COMBI

Gathering and handling Big data

From metagenomics to ecosystems behaviors

Damien Eveillard,
Computational Biology; Université de Nantes, France & Tara Global Ocean Systems Ecology and Evolution

December 2nd 2020
S4GES Workshop
Musing on the concept of Good Environmental Status: the complexity of the status & the status of complexity

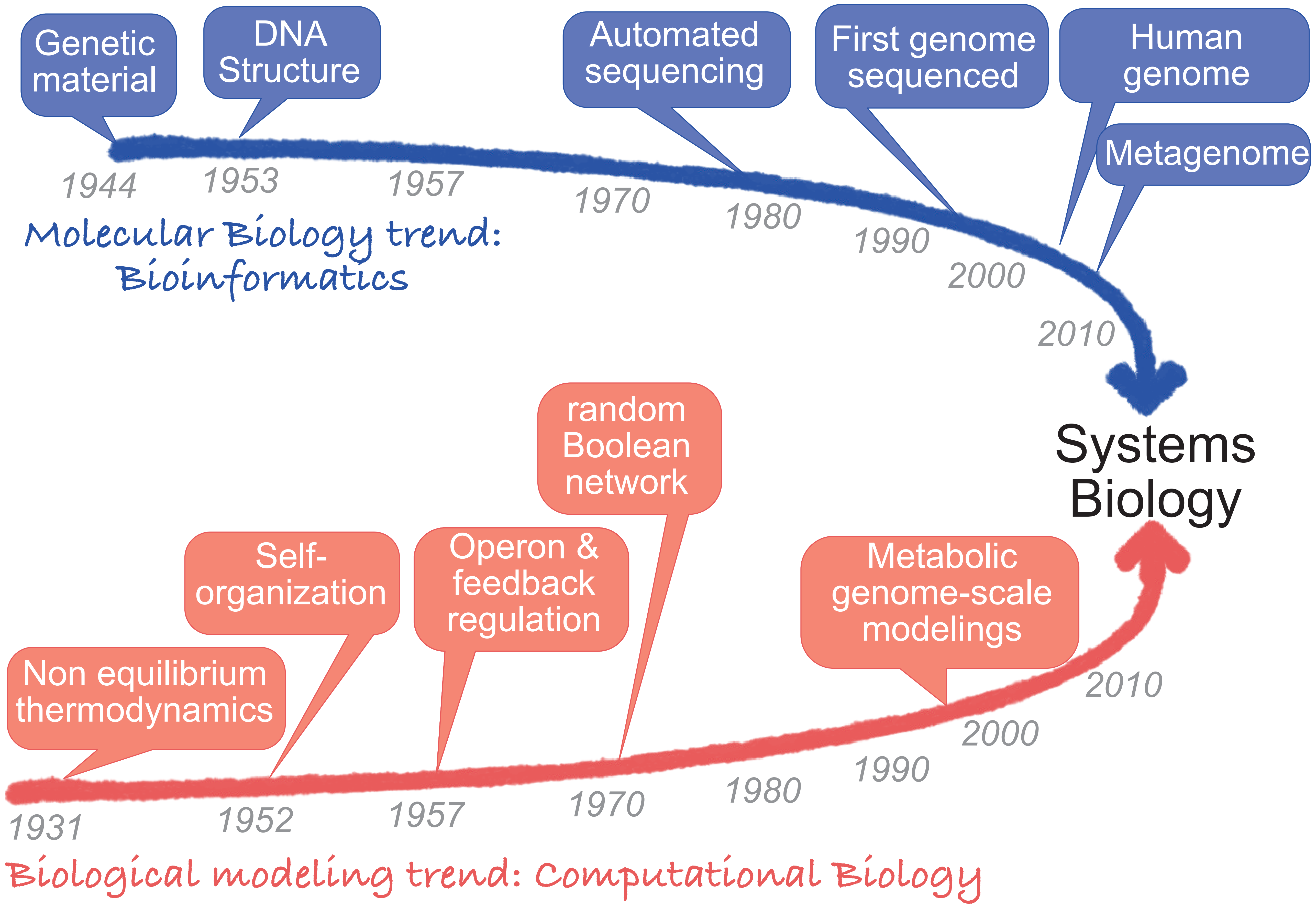
LS2N LABORATOIRE
DES SCIENCES
DU NUMÉRIQUE
DE NANTES

UNIVERSITÉ DE NANTES

cnrs

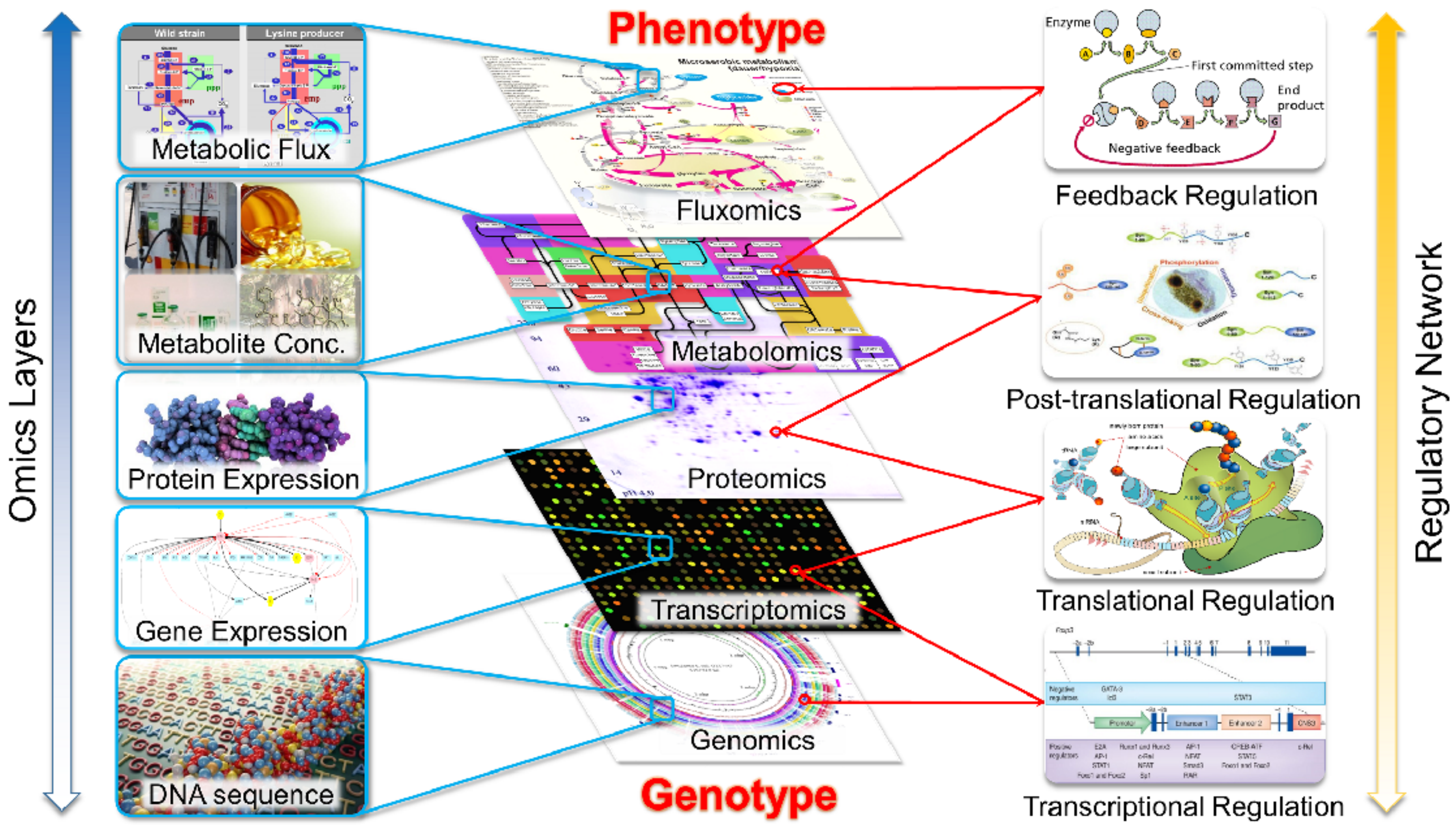
Systems Biology

Historical perspective driven by data and concepts



Adapted from Westefhoff & Palsson, Nat. Biotechnology 2004

Systems Biology in practice
(from discrete data to almost quantitative ones)



[Gao & Feng, PloS One, 2016]

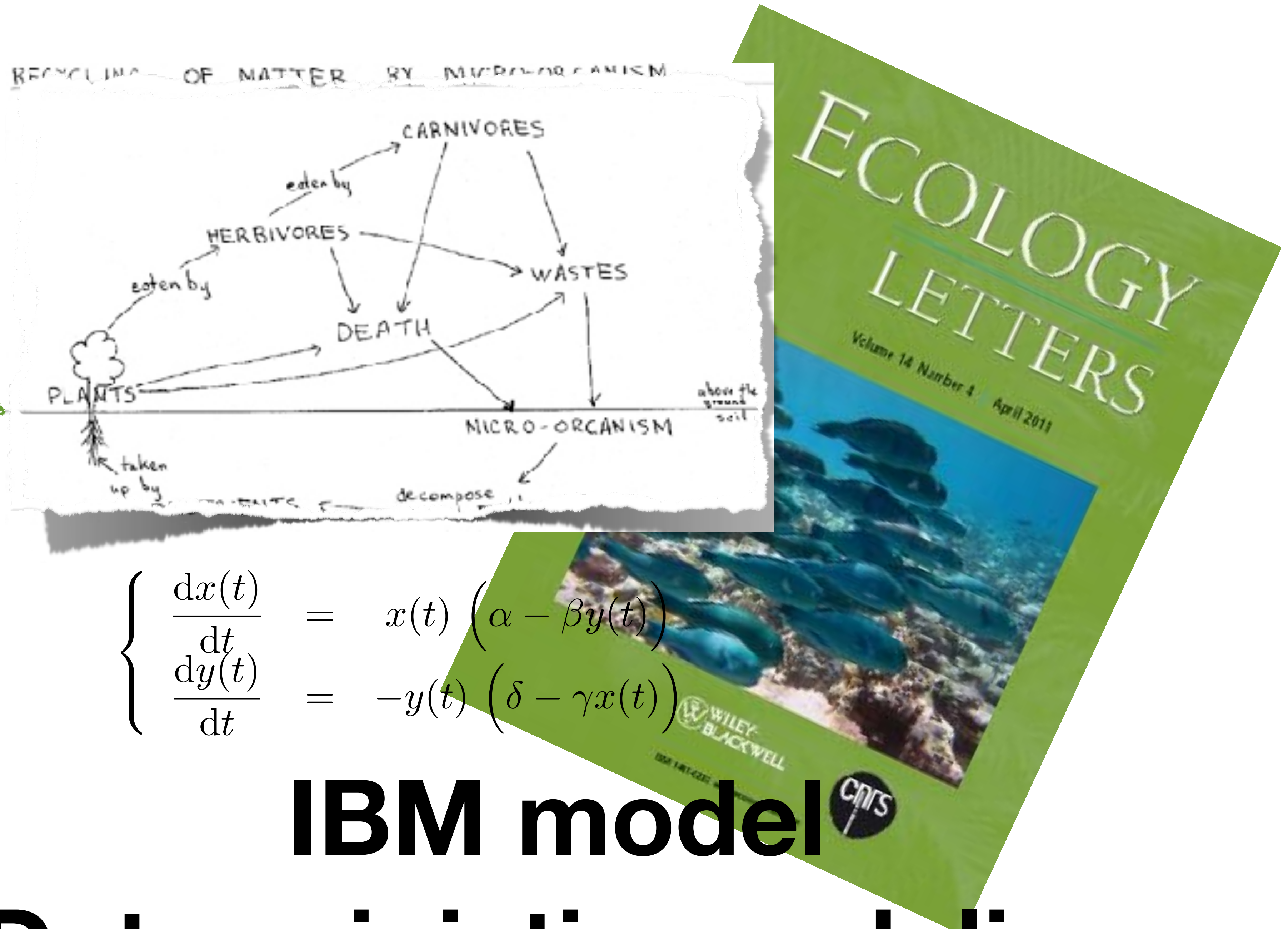
1. Composition of biological systems
2. Analysis of the biological system structure
3. Simulating the dynamics of the biological system
4. Control of the biological system

«Systems Ecology»

Metagenomic,
Metranscriptomic

new abundant observations

Complex ecosystems

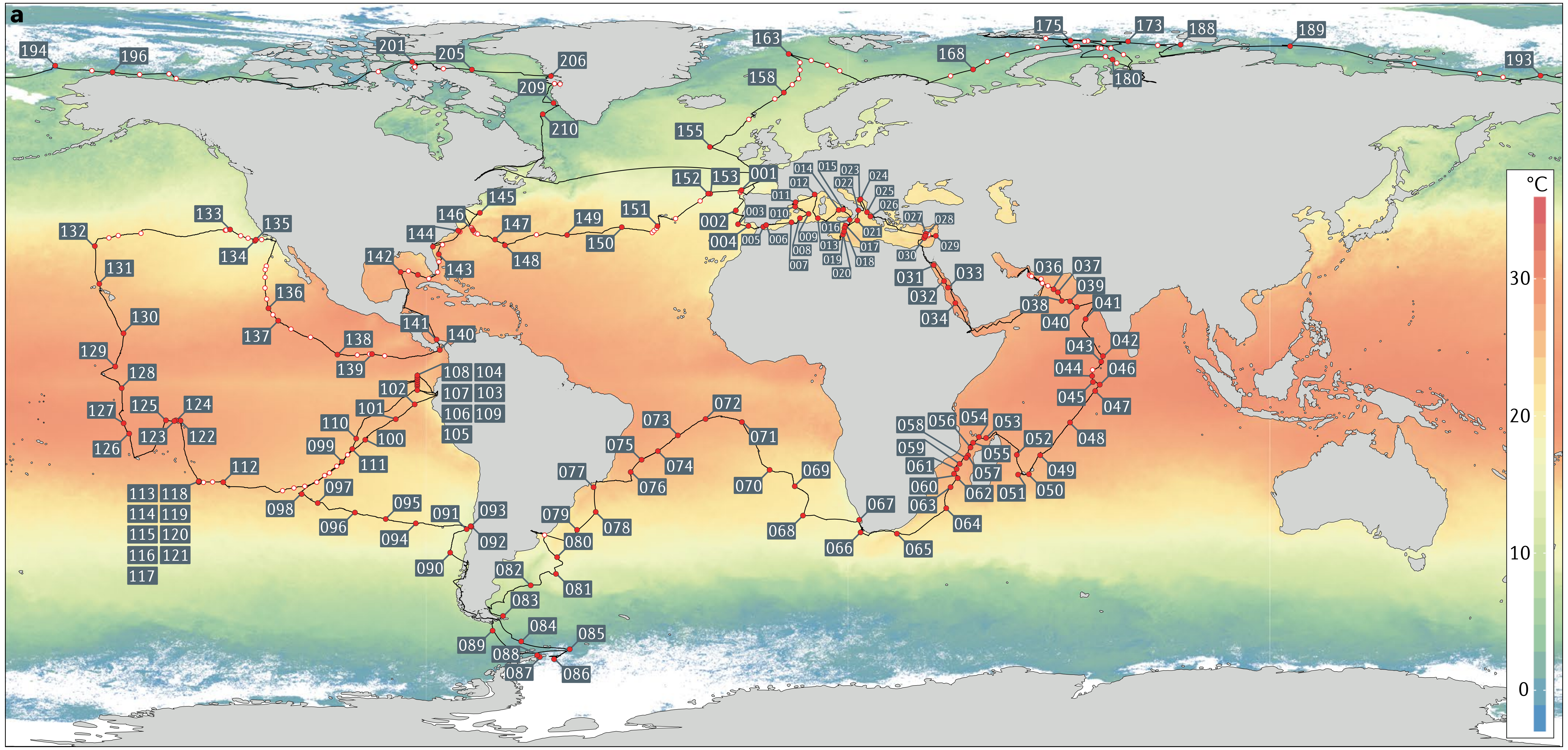


$$\begin{cases} \frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t)\left(\alpha - \beta y(t)\right) \\ \frac{\mathrm{d}y(t)}{\mathrm{d}t} = -y(t)\left(\delta - \gamma x(t)\right) \end{cases}$$

IBM model
Deterministic modeling
Relationship between abundance & diversity

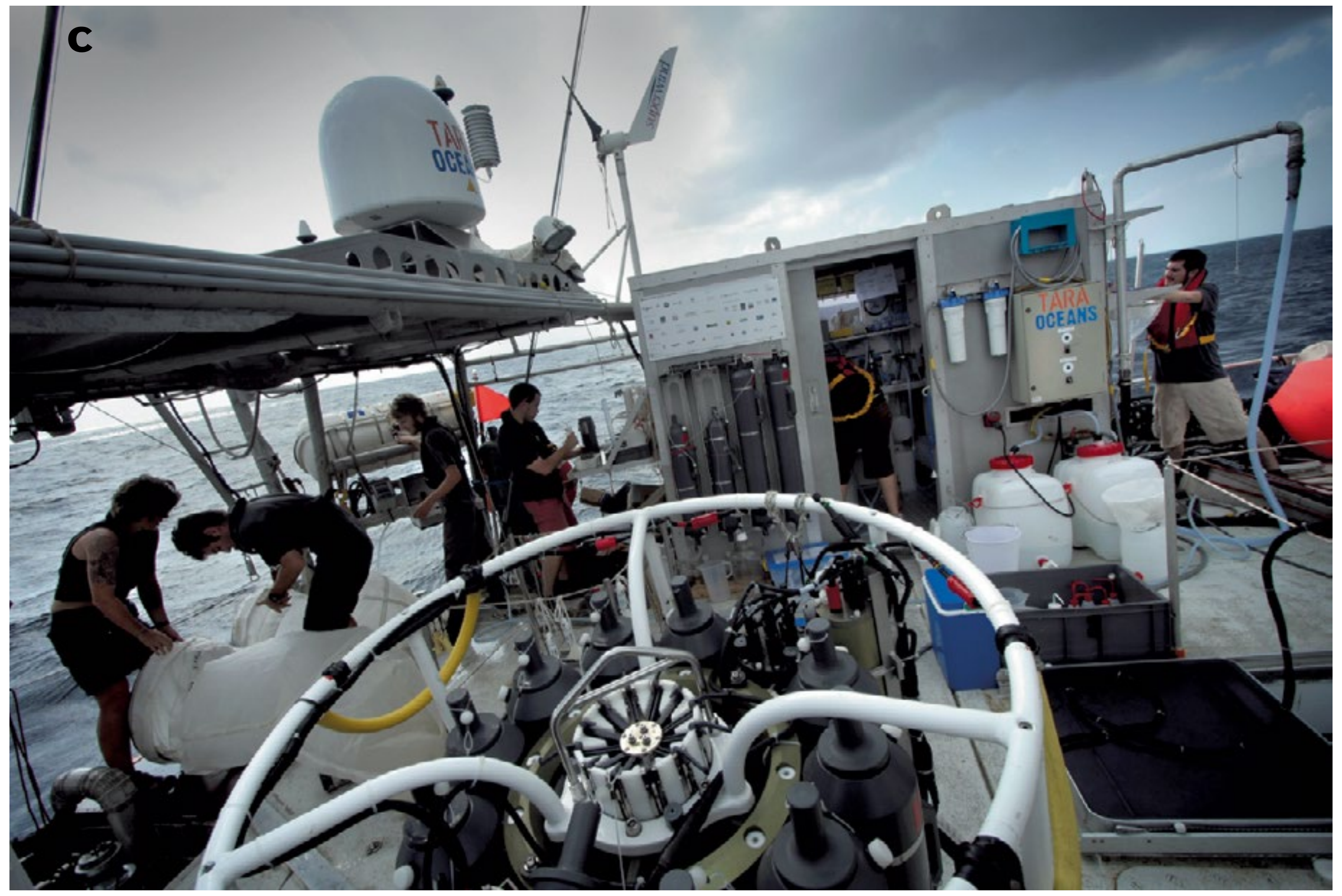
Extraction of emerging properties from complex ecological systems

A metagenomic case study: crushing numbers



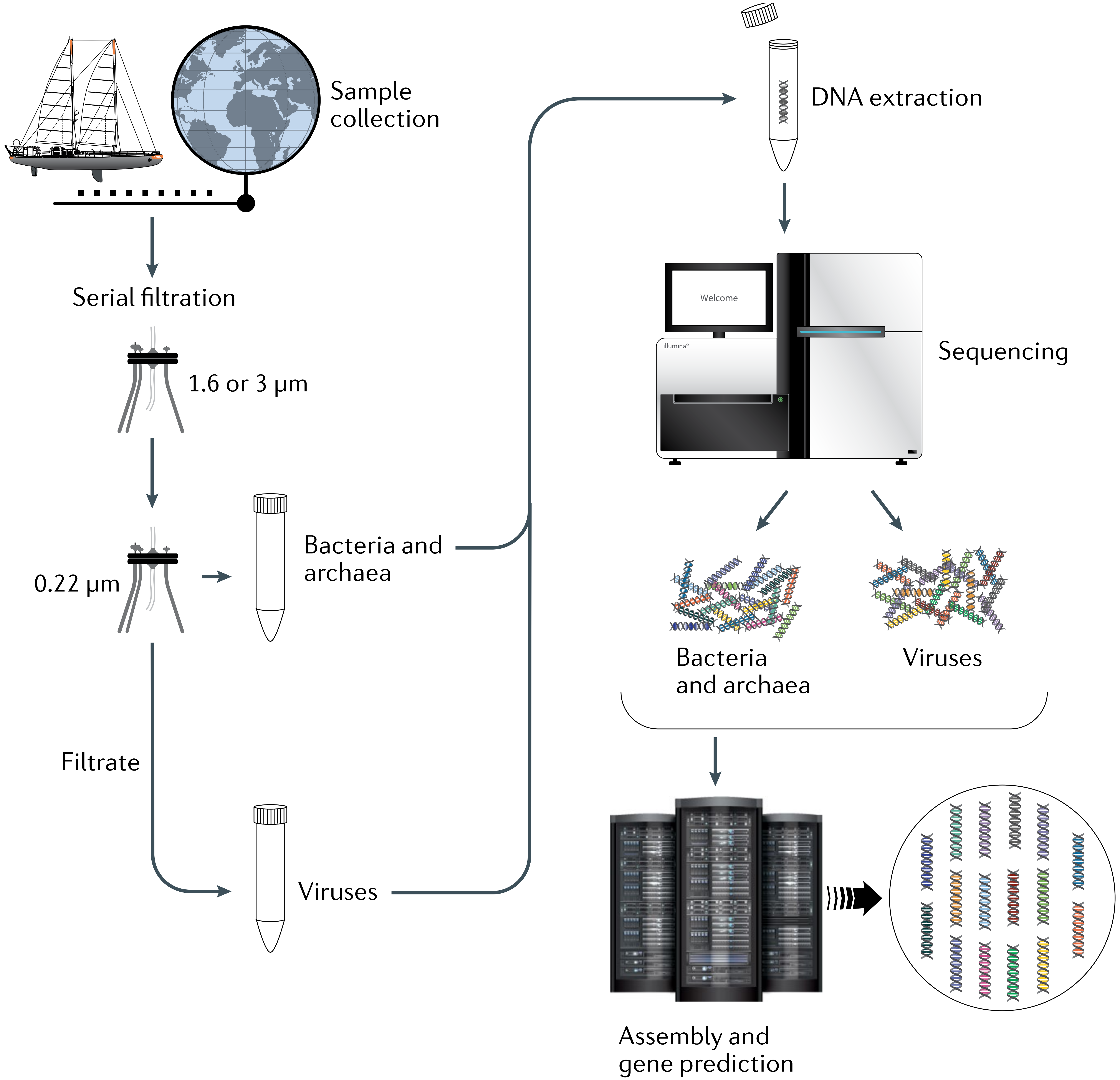
***Tara* Oceans (2009–2013)**

- 140,000 km sailed
- >35,000 plankton samples collected
- 210 sampling stations
- >60 terabases of DNA and RNA sequenced
- ~7 million images captured
- 120 crew members and scientists on-board
- 52 stopovers in 37 countries
- 35,000 schoolchildren on board at stopovers



Sunagawa et al., 2020,
Nature Microbiology
Review

Extraction of omics

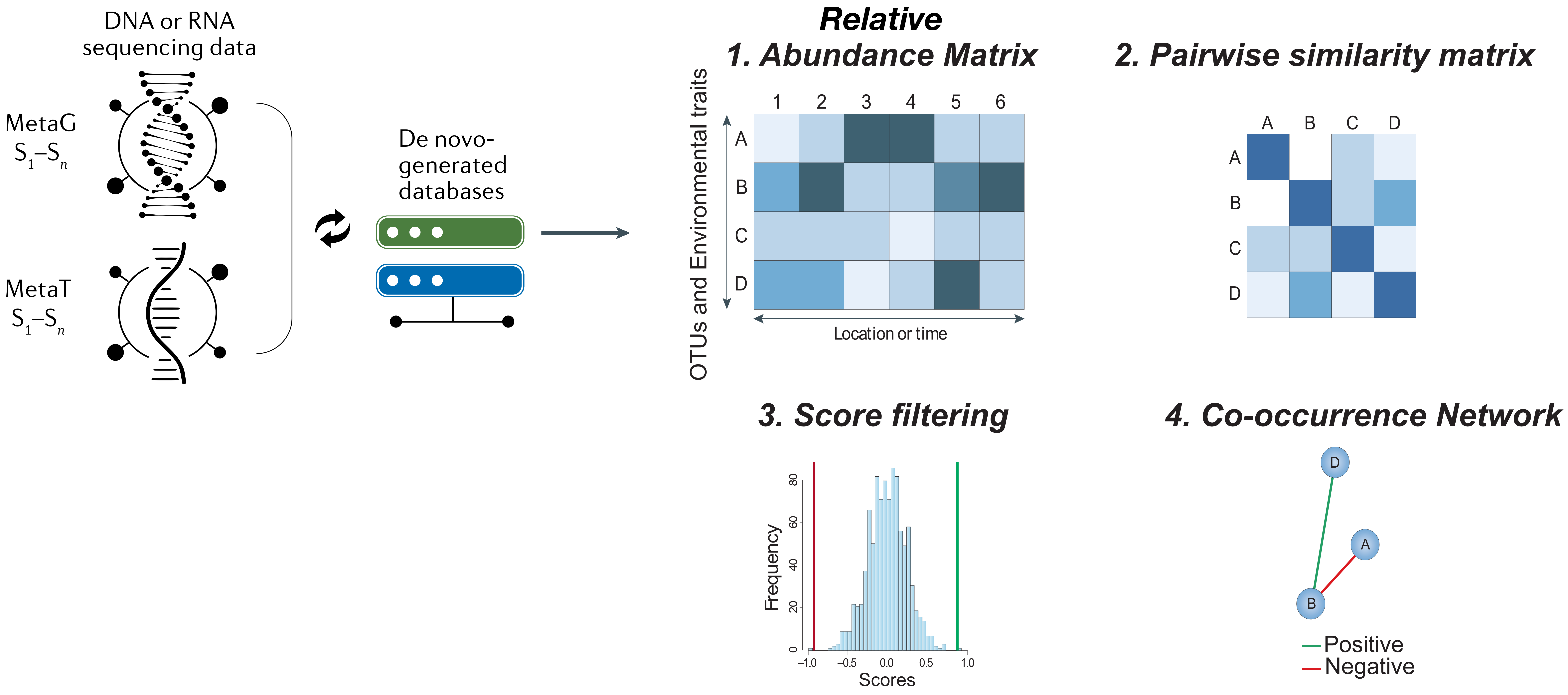


- Extraction of all genes (i.e., functions)
- Extraction of markers genes (i.e., diversity)

Sunagawa et al., 2020, Nature Microbiology Review

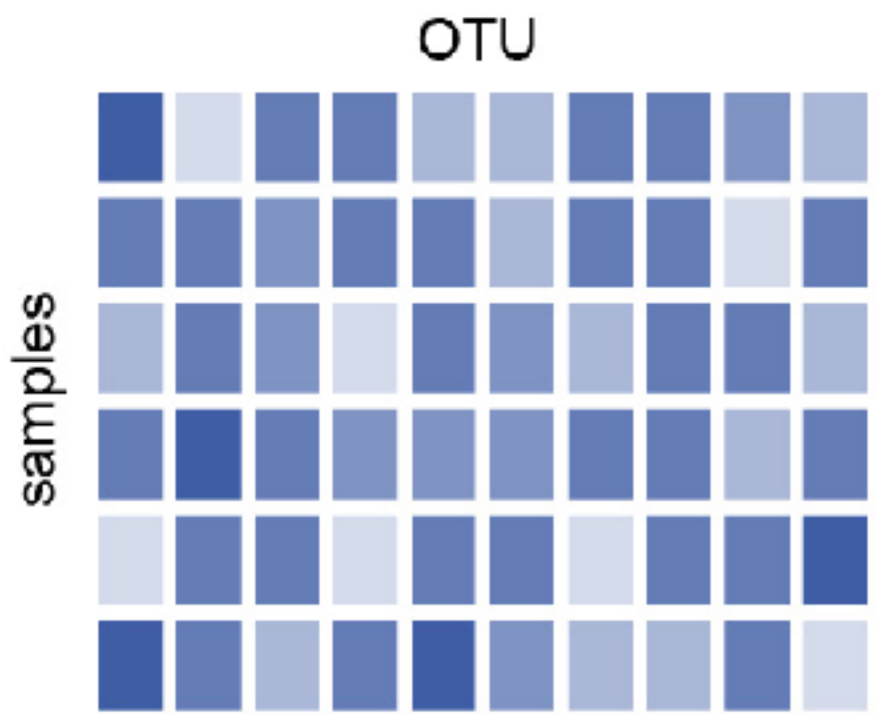
Graph: a discrete abstraction for omics

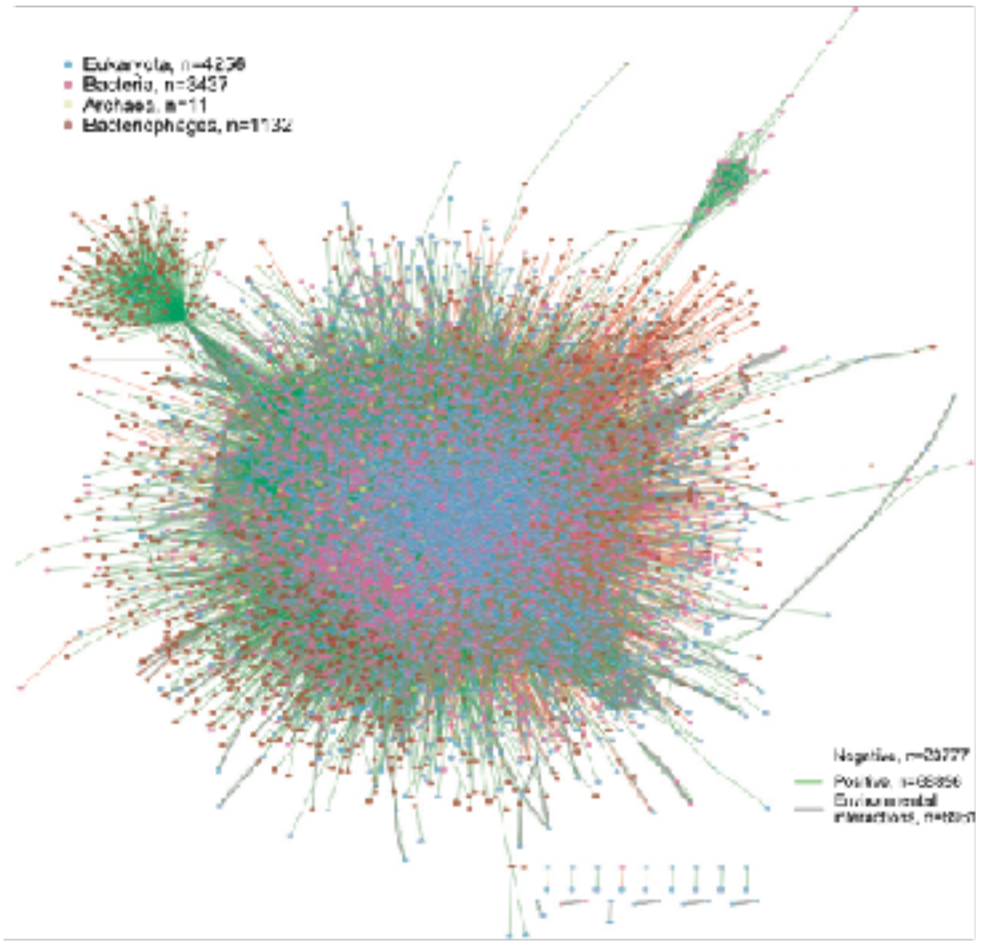
Building Co-occurrence networks

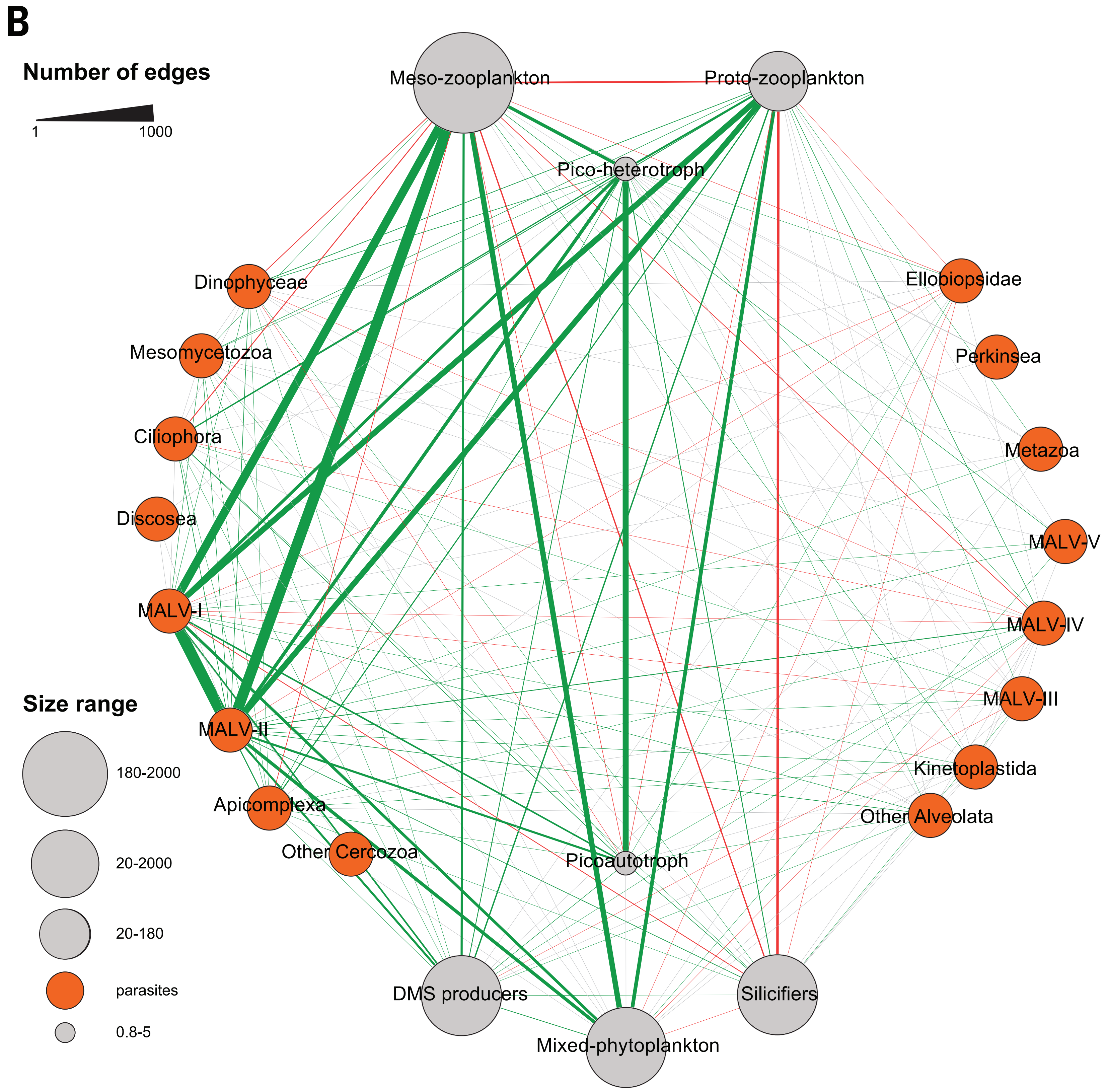


adapted from Faust & Jeroen, 2012, Nature Microbiology Reviews

Co-occurrence graph of TARA oceans

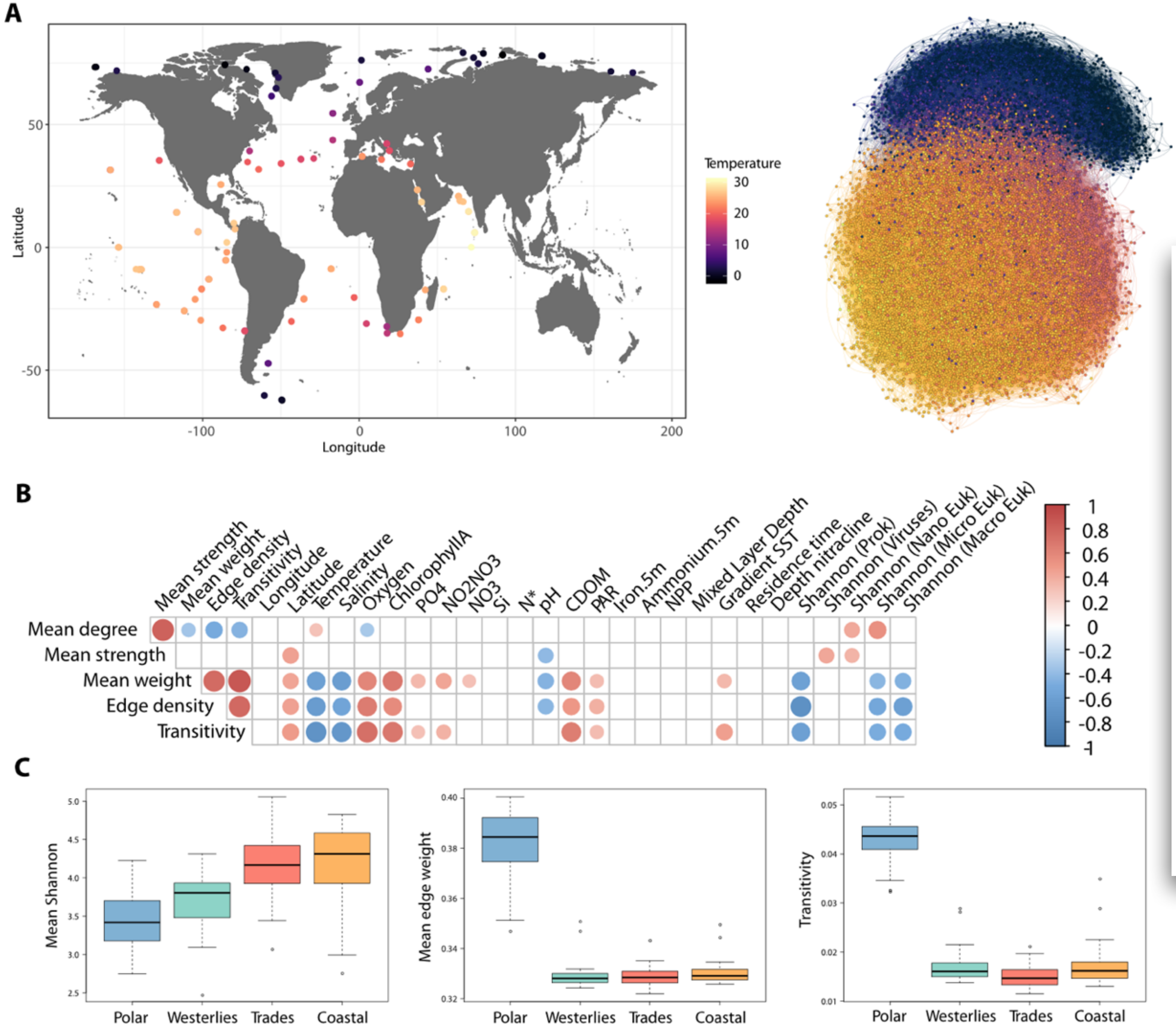


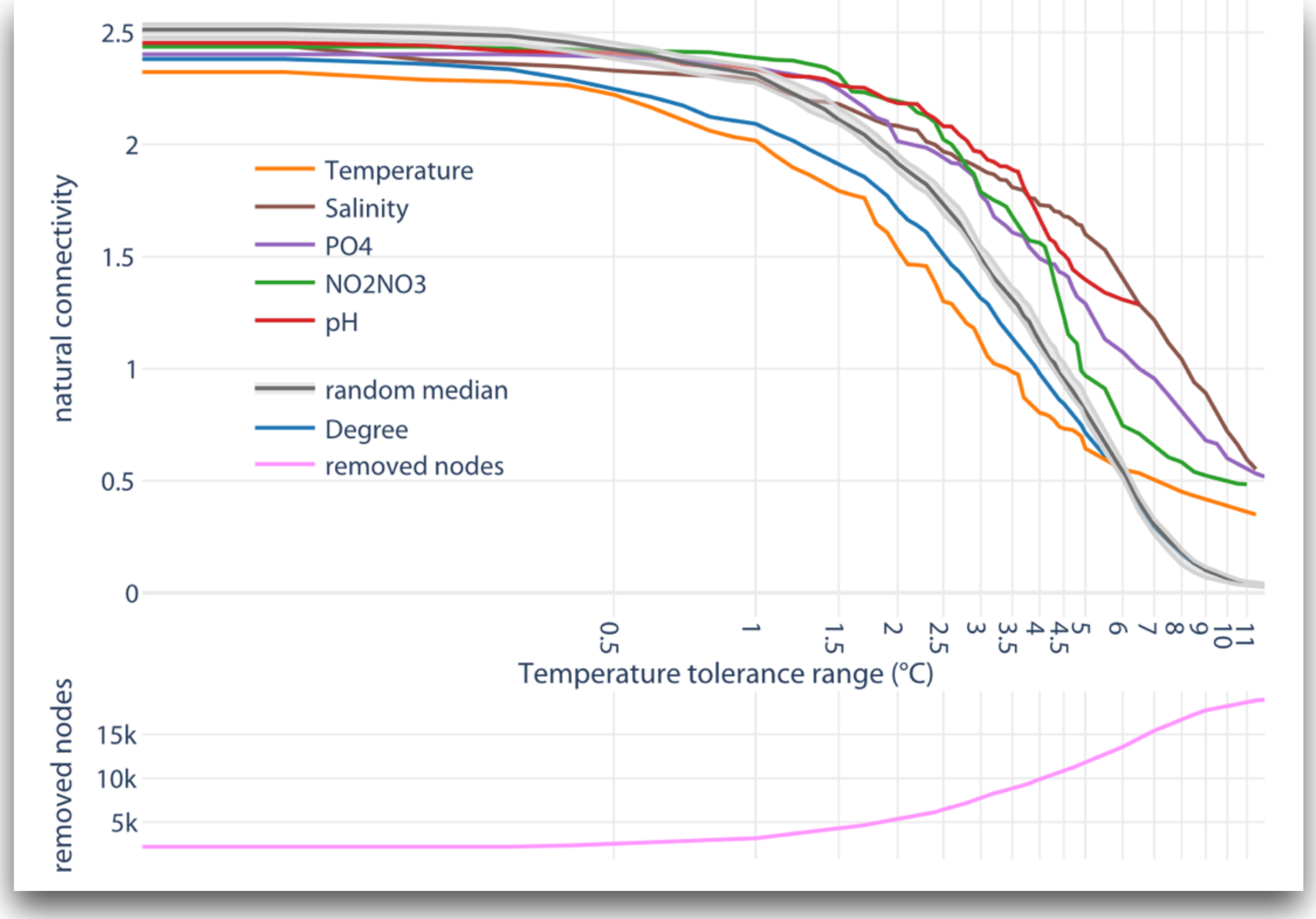




Lima Mendes et al., 2015, Science

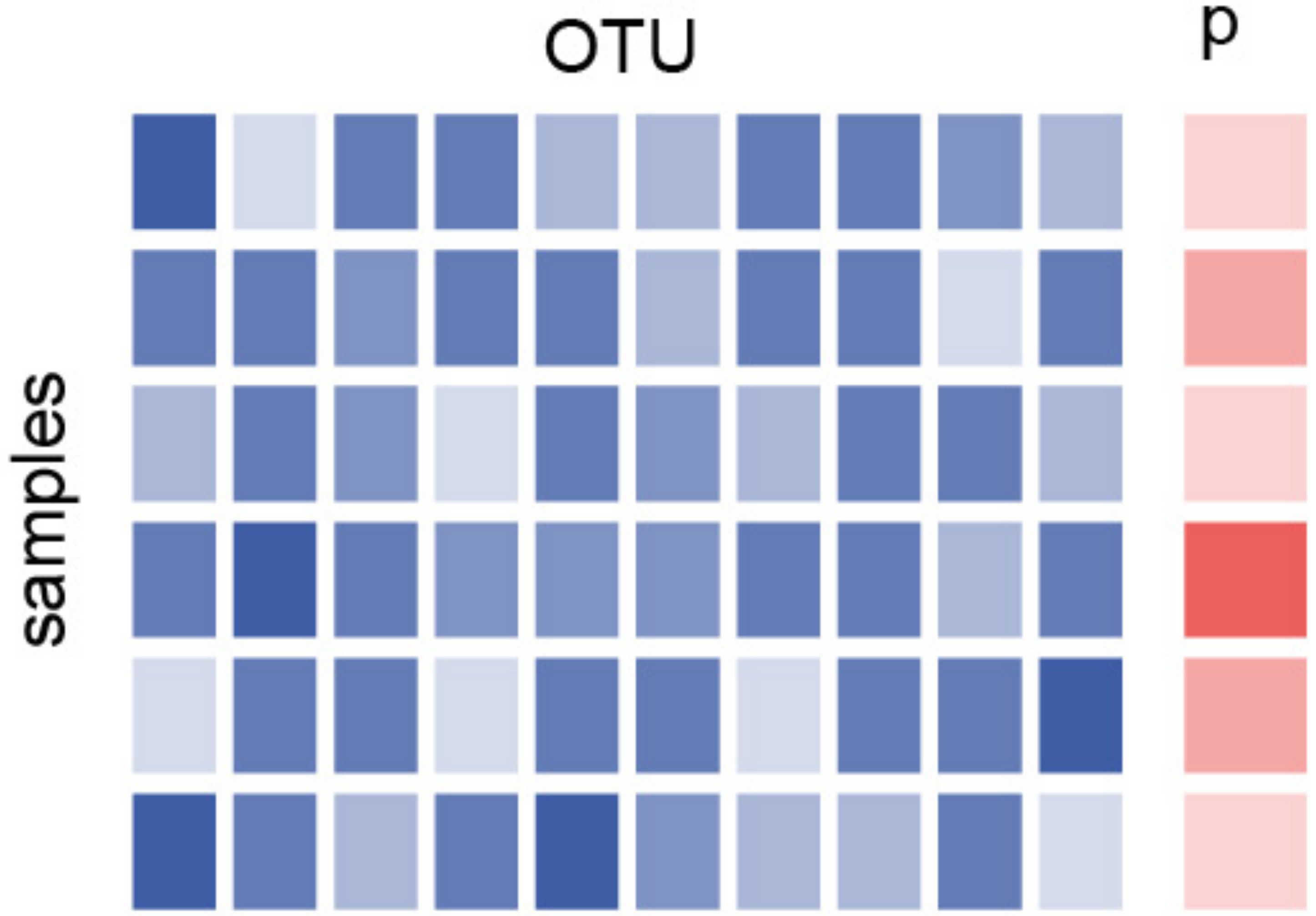
Global Ocean Graph shows a biogeography

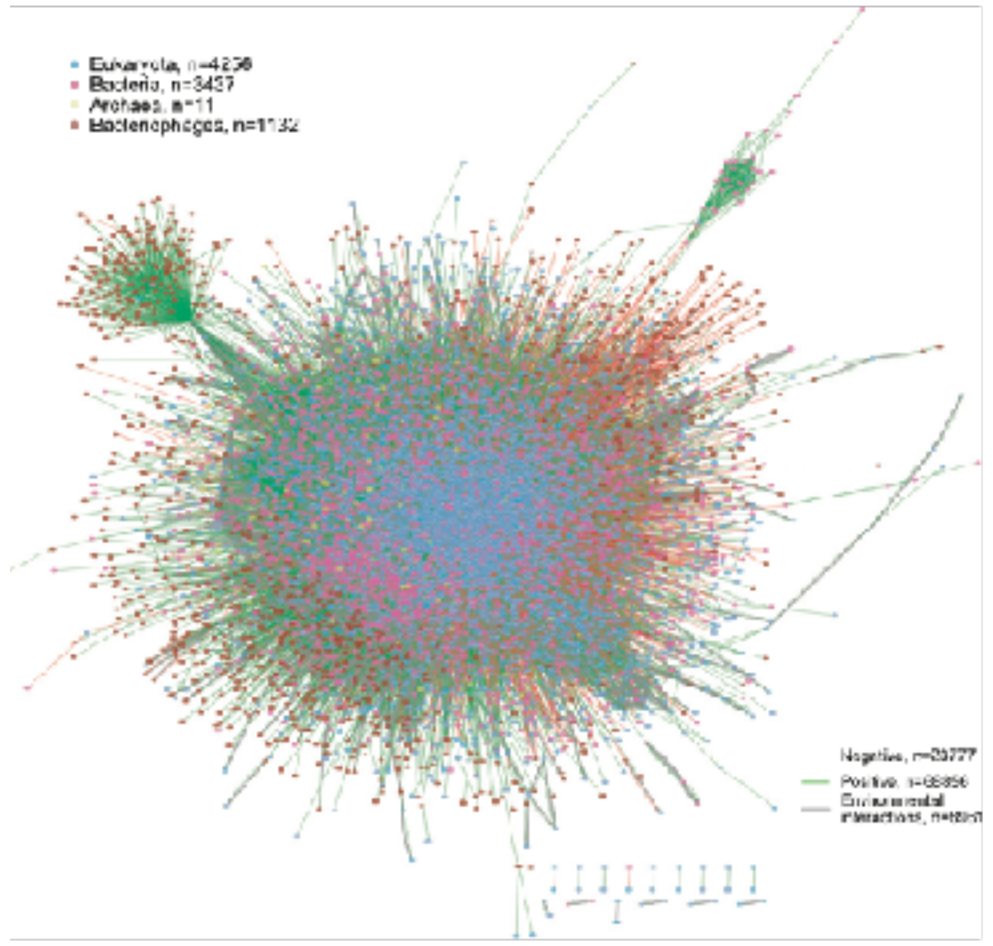




[Chaffron et al. 2020, Bioarxiv]

Integrating the graph topology with quantitative measurements?





- Nutrients
- Oxygen
- Light
- Temperature
- pH
- NPP
- Carbon export

Identification of subgraphs associated with these feature

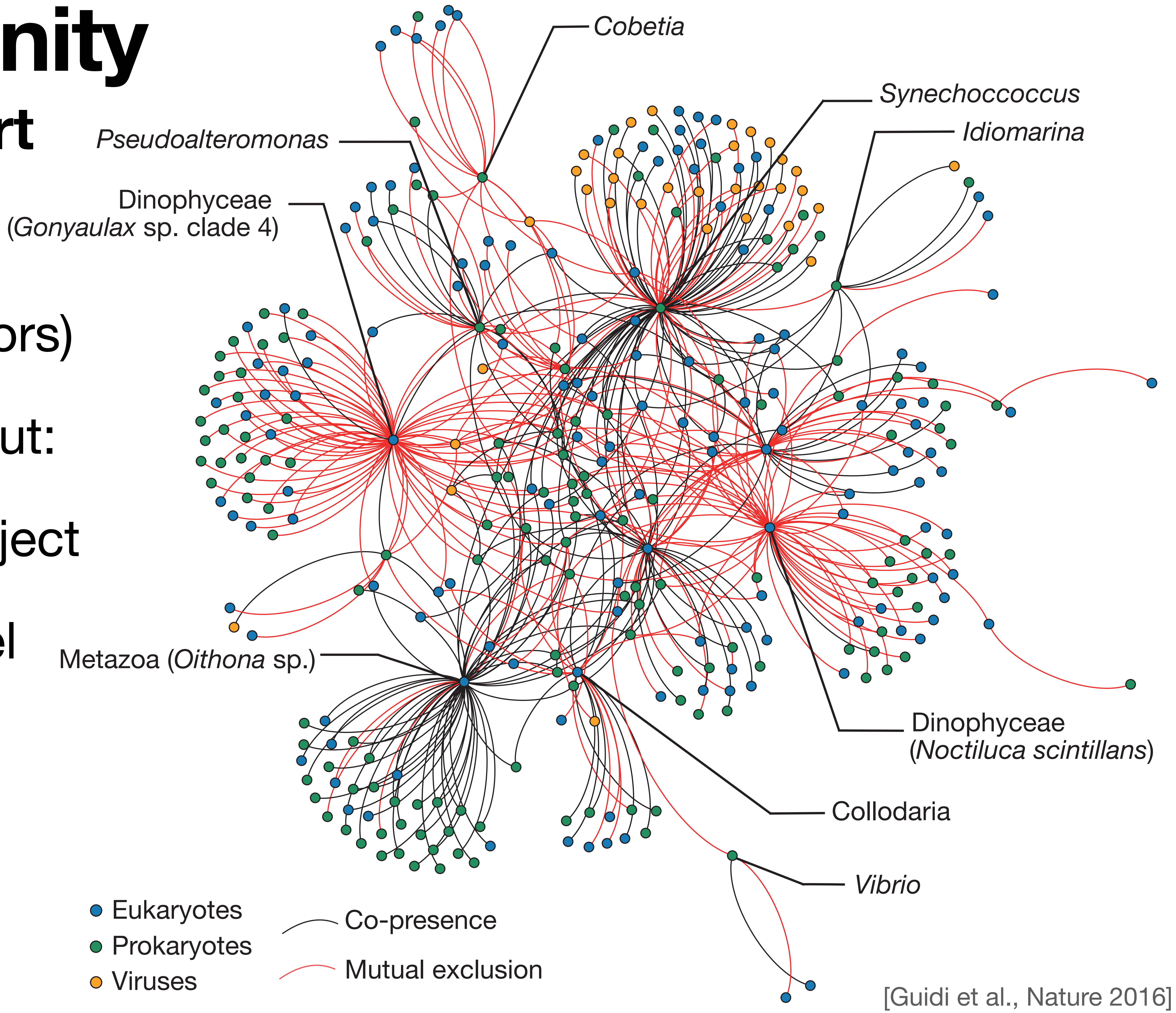
Building a community

associated to carbon export

- 51 genes are associated to C. export (using omics as predictors)
- Graph as a good abstraction but:

✓ Not just a discrete formal object

✓ It is not a network or a model



[Guidi et al., Nature 2016]

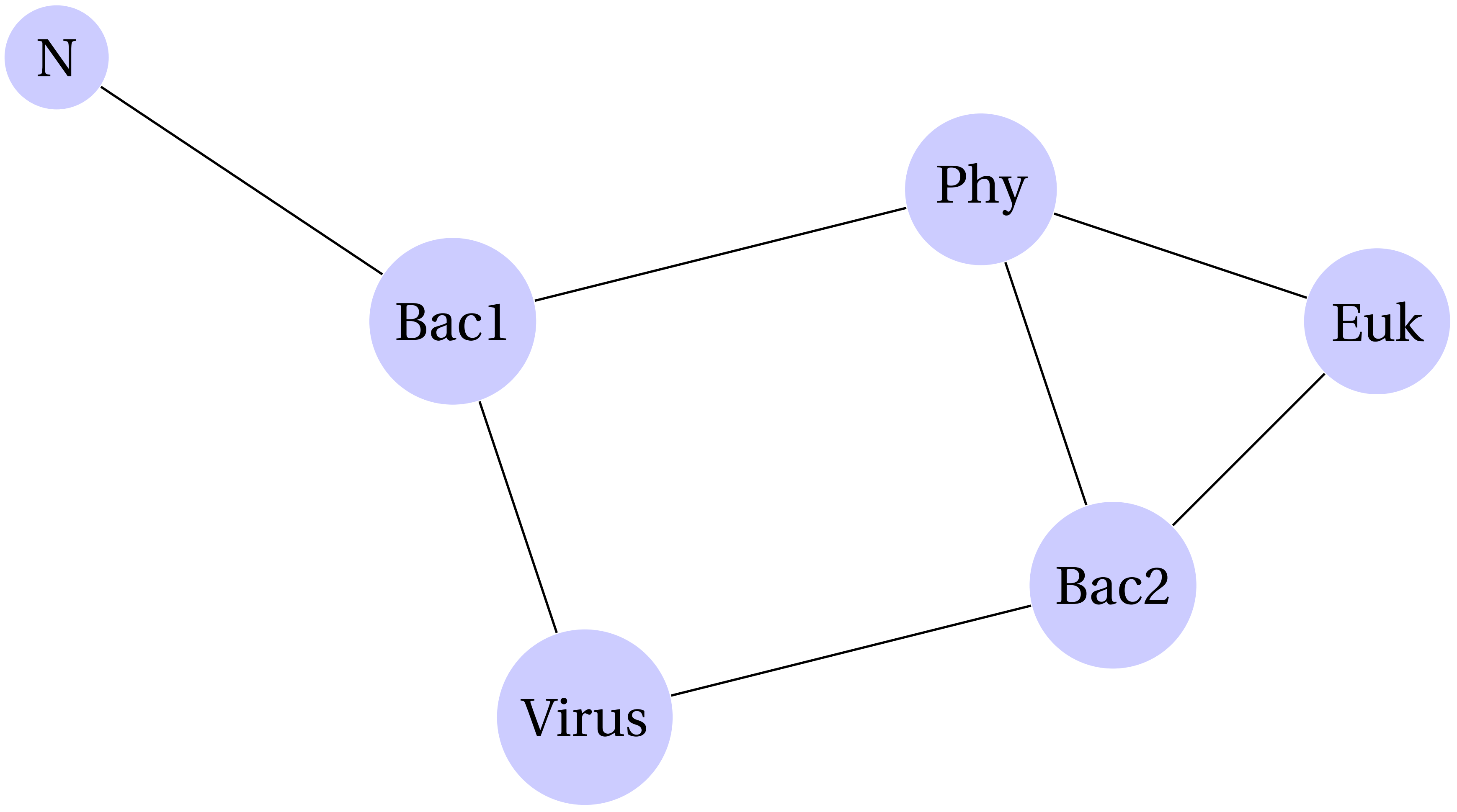
A short semantic point

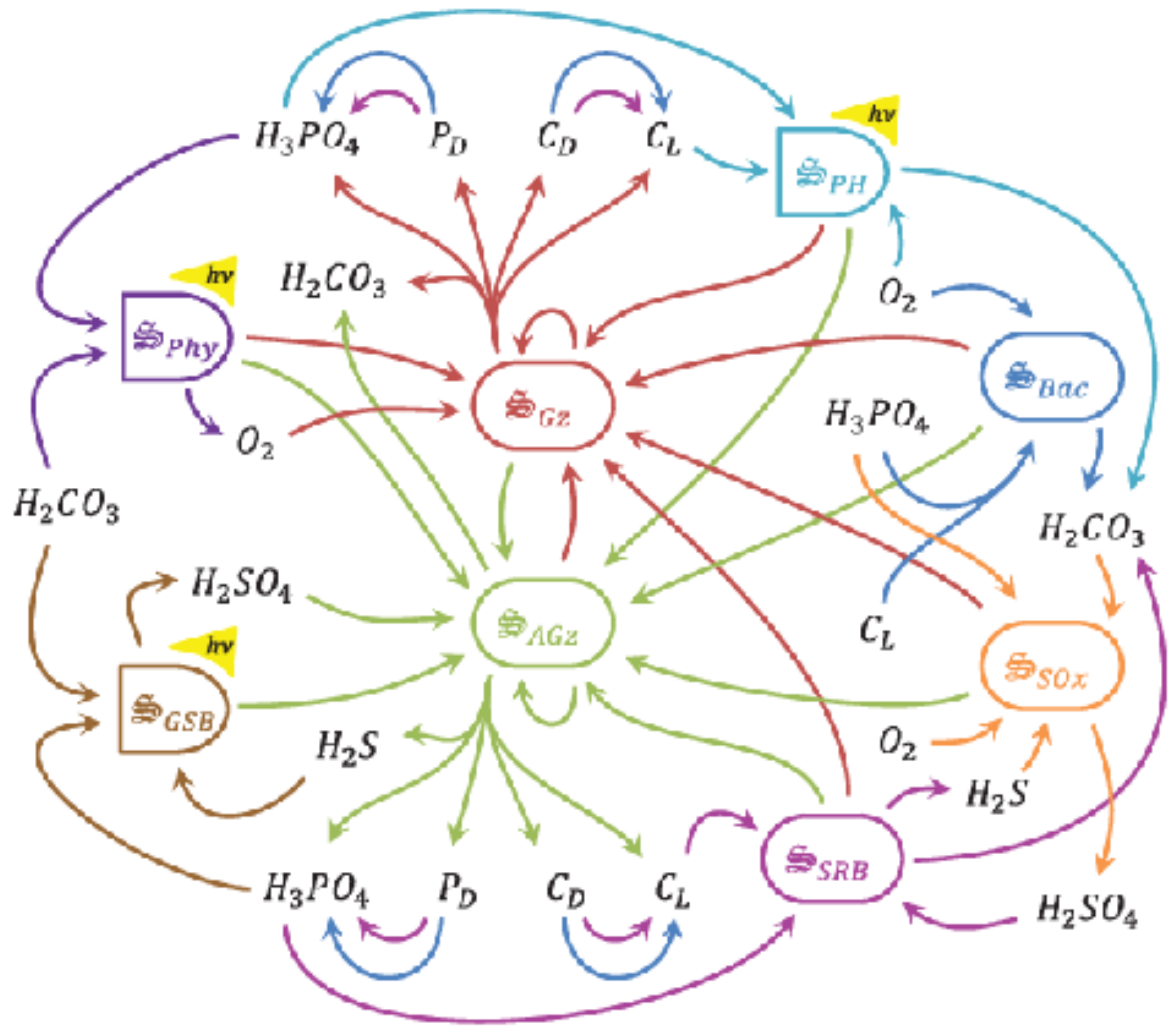
Graph VS. Network

- **Graph**: discrete abstraction that considers (colored) vertices connected by (un)directed (weighted) edges

- **Network**: abstraction of an automaton that considers states connected by flows

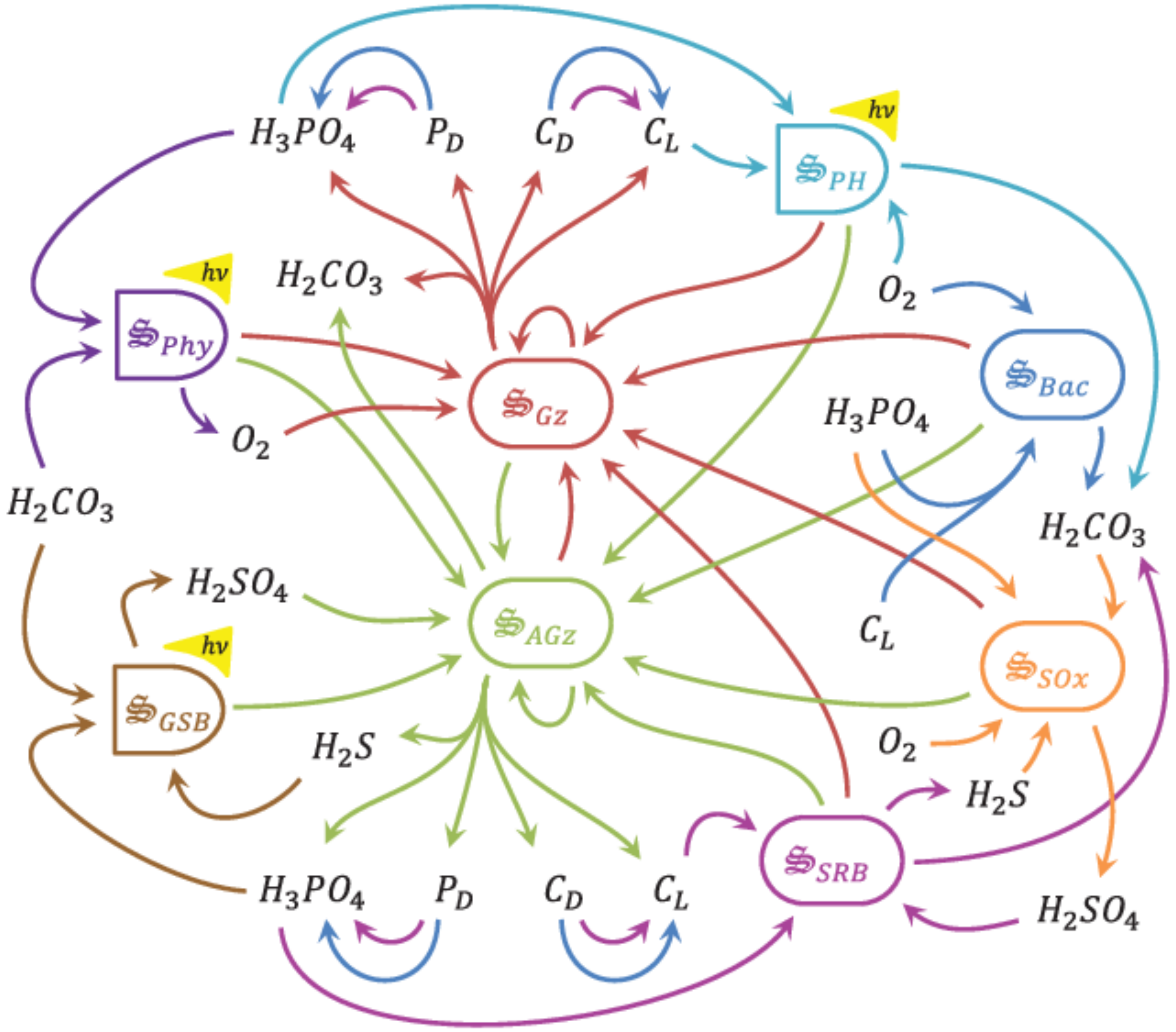
- **Model**: automaton that considers variables that evolve under (un)deterministic constraints



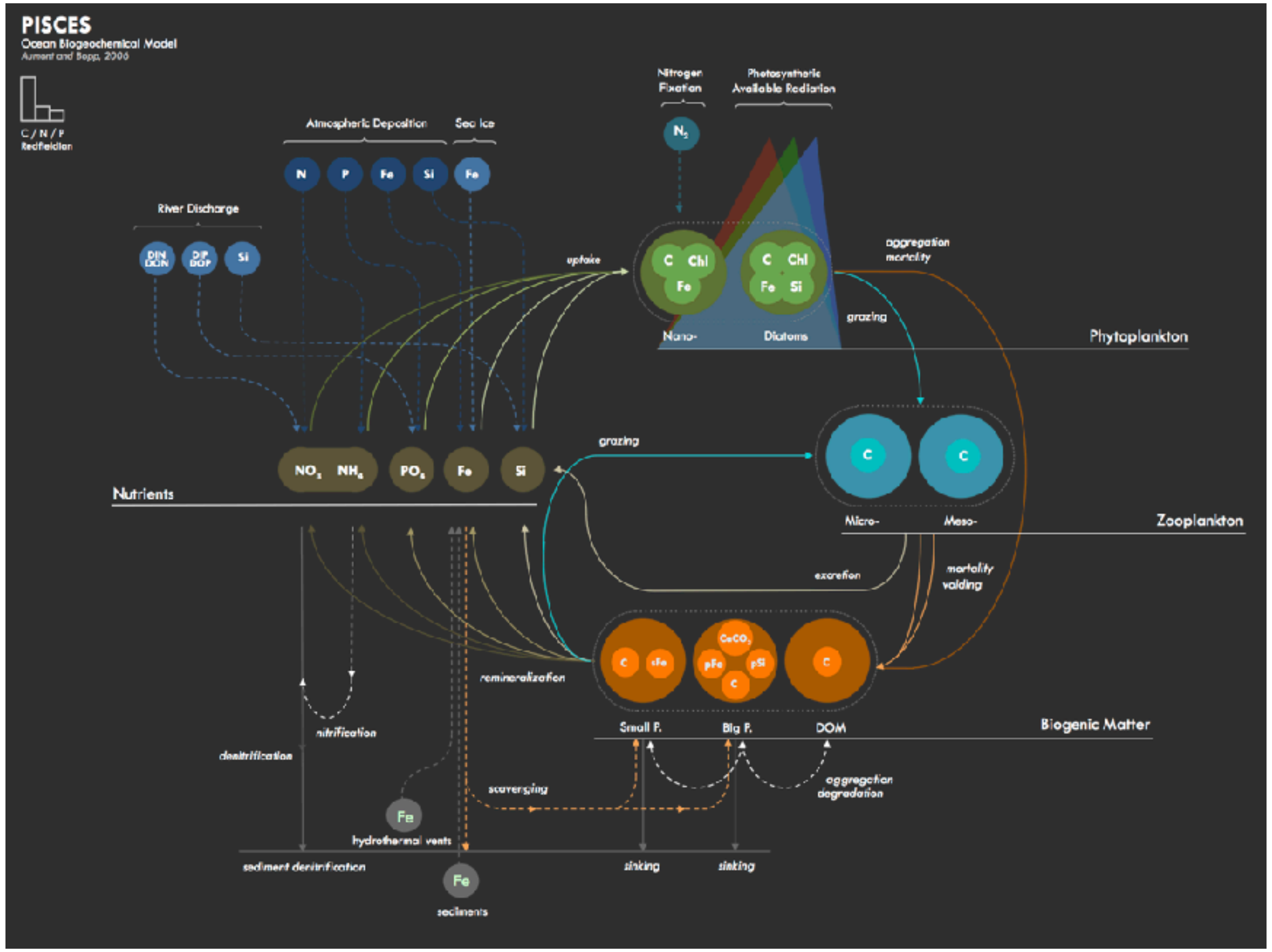


Dedicated biogeochemical modelings

a mechanistic network description

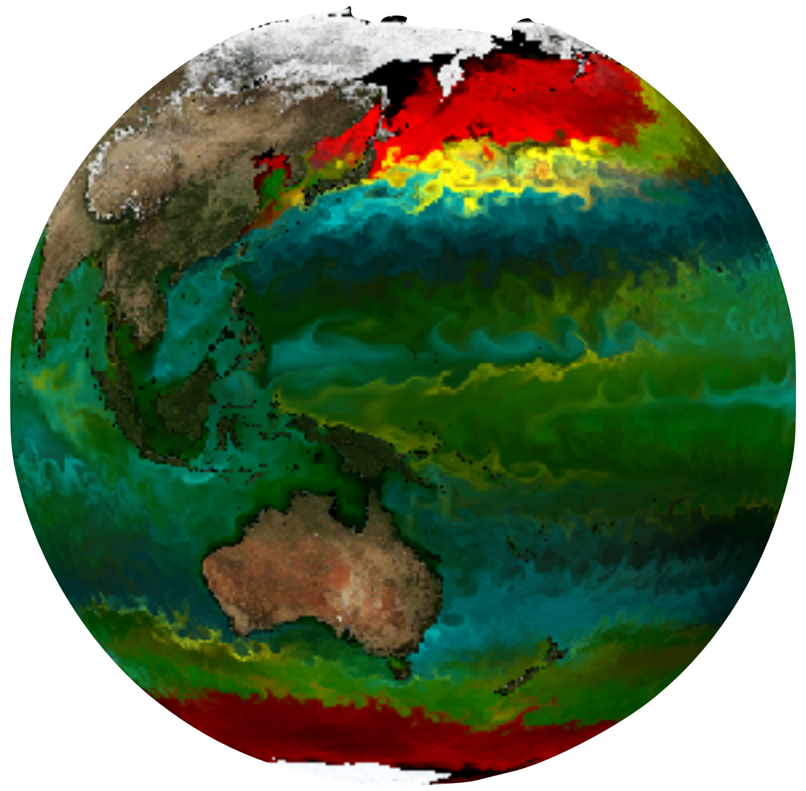


Valino & Huber, Frontiers Env. Science 2018

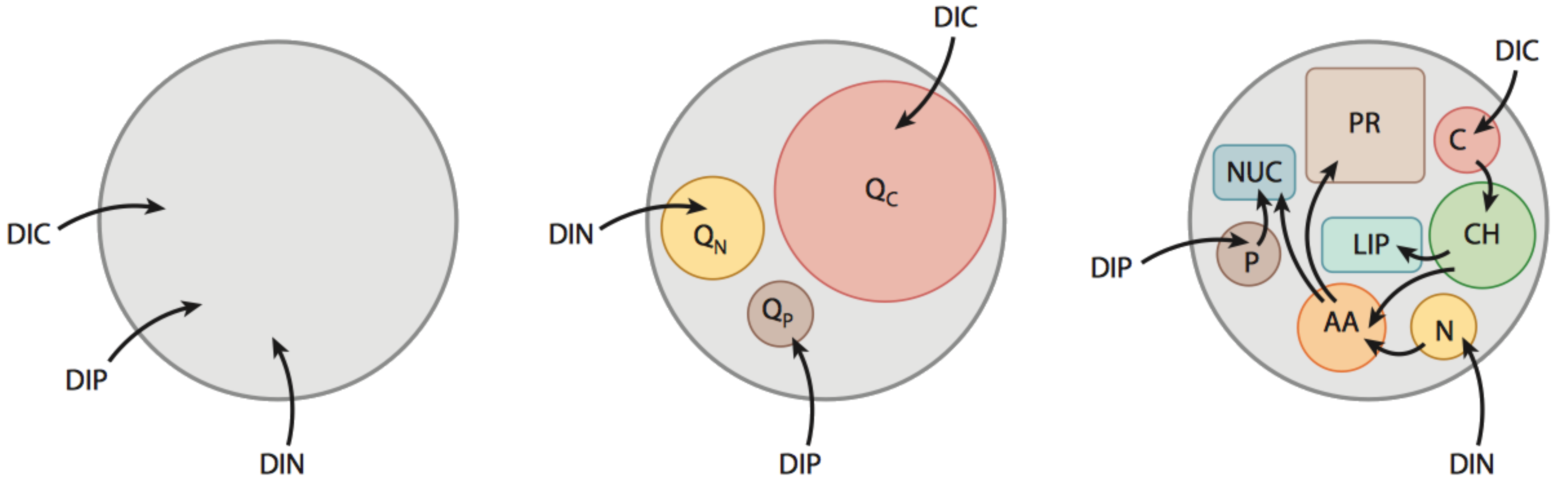


Aumont & Bopp, 2006

Adding biological complexity for global modeling

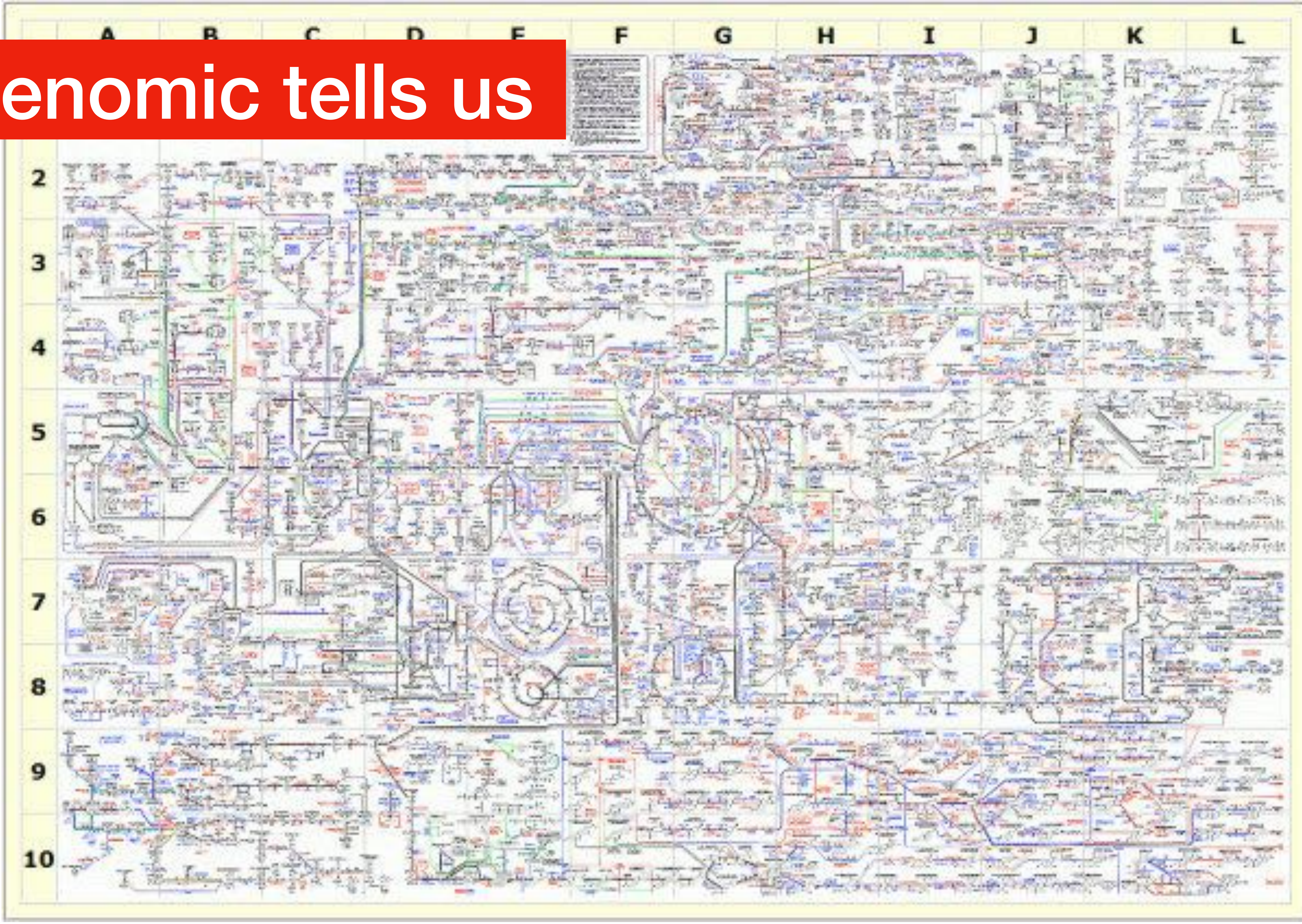


https://darwinproject.mit.edu/

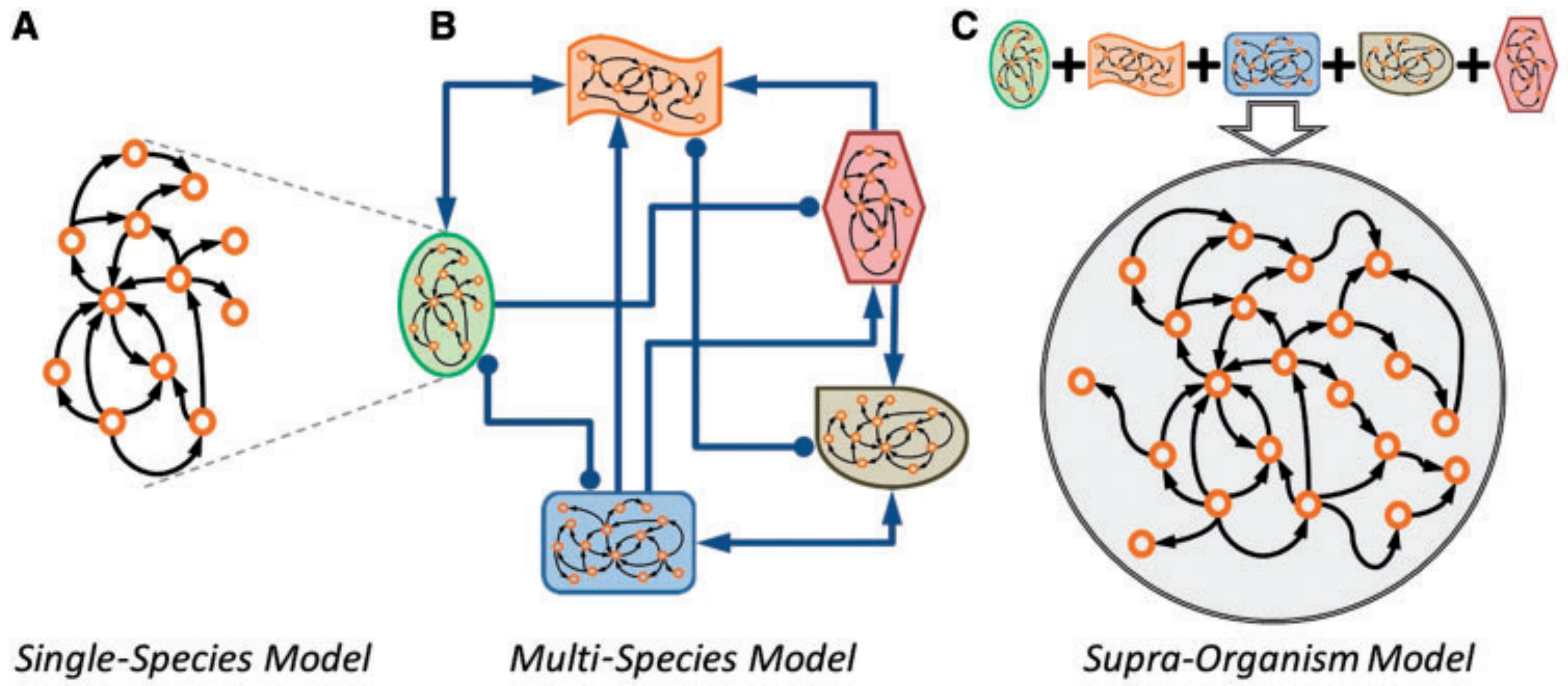


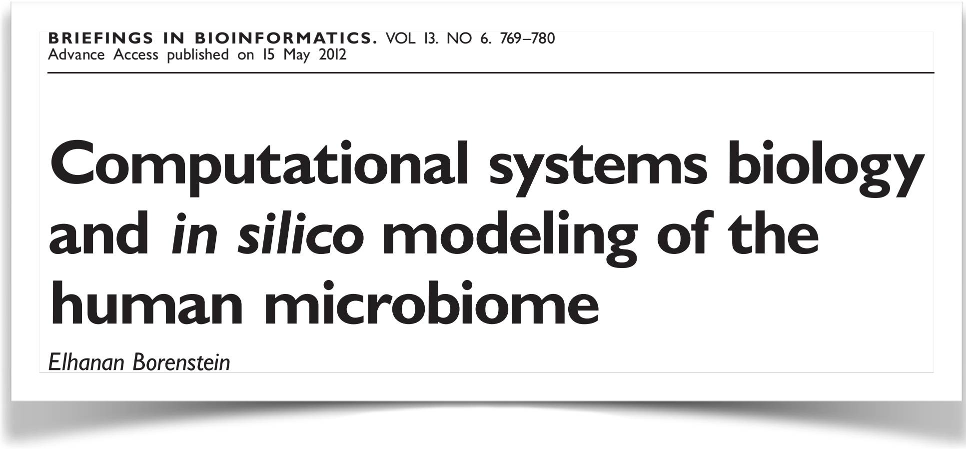
Follows MJ, Dutkiewicz S (2011). Modeling diverse communities of marine microbes. Ann Rev Mar Sci 3: 427–451.

What genomic tells us



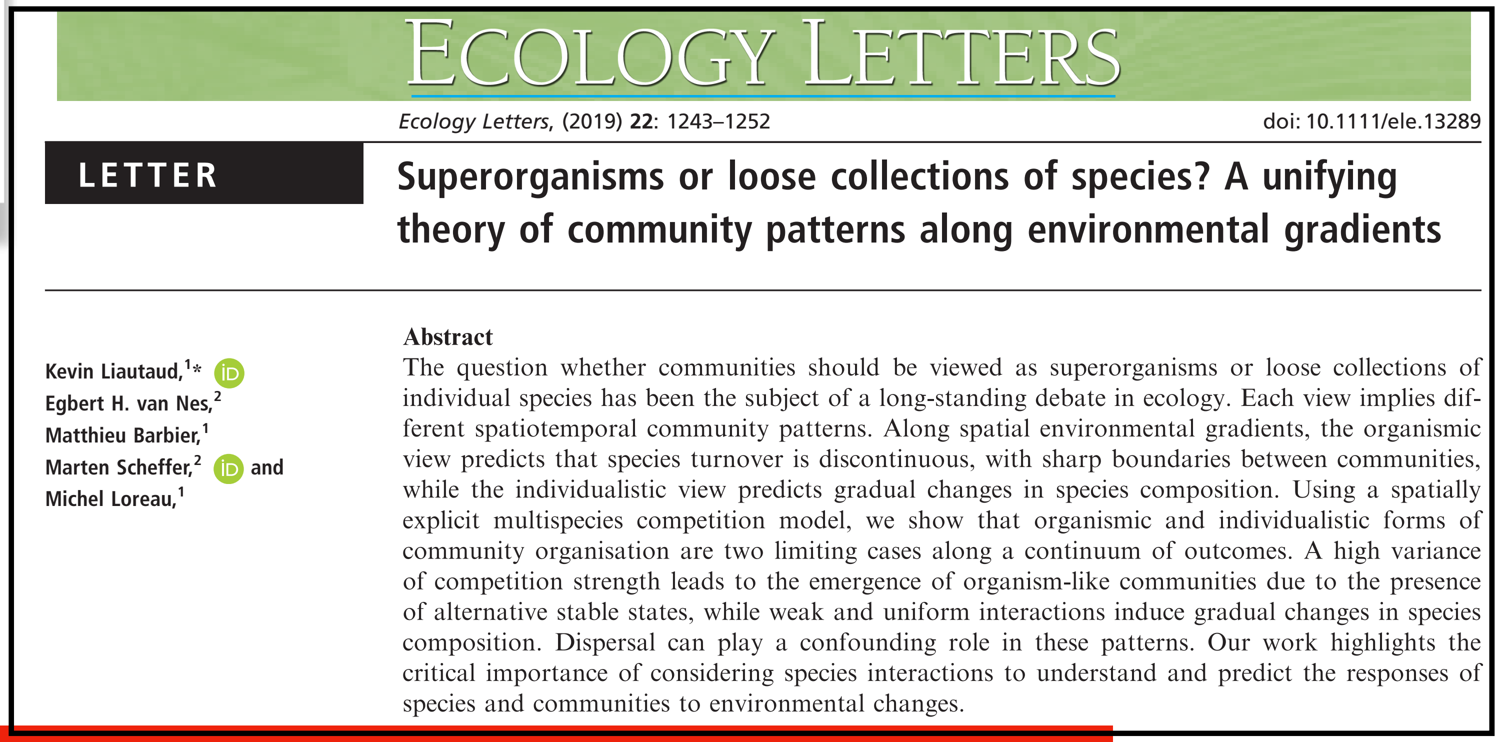
Hypothesis: Metabolic network of the Super-Organism to access the marine biogeochemistry




BRIEFINGS IN BIOINFORMATICS. VOL 13. NO 6. 769–780
Advance Access published on 15 May 2012

Computational systems biology and *in silico* modeling of the human microbiome

Elhanan Borenstein


ECOLOGY LETTERS
Ecology Letters, (2019) **22**: 1243–1252 doi: 10.1111/ele.13289

LETTER

Superorganisms or loose collections of species? A unifying theory of community patterns along environmental gradients

Kevin Liautaud,[1]*
Egbert H. van Nes,[2]
Matthieu Barbier,[1]
Marten Scheffer,[2] and
Michel Loreau,[1]

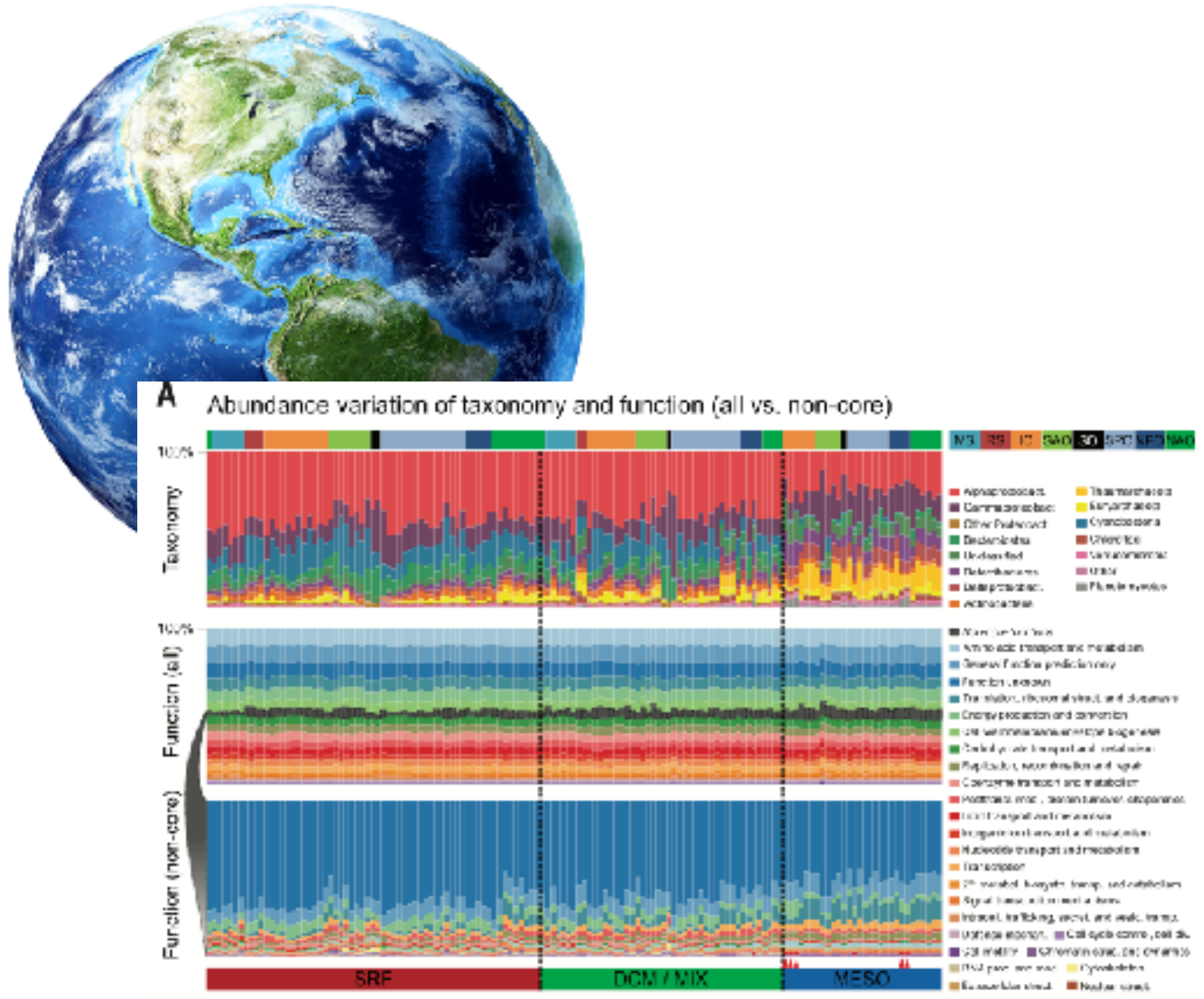
Abstract

The question whether communities should be viewed as superorganisms or loose collections of individual species has been the subject of a long-standing debate in ecology. Each view implies different spatiotemporal community patterns. Along spatial environmental gradients, the organismic view predicts that species turnover is discontinuous, with sharp boundaries between communities, while the individualistic view predicts gradual changes in species composition. Using a spatially explicit multispecies competition model, we show that organismic and individualistic forms of community organisation are two limiting cases along a continuum of outcomes. A high variance of competition strength leads to the emergence of organism-like communities due to the presence of alternative stable states, while weak and uniform interactions induce gradual changes in species composition. Dispersal can play a confounding role in these patterns. Our work highlights the critical importance of considering species interactions to understand and predict the responses of species and communities to environmental changes.

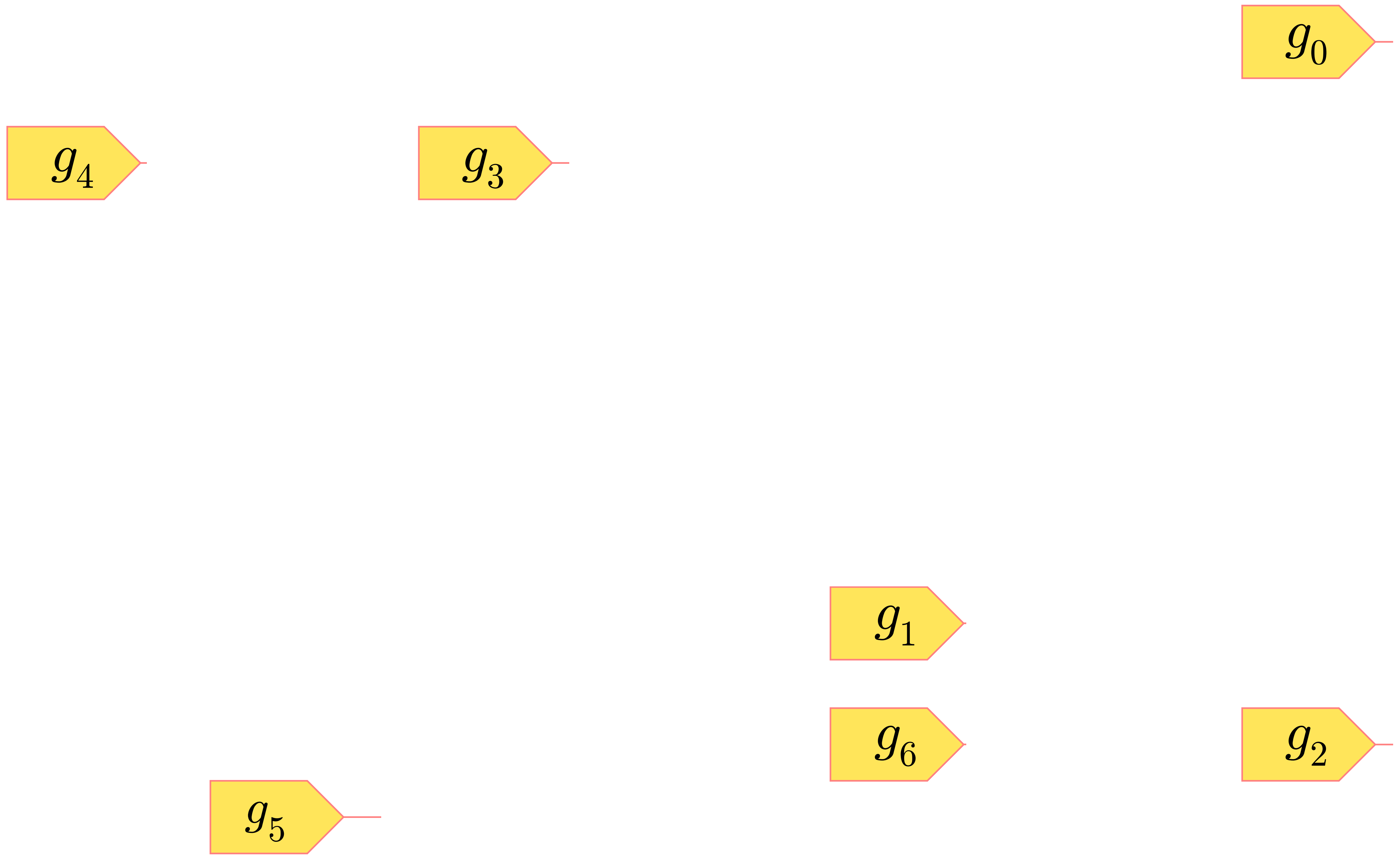
Superorganism Model assumption to assess marine biogeochemical cycles

Building Tara metabolic networks

Integrating Reactions, Enzymes and genes

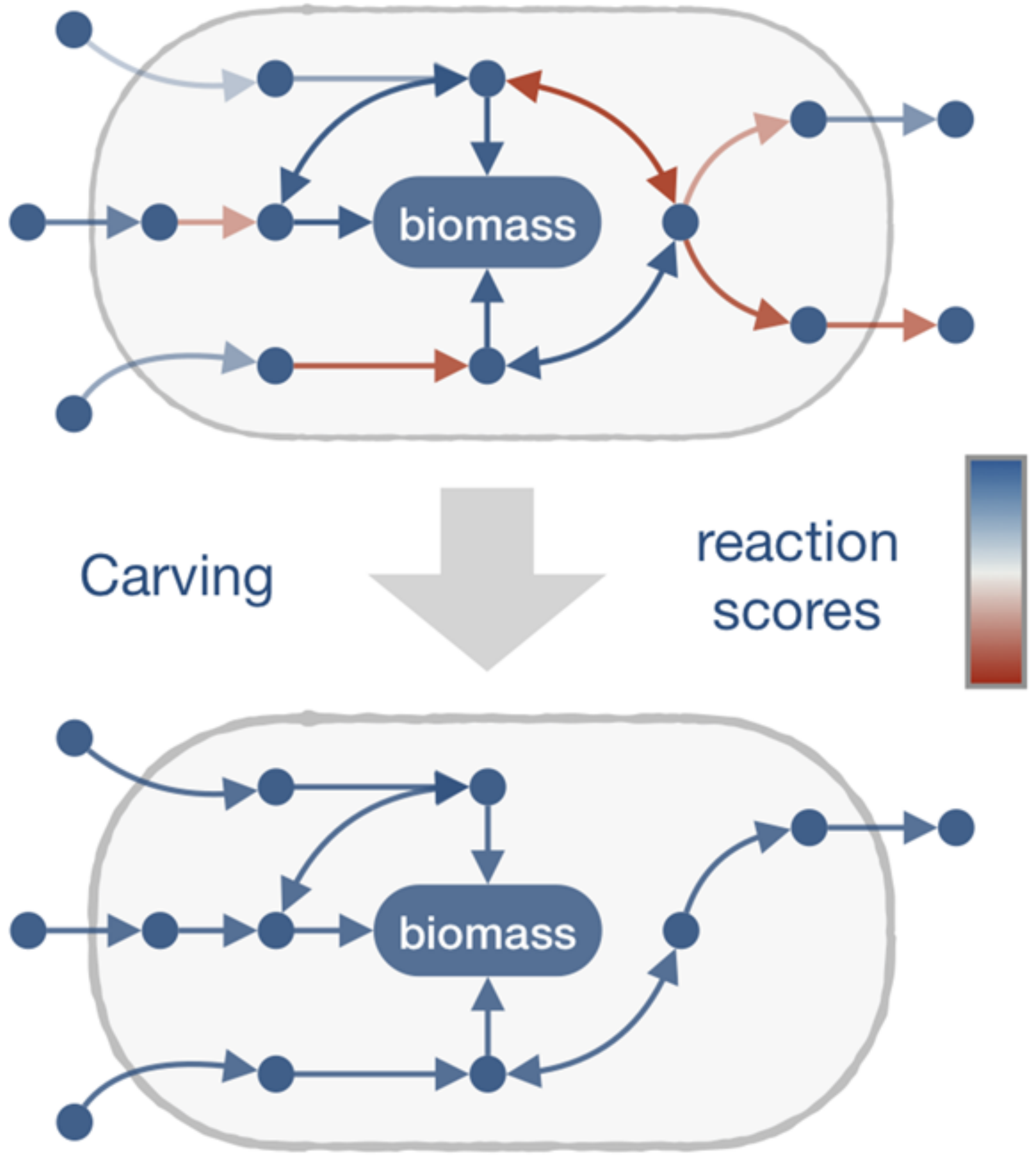


Fondation
tara océan
explorer et partager



Building Metabolic networks

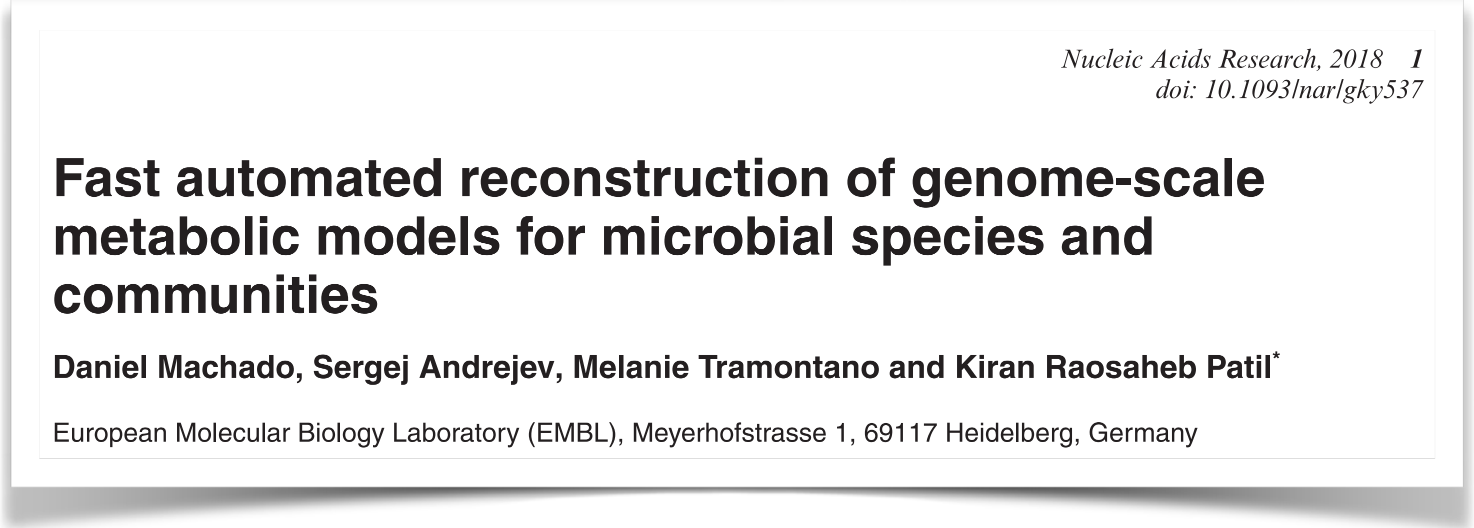
Starting from the Universal knowledge



Functional Metabolic Models

Targeting

- ✓ **Marine Autotrophic biomass objectives**
- ✓ **Marine Heterotrophic biomass objectives**
- ✓ **Marine Mixotrophic biomass objectives**

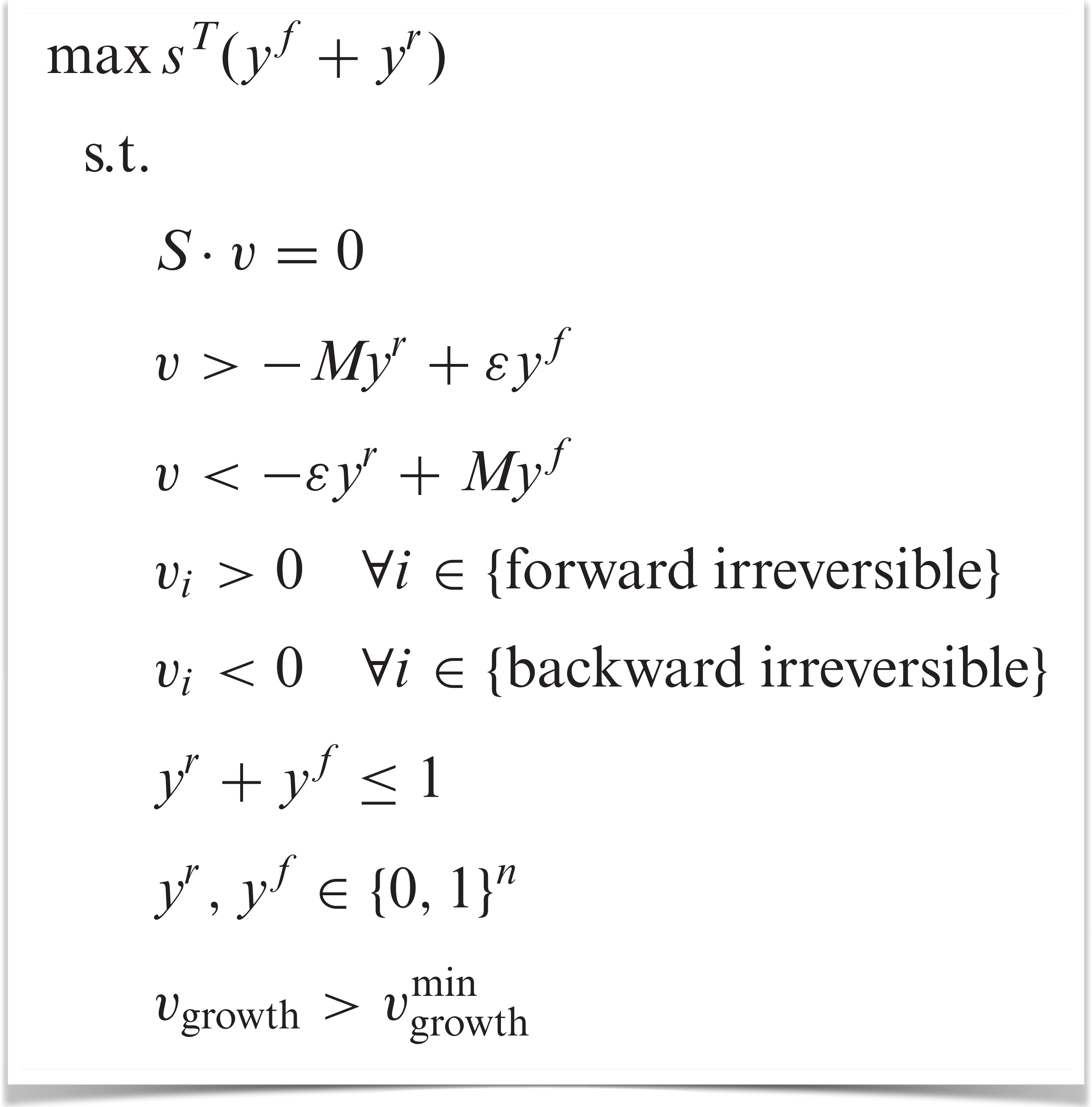


Nucleic Acids Research, 2018 1
doi: 10.1093/nar/gky537

Fast automated reconstruction of genome-scale metabolic models for microbial species and communities

Daniel Machado, Sergej Andrejev, Melanie Tramontano and Kiran Raosaheb Patil*

European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany

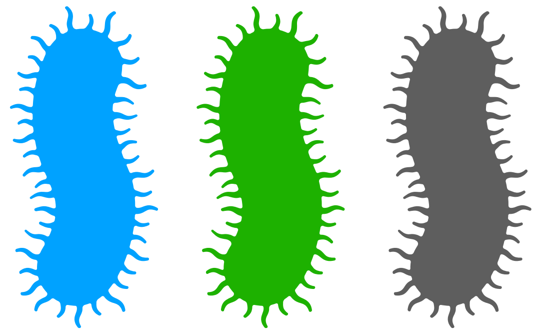


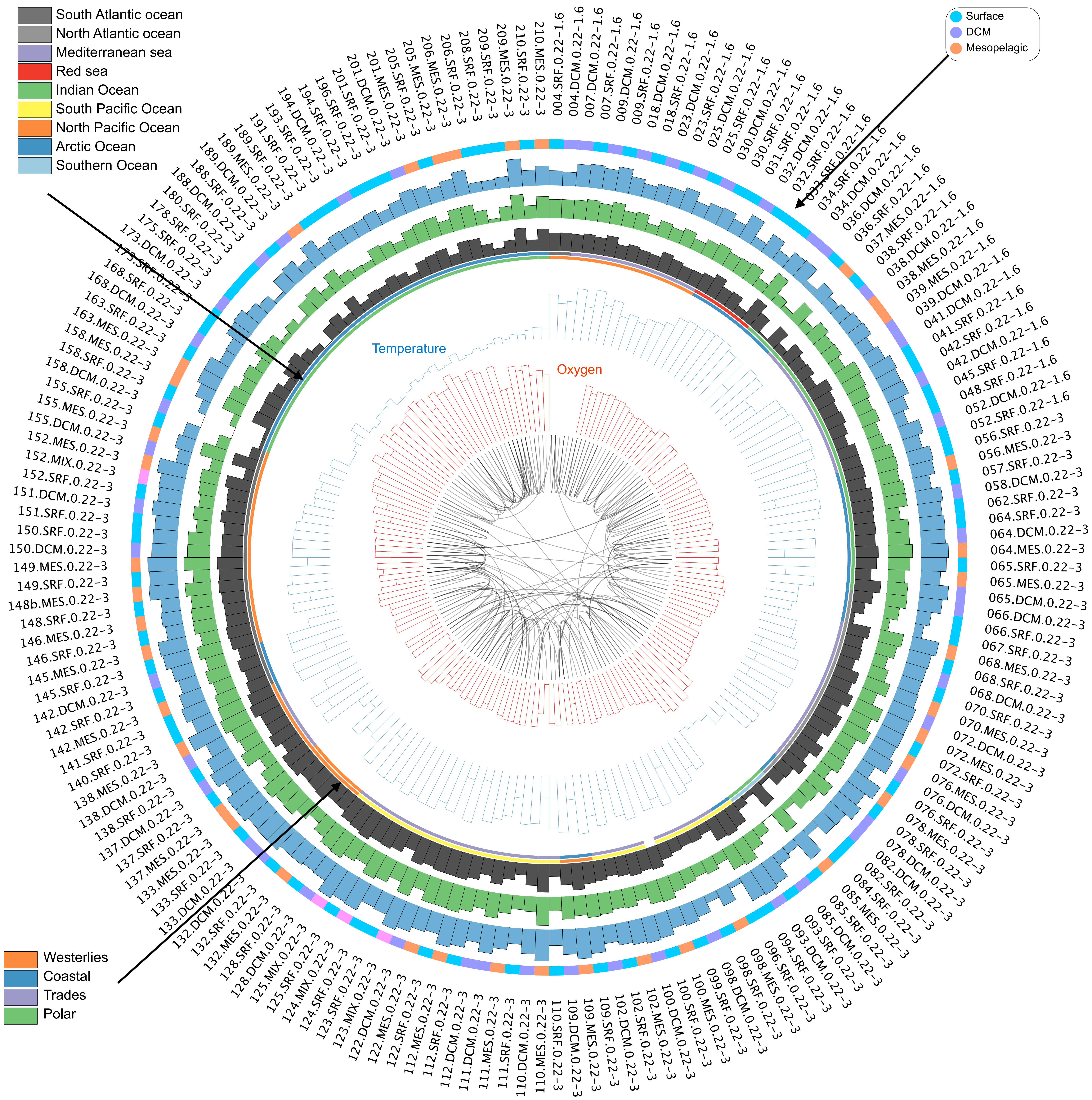
$$
\begin{aligned}
&\max s^T(y^f + y^r)\\
&\text{s.t.}\\
&\quad S \cdot v = 0\\
&\quad v > -My^r + \varepsilon y^f\\
&\quad v < -\varepsilon y^r + My^f\\
&\quad v_i > 0 \quad \forall i \in \{\text{forward irreversible}\}\\
&\quad v_i < 0 \quad \forall i \in \{\text{backward irreversible}\}\\
&\quad y^r + y^f \leq 1\\
&\quad y^r, y^f \in \{0, 1\}^n\\
&\quad v_{\text{growth}} > v_{\text{growth}}^{\text{min}}
\end{aligned}
$$

Considering:

- stoichiometry of the reactions,
- mass action laws,
- reactions entropy
- optimize the growth rate

Prokaryotic metabolic world

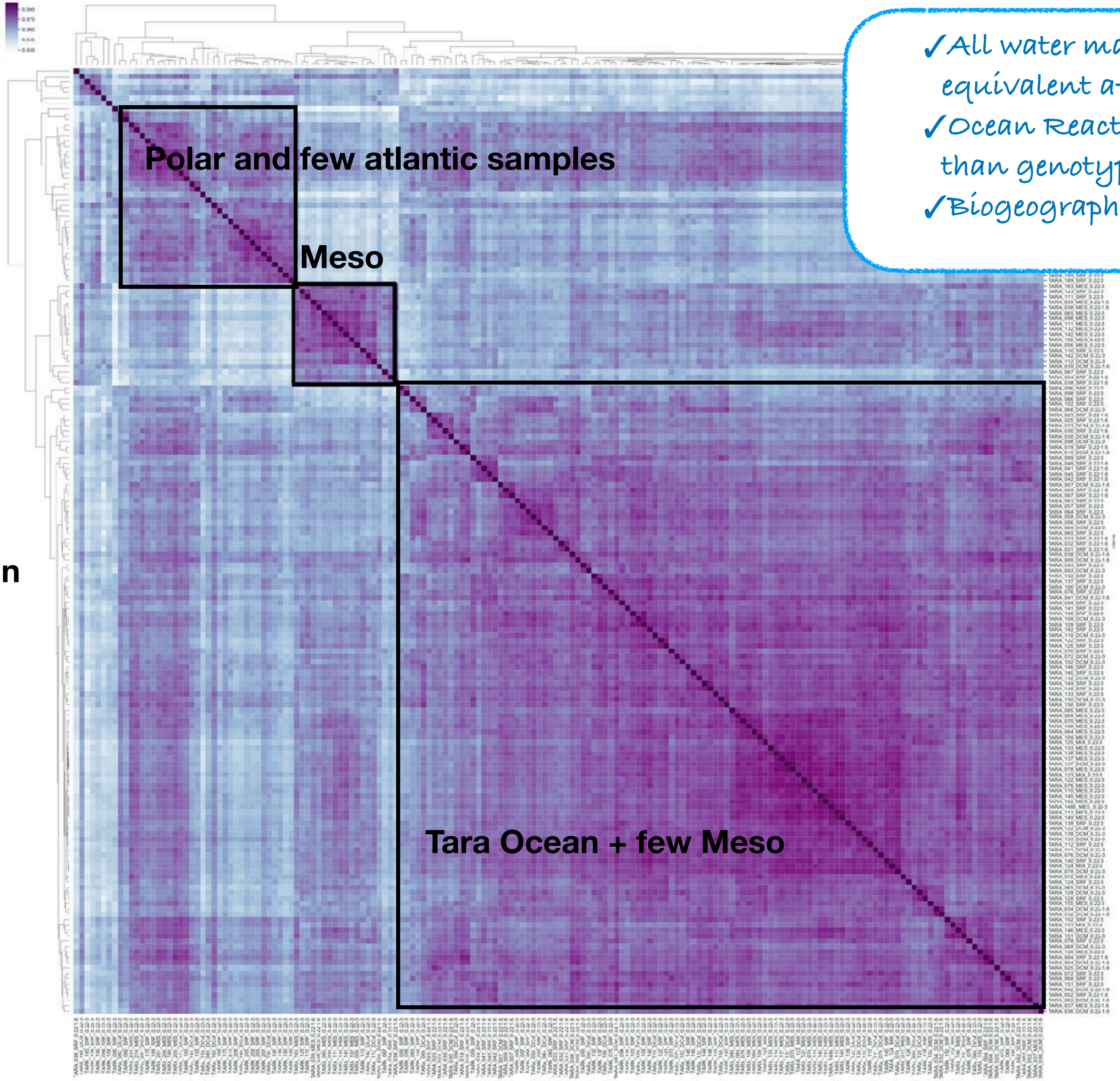




- **Heterotrophic**,
- **Autotrophic**,
- **Mixotrophic**

Metabolic networks (~3600 metabolic reactions) show differences but overall consistencies

Metabolic Reaction Jaccard Index



Polar and few atlantic samples

Meso

Tara Ocean + few Meso

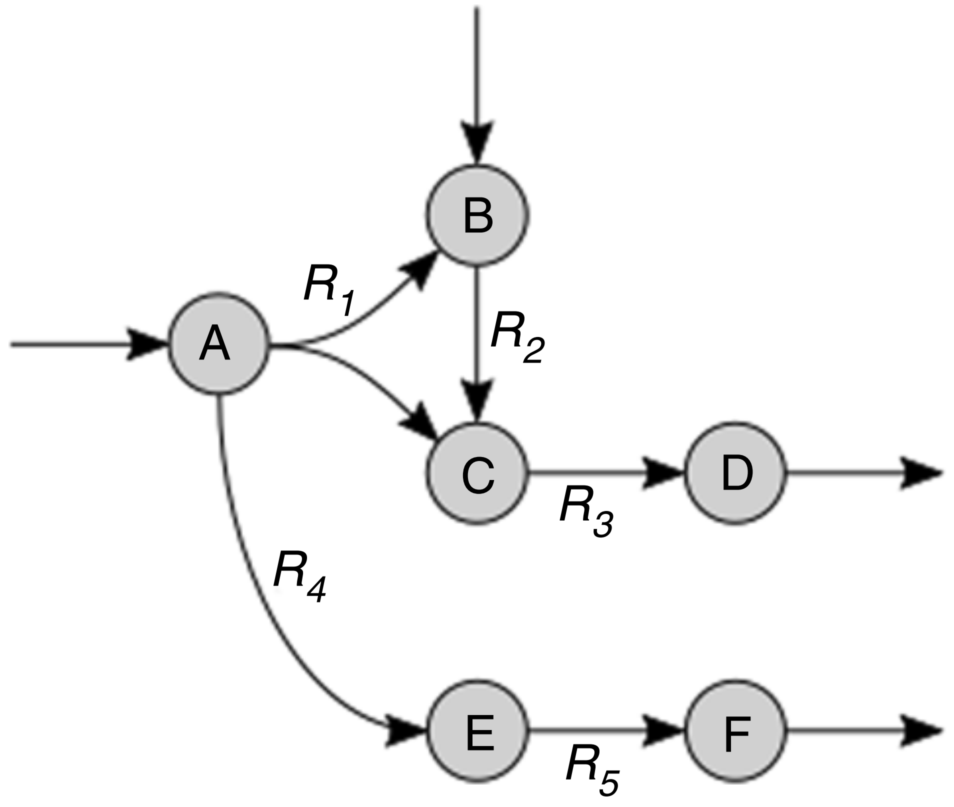
- ✓All water masses are not geochemically equivalent after metabolic pipeline
- ✓Ocean Reactome shows similar patterns than genotypes
- ✓Biogeography from reactome

Ready for analyzing the Ocean phenotypes
Need for a dedicated modeling

Analyzing each water mass metabolic behavior

(from network to model)

Metabolic network



$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -3 & -2 & \cdots & 0 & 0 & 0 \\ -1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 3 & -1 & \cdots & -1 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & -1 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ \vdots \\ \vdots \\ v_{n-2} \\ v_{n-1} \\ v_n \end{pmatrix} =$$

$$\mathbf{Sv} = 0$$

$$lb \leq v_i$$

$$v_i \leq ub$$

Stoichiometric Matrix

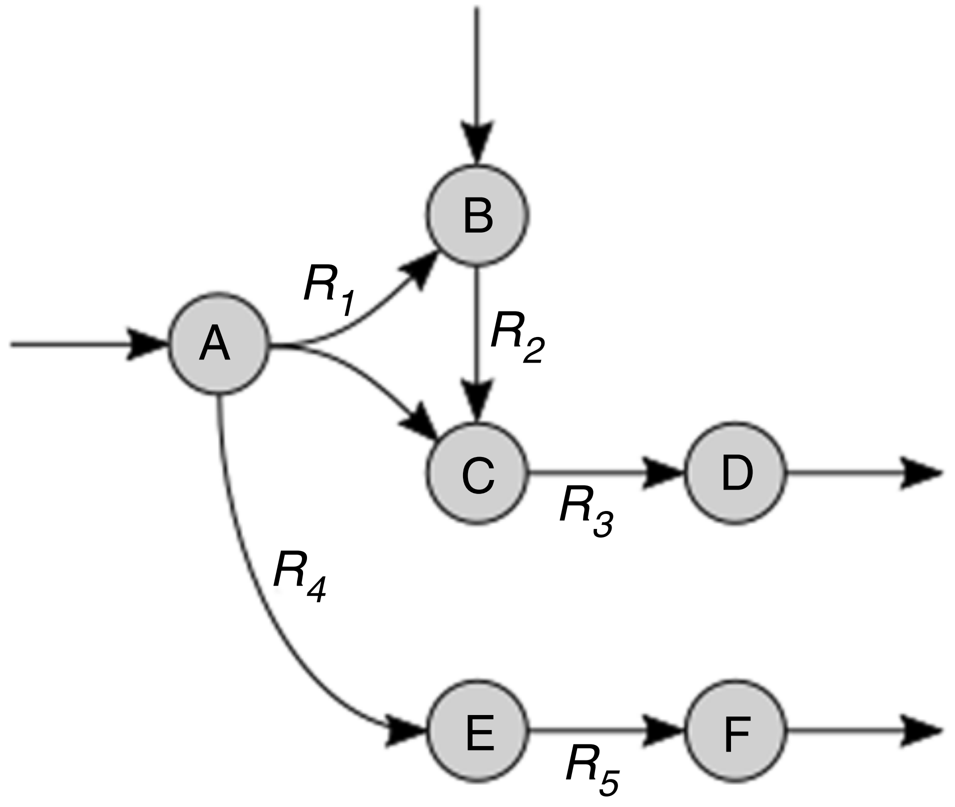
Flux ***v***ector*v(t)*

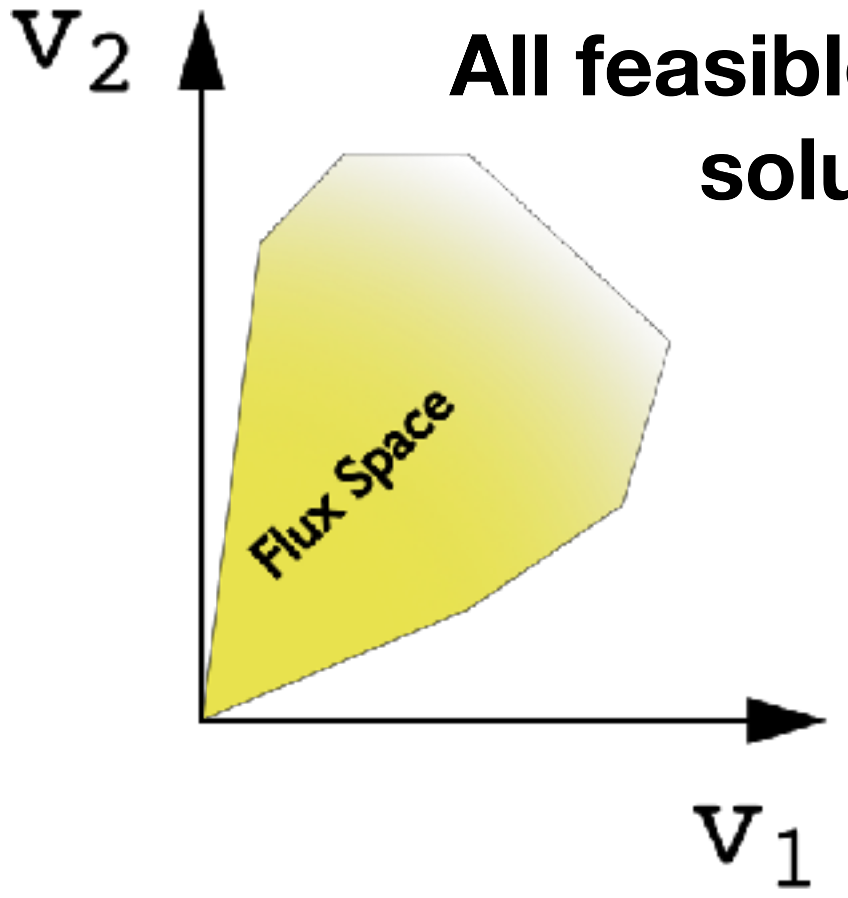
Flux Space

Analyzing each water mass metabolic behavior

(from network to model… and back to graph!)

Metabolic network



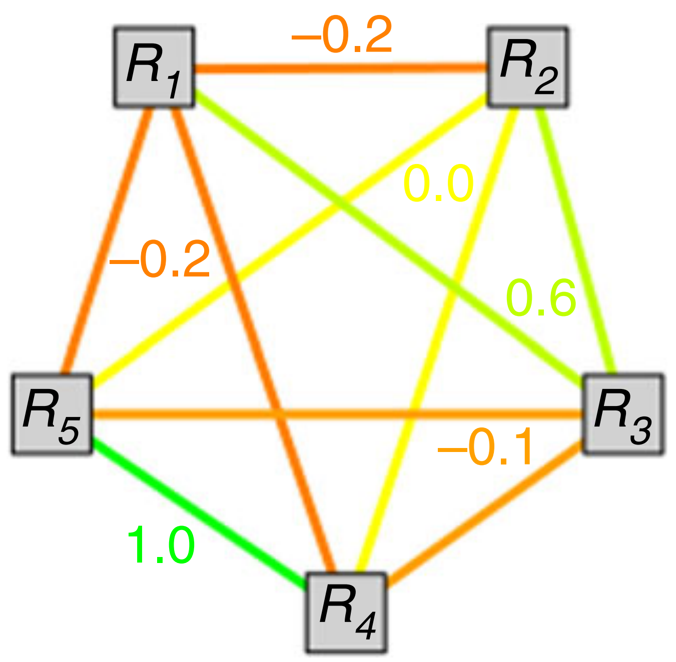


All feasible metabolic solutions

Exploration of the Flux Space

(i.e., subsampling procedure)

Flux correlations

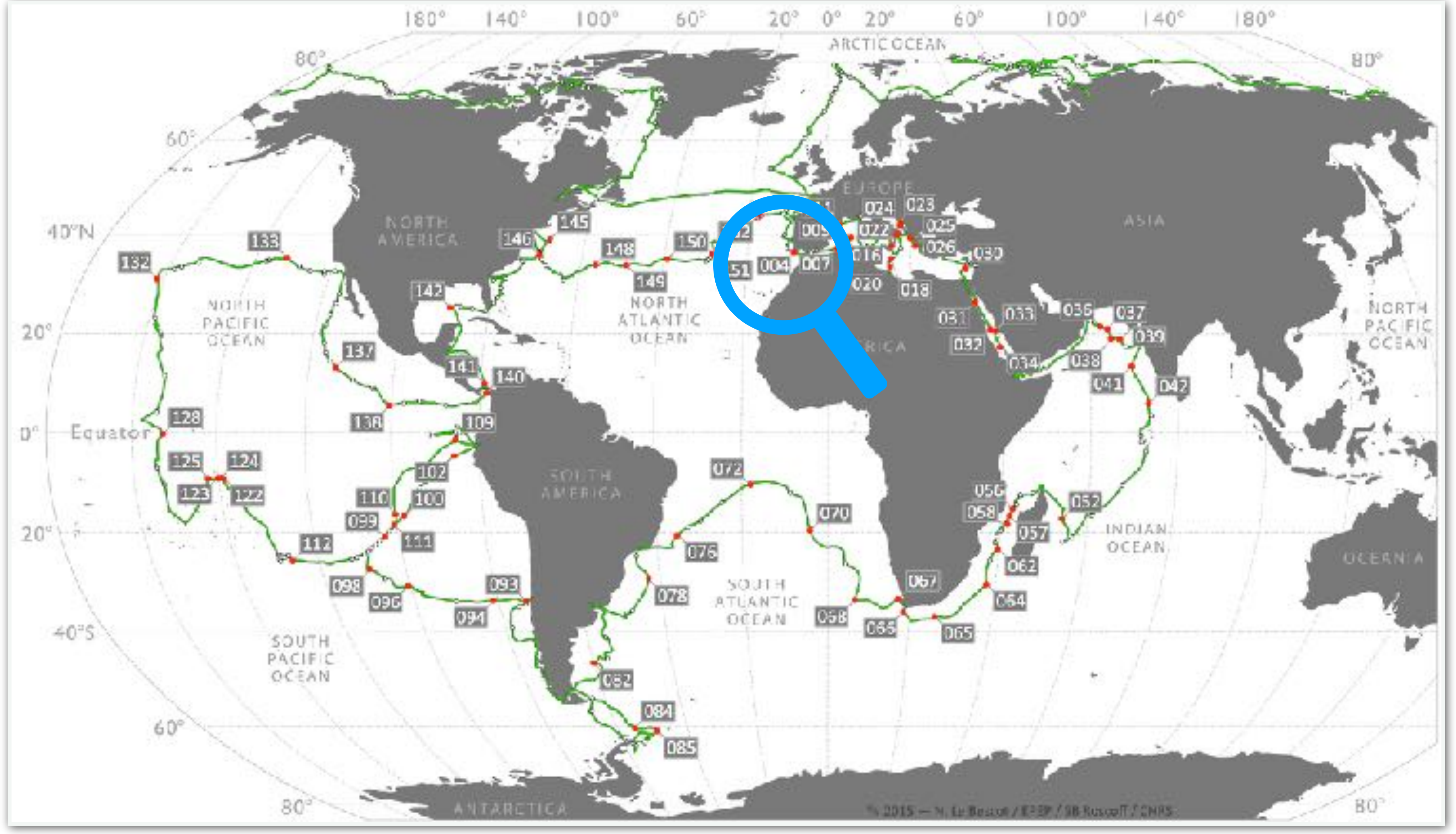


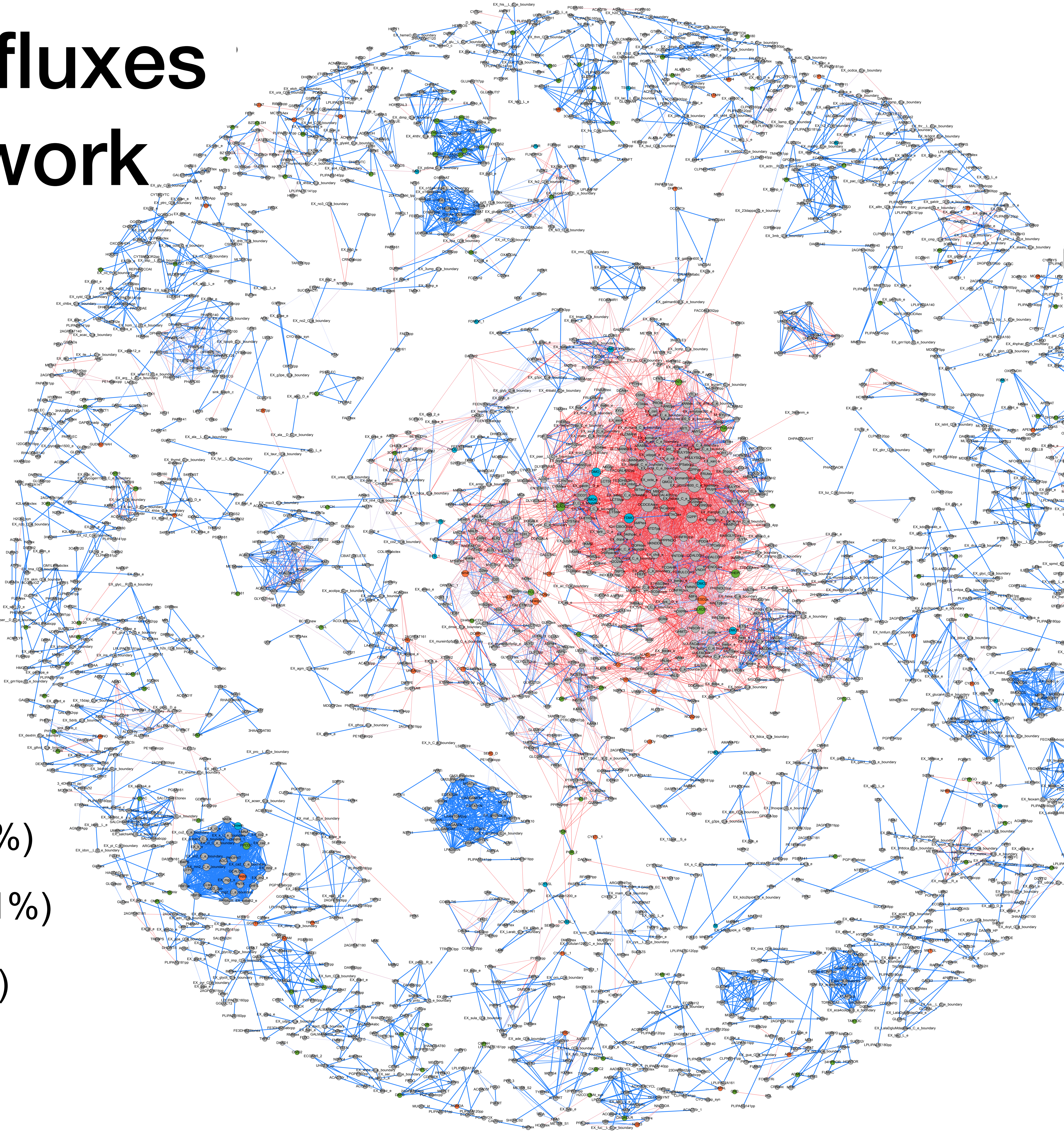
Correlation between fluxes in reactions

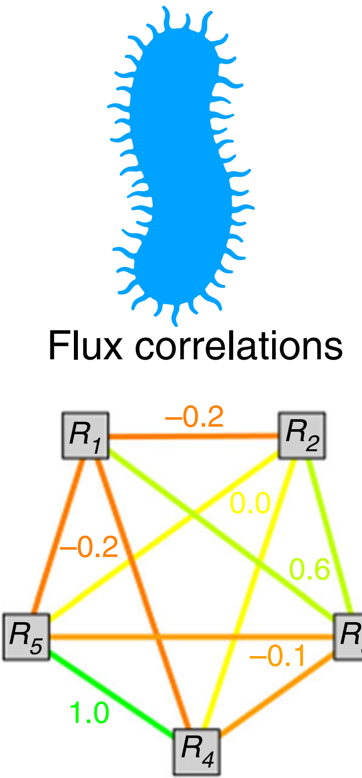
~3 600 reactions per Tara Ocean samples (1,5 times a Human organ)

~ 6,8 millions of flux correlations per Tara Océan samples

Dependencies between fluxes in given metabolic network







correlation value

1

-0.9 // 0.9

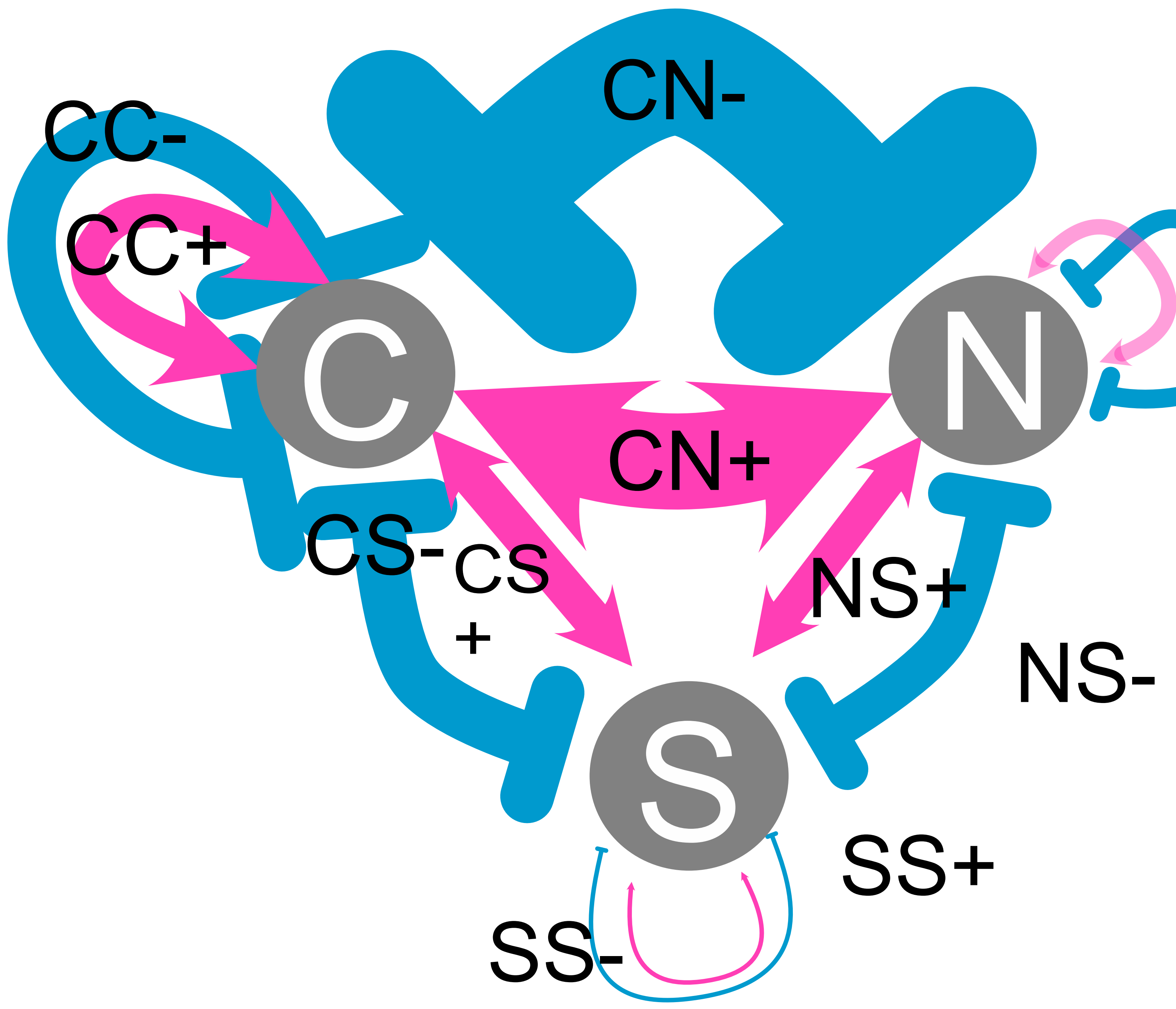
-1

metabolism

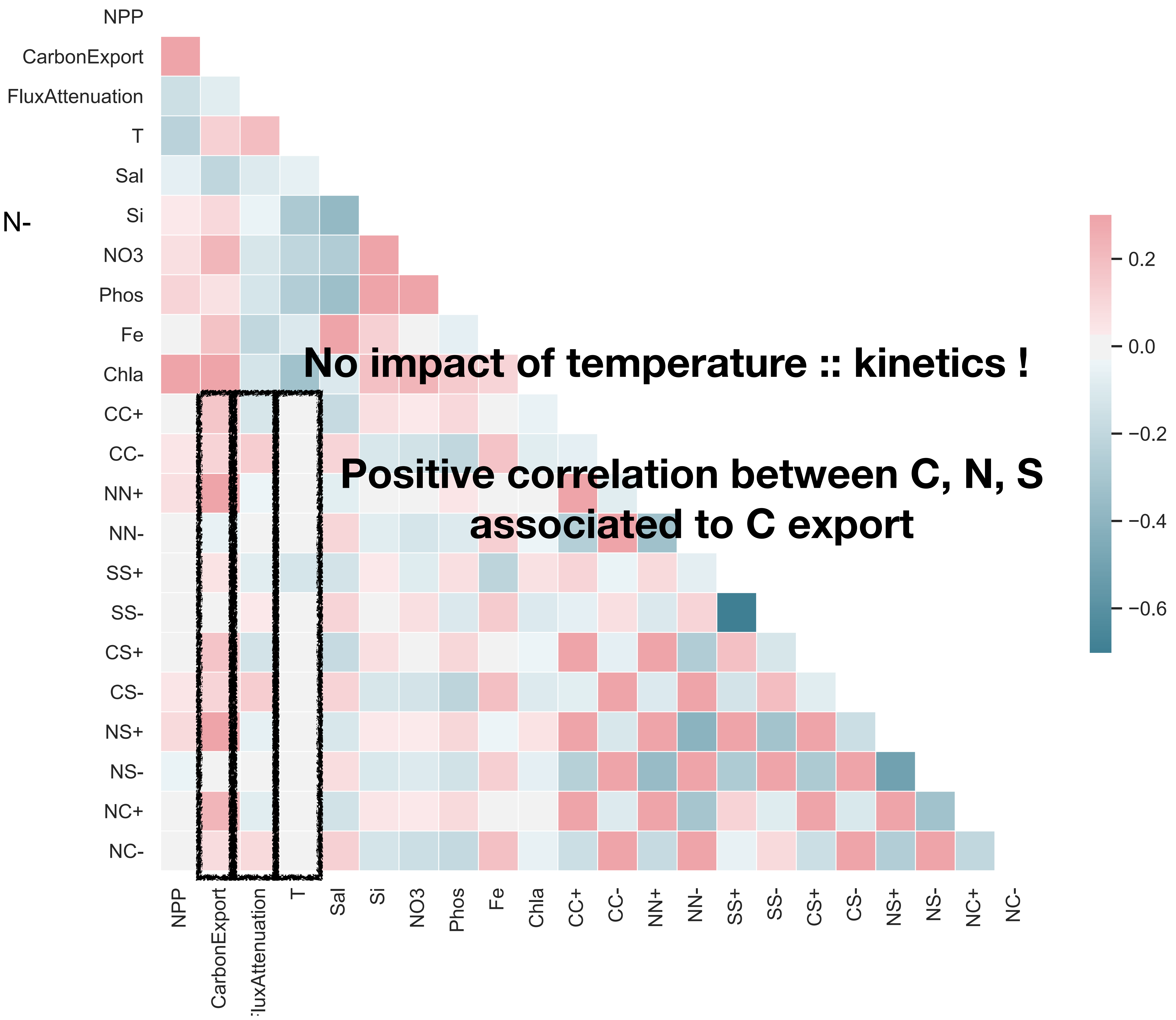
- 🟢 Carbon metabolism (5.05%)
- 🟠 Nitrogen metabolism (2.91%)
- 🔵 Sulfur metabolism (1.08%)
- ⚪ Everything else (90.96%)


correlation value
1
-0.9
0.9
-1
metabolism
Carbon metabolism (5.05%)
Nitrogen metabolism (2.91%)
Sulfur metabolism (1.08%)
Everything else (90.96%)

Prok. Biogeochemistry in the global ocean (Carbon, Nitrogen, Sulfate)



Highest NPP water mass



No impact of temperature :: kinetics !

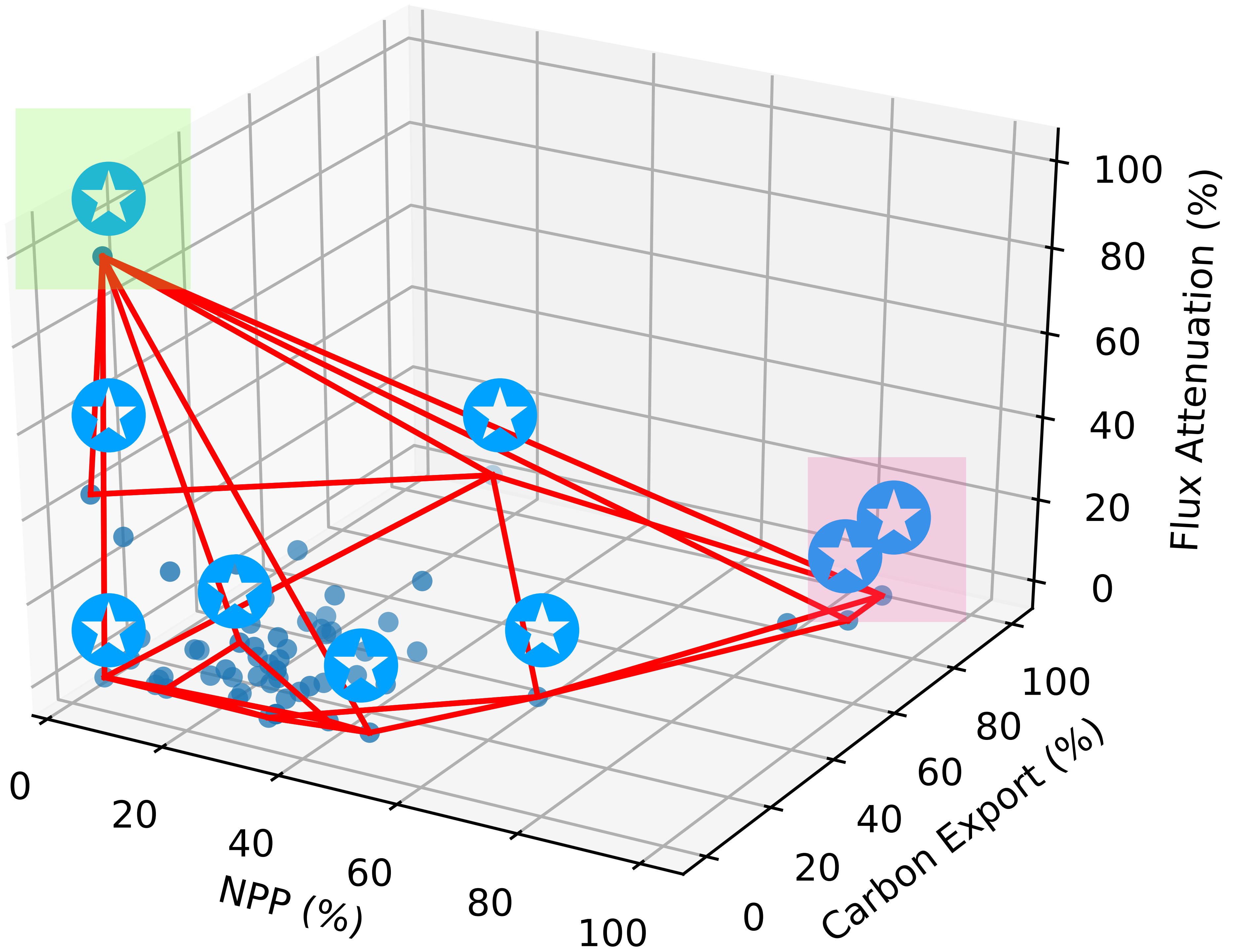
Positive correlation between C, N, S associated to C export

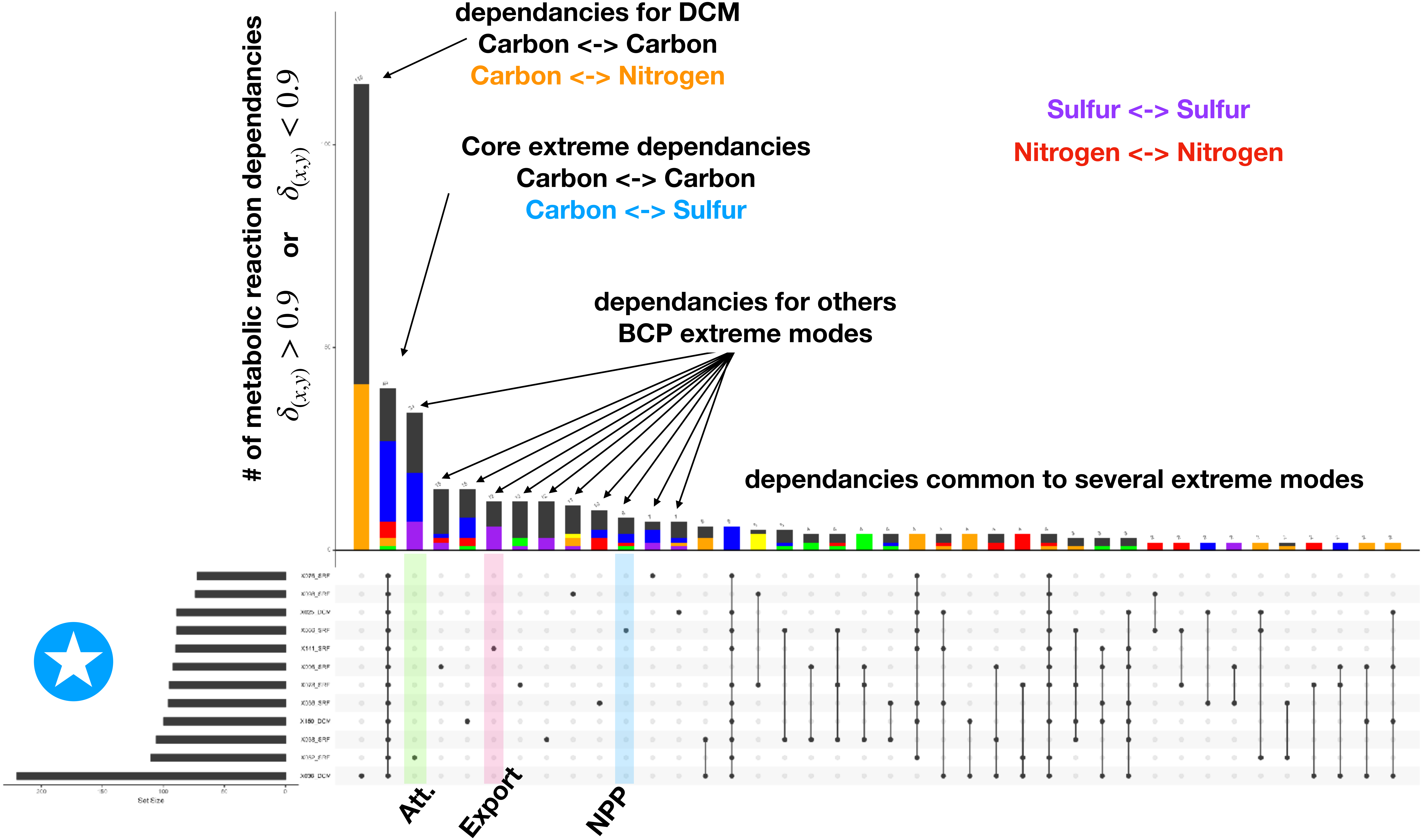
« Simplexification » of the Biological Carbon Pump

BCP is described by

- carbon export
- net primary production (NPP)
- remineralization

The Global Ocean is enclosed in a convex hull that emphasizes extreme stations ✪





dependancies for DCM
Carbon <-> Carbon
Carbon <-> Nitrogen

Core extreme dependancies
Carbon <-> Carbon
Carbon <-> Sulfur

dependancies for others
BCP extreme modes

dependancies common to several extreme modes

Sulfur <-> Sulfur

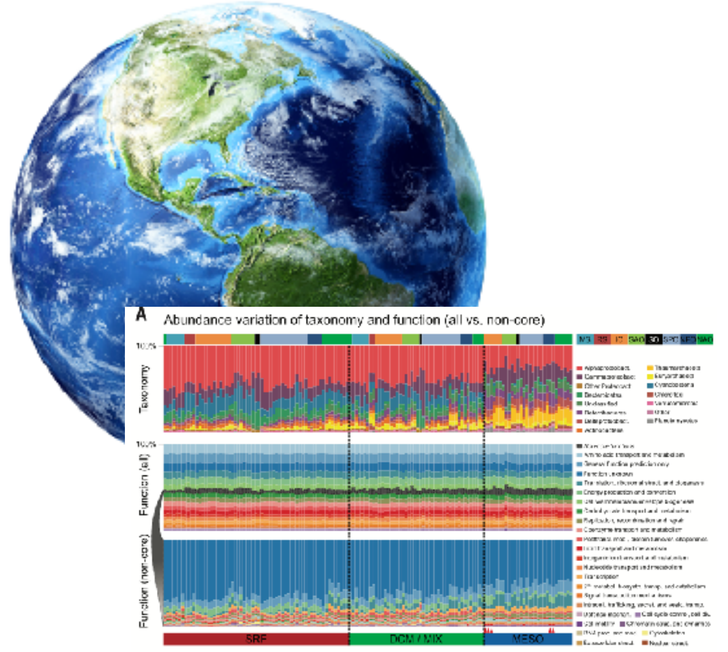
Nitrogen <-> Nitrogen

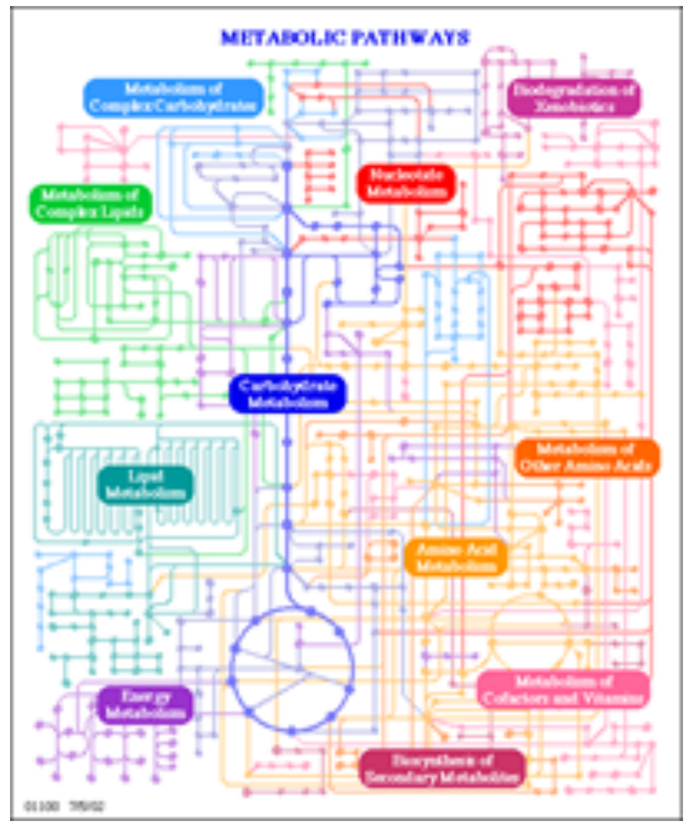
Aims of Constraints-based modeling applied on Ocean omics data

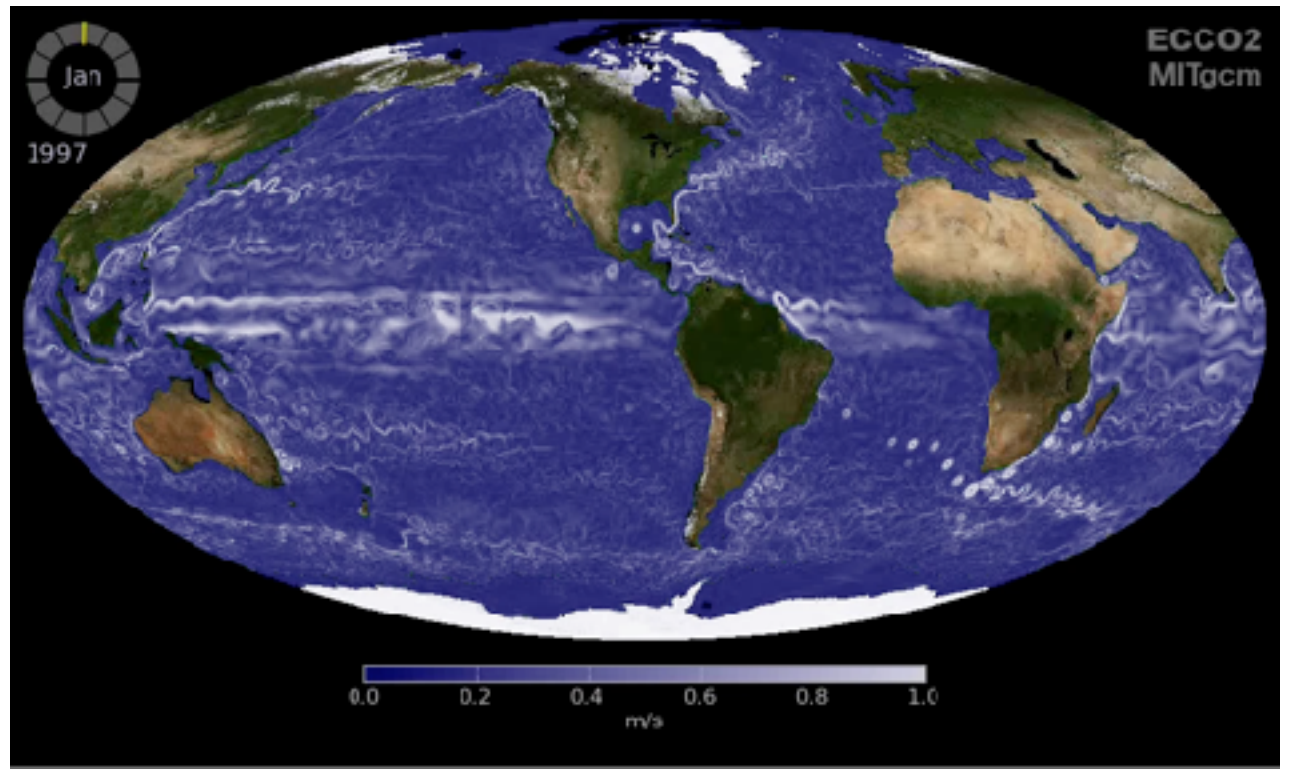
- From big data to a proper abstraction
- Inference of biogeochemistry of the ocean from omics (revisiting the microbial loop)
- Mechanistic **and** meta-genome-scale modeling

<u>Perspectives</u>:

- Proper mining of the key metabolic reactions
- Identify organisms responsible for biogeochemical keystone reactions (rare or abundant? role of diversity in biogeochemistry)
- Adding mesoscale data & **time!!!!!!**







Thank you

- Samuel Chaffron, Julie Haguait, Abdelhalim Larhlimi, Antoine Regimbeau, Camille Trottier
- Marko Budinich
- Philippe Bordron
- Lionel Guidi
- Matt Sullivan, James Wainaina
- Alejandro Maass, Dante Travisany
- Member of the Tara Ocean consortium

UNIVERSITÉ DE NANTES

LS2N LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES

CNRS • SORBONNE UNIVERSITÉ
Station Biologique de Roscoff

Inserm

IMEV INSTITUT DE LA MER DE VILLEFRANCHE

THE OHIO STATE UNIVERSITY

CMM
Center for Mathematical Modeling

Fondation tara océan
explorer et partager