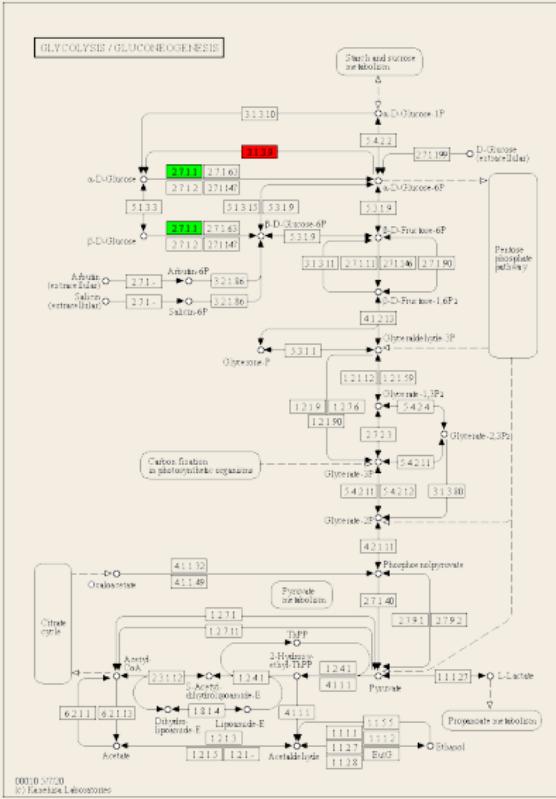
More than happy to discuss!

<https://hariszaf.github.io>

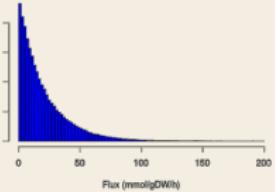
<mailto:haris.zafeiropoulos@kuleuven.be>



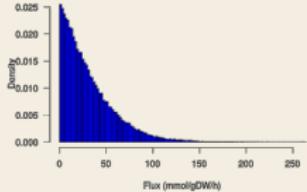
Flux sampling output marginal distributions and copulas



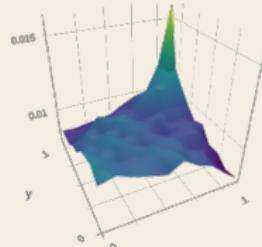
HEX



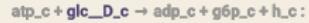
G6PPer



HEX - G6PPer



HEX:

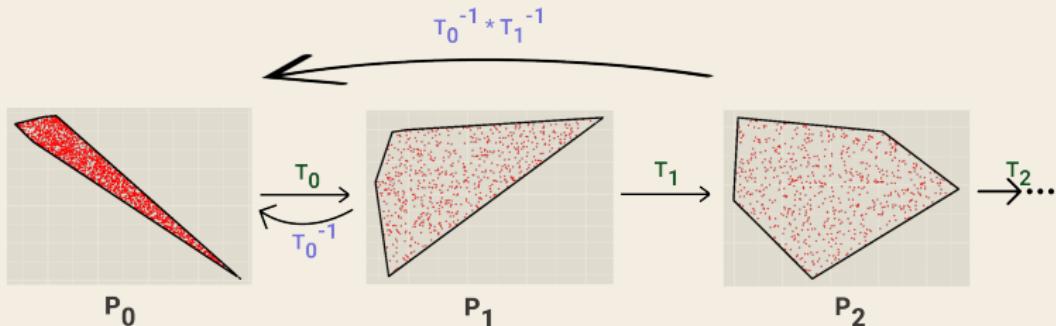


G6PPer:





A new Markov Chain Monte Carlo 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

From concentrations to fluxes to study changing environments

We can describe the mass balance of a chemical compound as the difference between the sum of the fluxes of all the reactions that form it and the sum of all that degrade it.

$$\frac{d\omega_i}{dt} = \sum_k s_{ik} v_k = \langle s_i, v \rangle$$

and thus:

$$\frac{d\omega}{dt} = Sv$$

The region of steady states

moving to full dimensional polytope

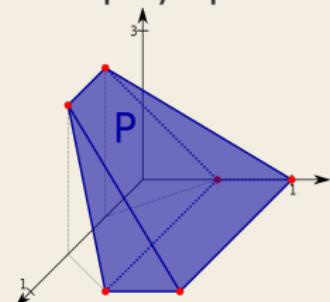
The *constraints* on the reactions fluxes.

$$Sv = 0, \quad (1) \quad v = Nx$$

$$v_{lb} \leq v \leq v_{ub}$$

$$S \in \mathbb{R}^{m \times n}, v \in \mathbb{R}^n$$

As a *full dimensional polytope*



$$P := \{x \in \mathbb{R}^d | Ax \leq b\}$$

$N \in \mathbb{R}^{n \times d}$ denotes the matrix of the null space of S , i.e. $SN = 0_{m \times d}$.

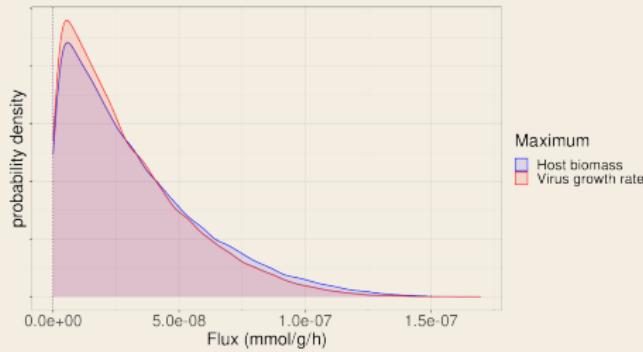
By replacing v with Nx in Equation 1, we get the full dimensional polytope P , where
 $A = \begin{pmatrix} I_n N \\ -I_n N \end{pmatrix}$ and $b = \begin{pmatrix} v_{ub} \\ v_{lb} \end{pmatrix} N$, (in \mathbb{R}^d).

Find possible targets against SARS-CoV-2

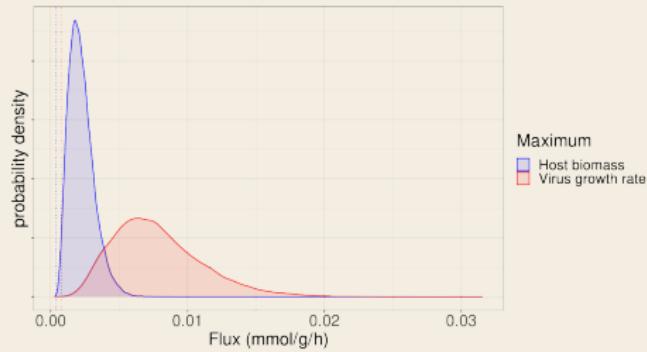
a flux sampling application



Reaction TYMSULT



Reaction GK1



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

For more about this example case, you may check this [blog-post](#).