

Co-activity networks reveal the structure of planktonic symbiosis in the global ocean

Nils Giordano, Samuel Chaffron

Computational Biology team (COMBI)
Laboratoire des Sciences du Numérique de Nantes (LS2N, UMR 6004)

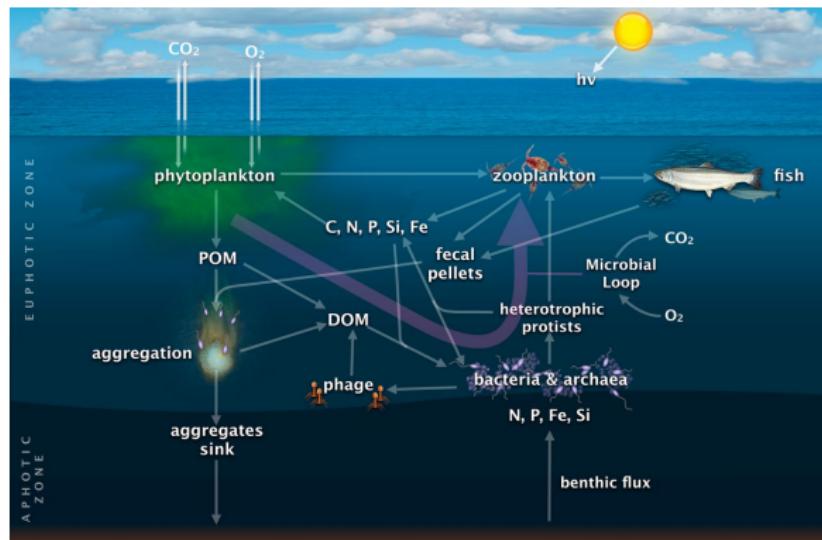
RCAM'19
September, 30th 2019 (Paris)



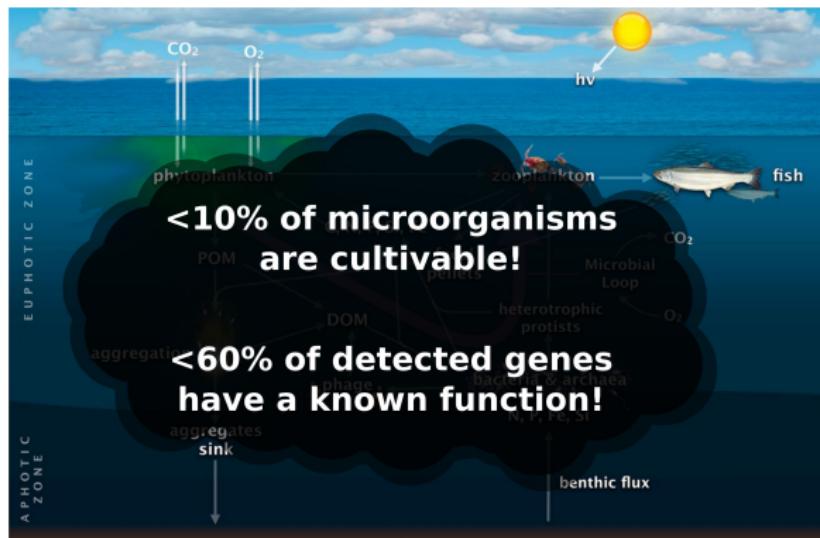
LABORATOIRE
DES SCIENCES
DU NUMÉRIQUE
DE NANTES



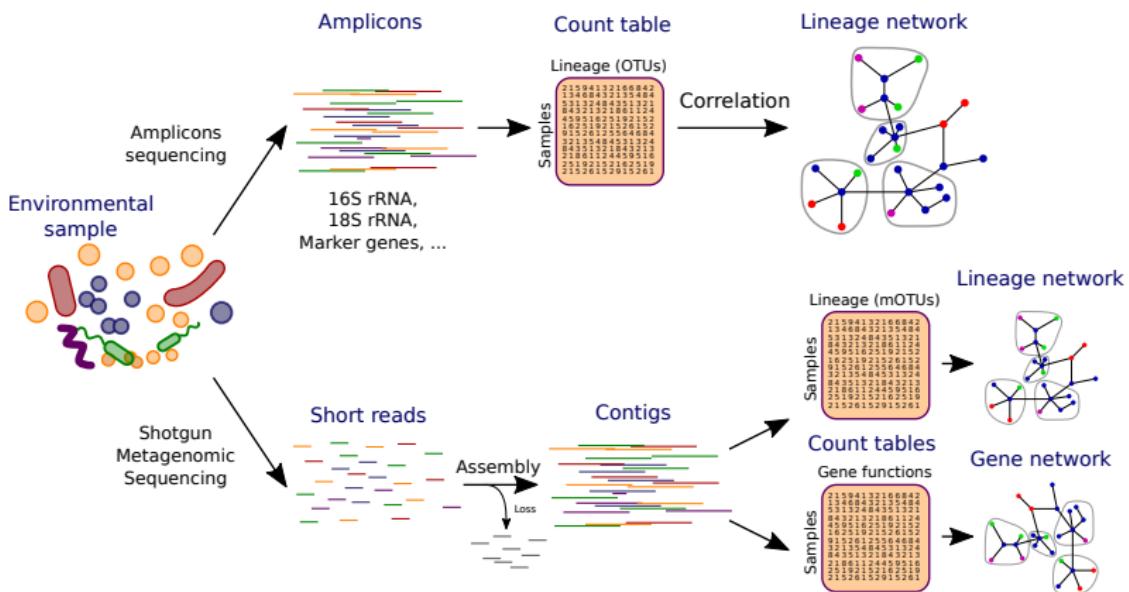
MARINE MICROBIAL COMMUNITIES PLAY CRUCIAL ECOLOGICAL ROLES



MARINE MICROBIAL COMMUNITIES PLAY CRUCIAL ECOLOGICAL ROLES

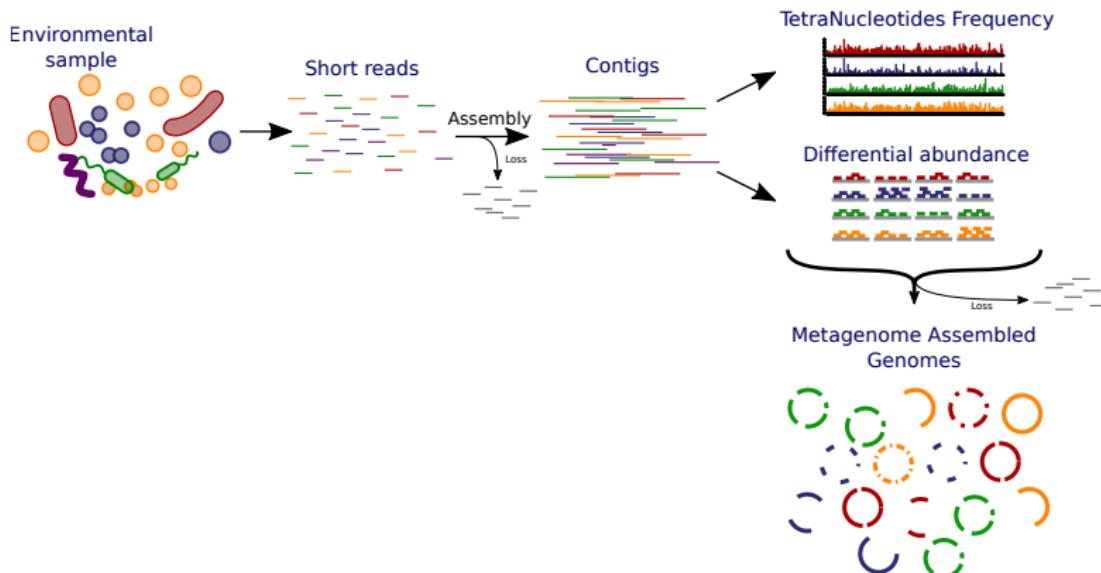


CULTURE-INDEPENDENT ENVIRONMENTAL GENOMICS

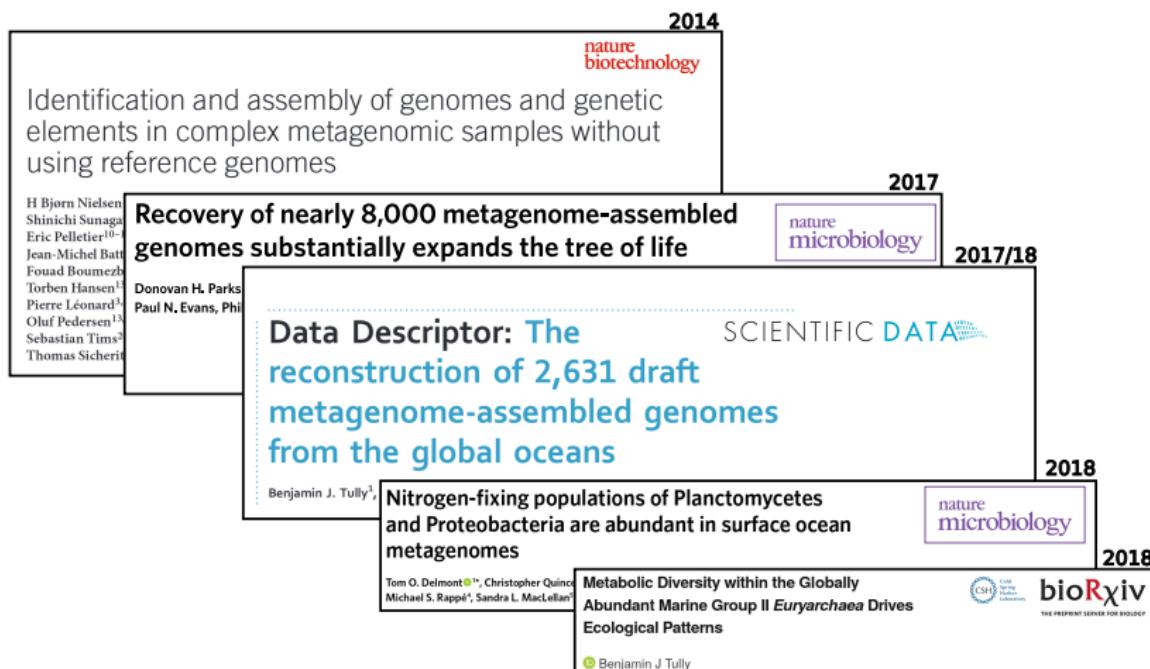


No link between taxonomy and functions?

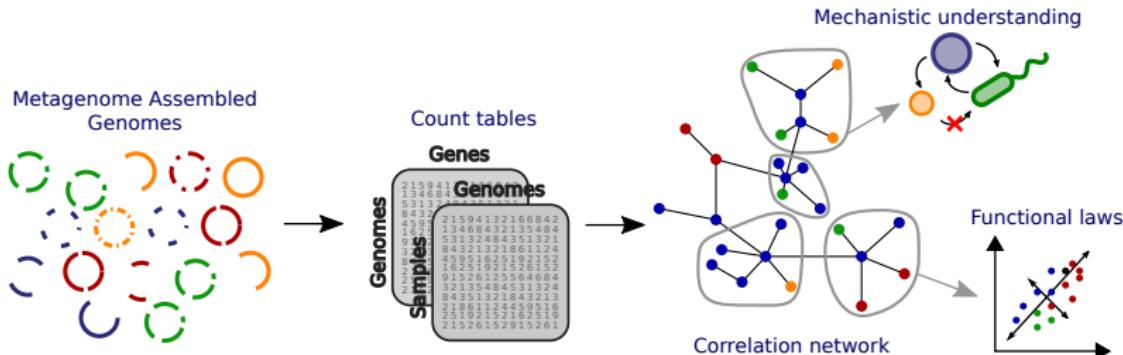
HIGH QUALITY GENOMES CAN BE ASSEMBLED FROM METAGENOMES



LITERATURE IS FILLING WITH NEW METAGENOME ASSEMBLED GENOMES (MAGs)



We want to predict, characterize, and **explain** the community structure of non-cultivable marine microorganisms.



- How to make an association network of MAGs?
- How to use this network to explain the associations and predict interactions?

DATASETS

CULTIVATED AND NON-CULTIVATED GENOMES

| Source | Nb. | After derepl.* | Samples | Techniques |
|---------------------|-------|----------------|-----------------------------------|--------------------|
| MarRef | 943 | 602 | | |
| MarDB | 12963 | 4445 | | |
| Aquatic Progenomes | 566 | 265 | | |
| Parks et al 2017 | 1765 | 576 | Tara (all size fractions), others | SA+Metabat (no DC) |
| Tully et al 2017/18 | 2597 | 1298 | Tara (all depth/size fractions) | SA+Merging+Metabat |
| Delmont et al 2018 | 957 | 472 | Tara (SUR+DCM, prok. only) | CA+CONCOCT/Anvio |
| Total | 19791 | 7658 | | |

*Dereplication: 95% ANI over $\geq 60\%$ of the genome (tool: dRep)

METAGENOMES

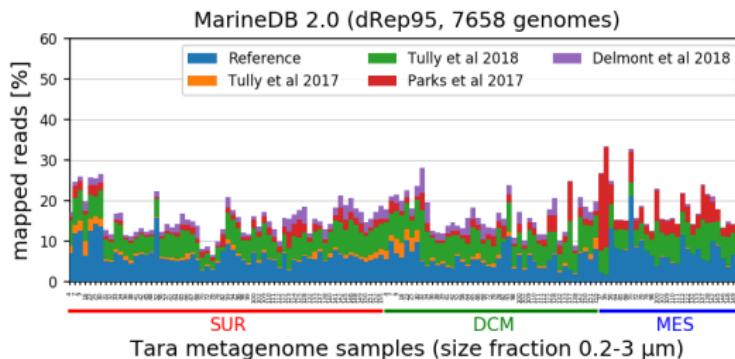


Tara expeditions dataset (2009-2013)

- 3 depth (SUR, DCM, MES)
- Size-filtered samples (0.2-3 μm)
- MetaDNA (N=171) and RNA* (N=179)

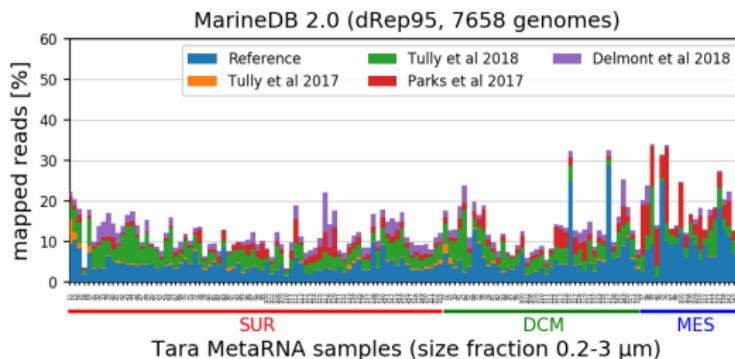
*ribo-depleted MetaRNA (ribosomal RNA chemically removed)

COVERAGE ACROSS TARA SAMPLES



MetaDNA

Total: 16.6%
SUR: 15.8%
DCM: 16.5%
MES: 18.5%

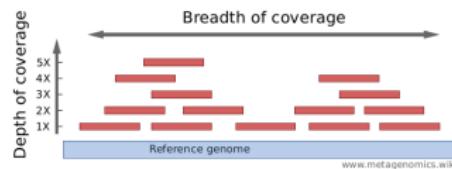
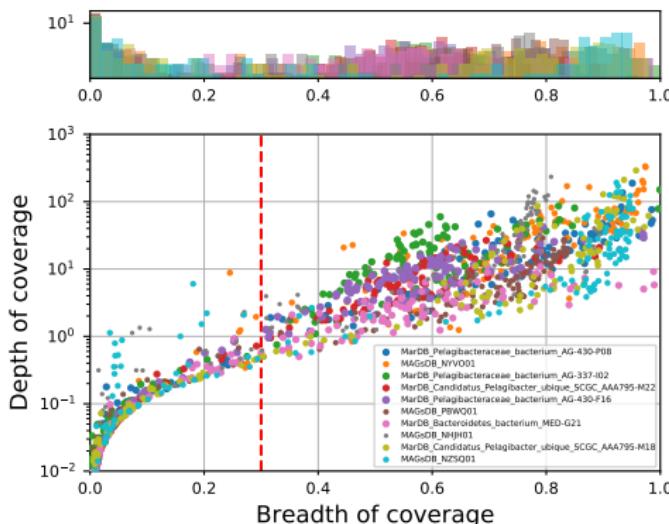


MetaRNA

Total: 13.5%
SUR: 11.8%
DCM: 13.5%
MES: 19.7%

CO-ABUNDANCE NETWORK

HOW TO DEFINE PRESENCE/ABSENCE OF A GENOME?



- 30% threshold on coverage breadth (horizontal coverage)

CO-ABUNDANCE NETWORK

NORMALIZATION

- Shotgun Metagenomic Data are **compositional**
 - For a sample x with D features (= read count on genome i):

$$x = [x_1, \dots, x_D] ; \sum x = c$$

- Centered Log-Ratio (CLR)

$$clr(x) = \left[\log\left(\frac{x_1}{G(x)}\right), \dots, \log\left(\frac{x_D}{G(x)}\right) \right]$$

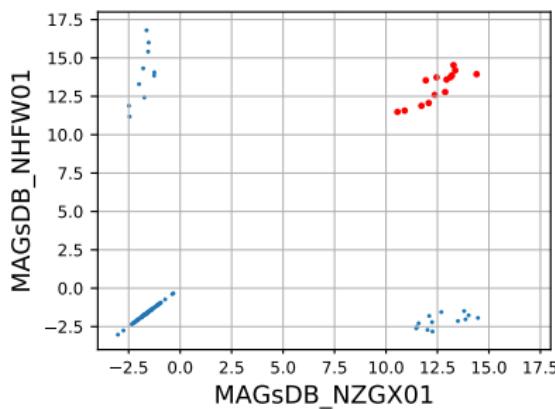
with $G(x) = \sqrt[D]{x_1 \cdot \dots \cdot x_D}$

- Problem with zeros → pseudocount

CO-ABUNDANCE NETWORK

CORRELATION

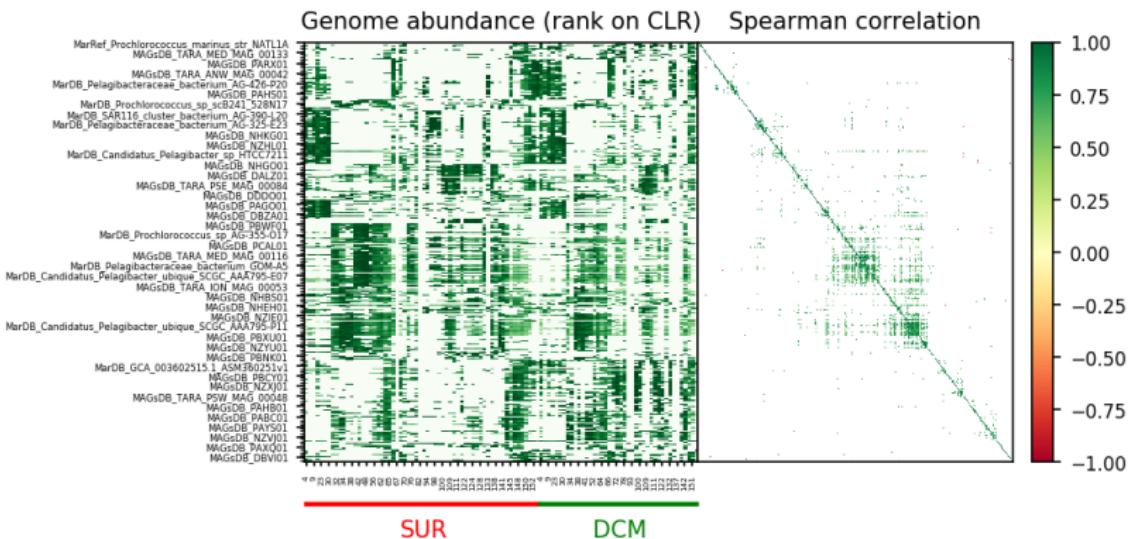
- Spearman correlation on 'co-occurring' datapoints only
 - Constraint: co-occurrence in at least 10% of samples (N=11)



- Benjamini-Hochberg FDR correction ($FDR \leq 0.01$)

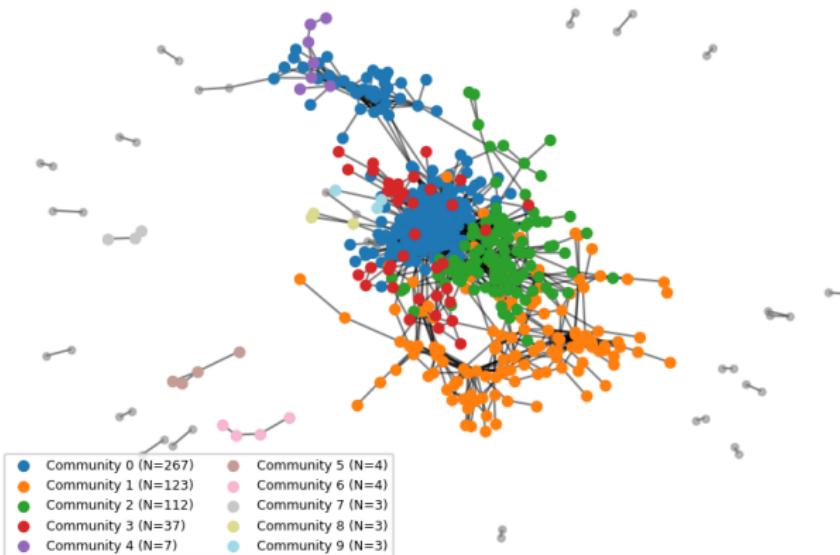
CO-ABUNDANCE NETWORK

RESULTS



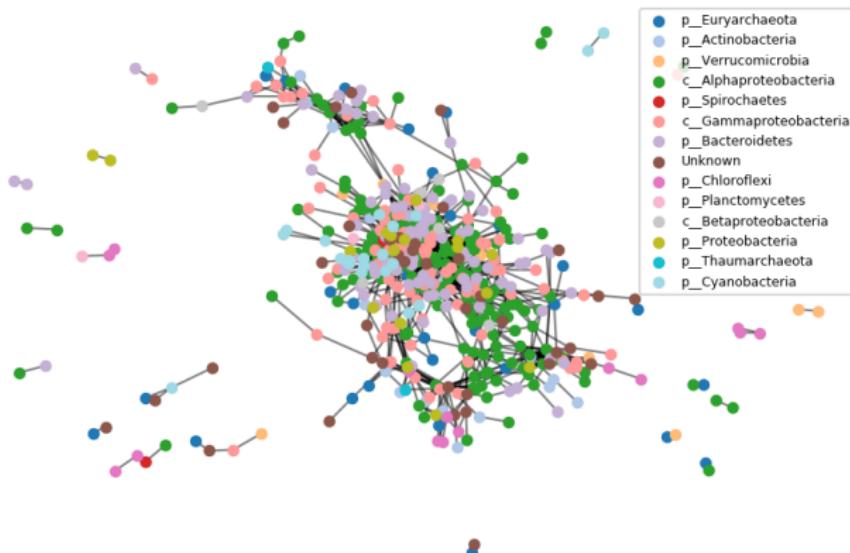
623 genomes with at least 1 significant correlation.

COMMUNITIES OF CO-ABUNDANT MAGS



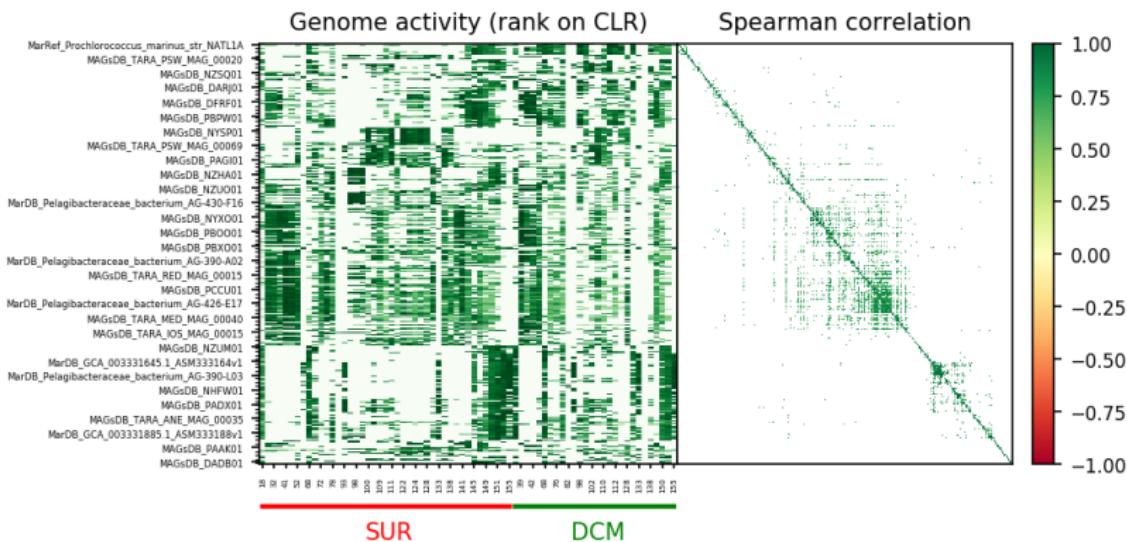
Clauset-Newman-Moore greedy modularity

COMMUNITIES OF CO-ABUNDANT MAGS



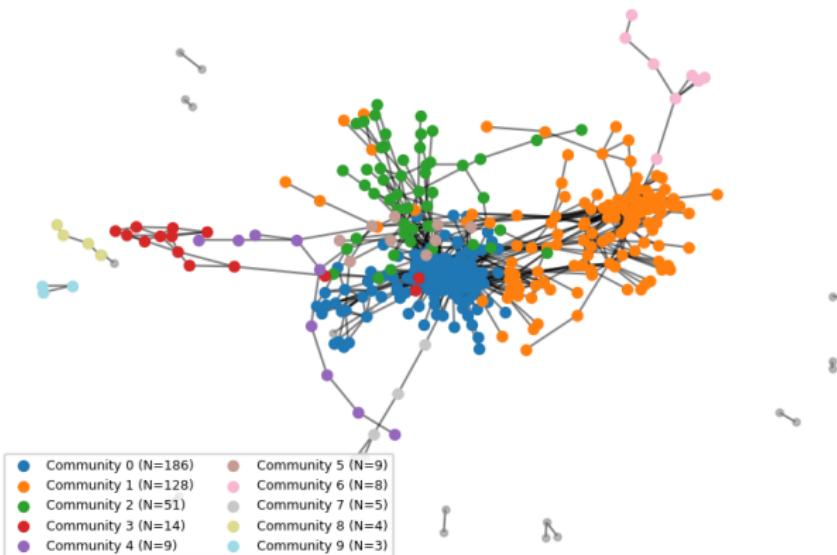
CO-ACTIVITY NETWORK

RESULTS



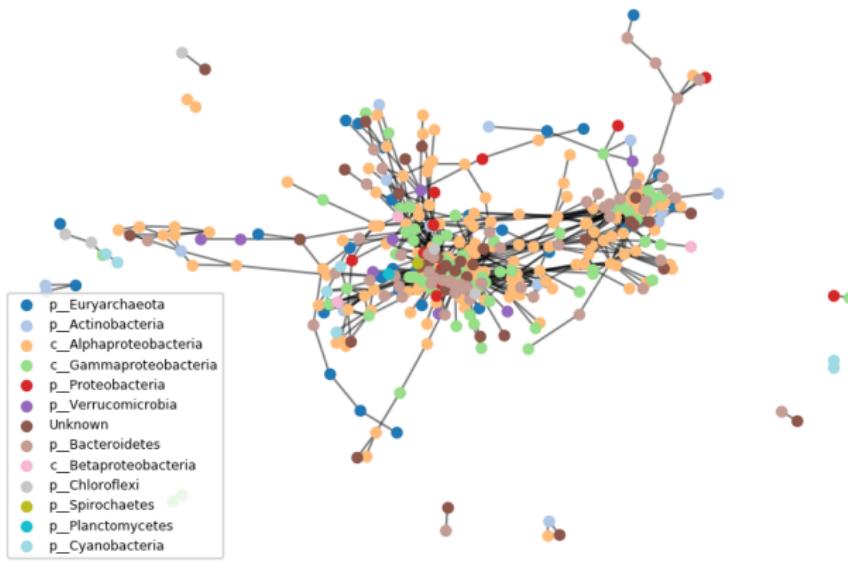
438 genomes with at least 1 significant correlation.

COMMUNITIES OF CO-ACTIVE MAGS



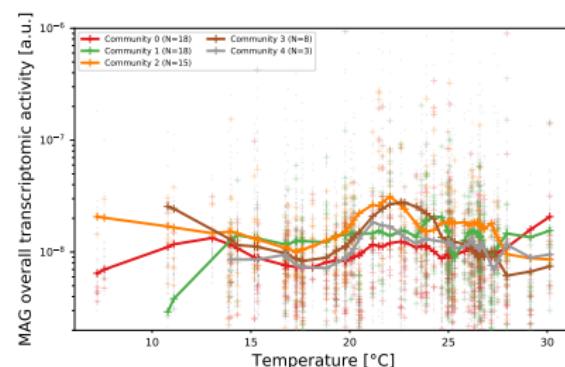
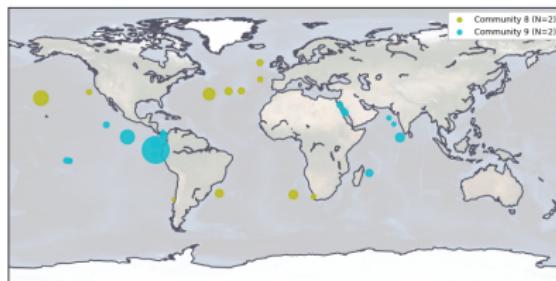
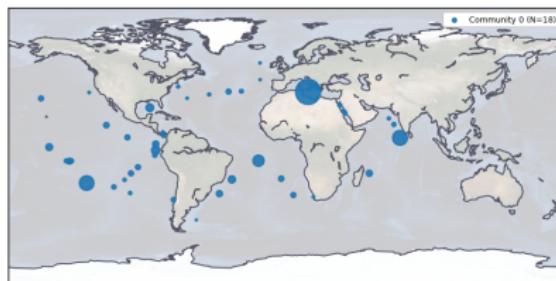
Clauset-Newman-Moore greedy modularity

COMMUNITIES OF CO-ACTIVE MAGS



Taxonomy assortativity: 0.039

LINKING COMMUNITIES TO ENVIRONMENTAL PARAMETERS

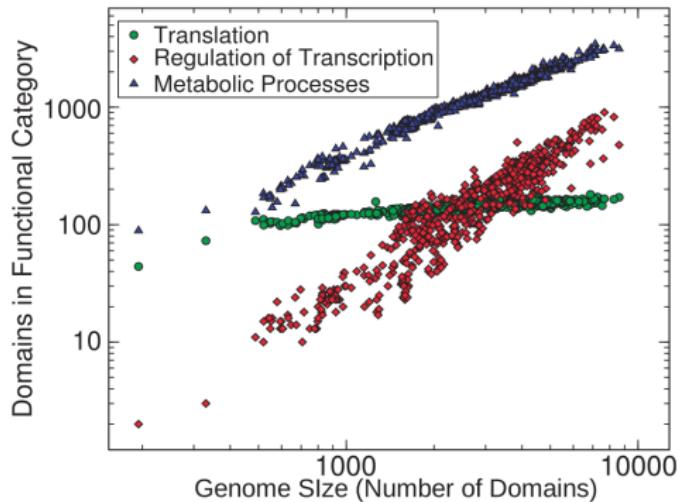


- Global/local communities
- Ecological niches

In progress...

MAGS VERSUS LAB-CULTIVATED GENOMES?

LAB-CULTIVATED GENOMES FOLLOW SCALING LAWS



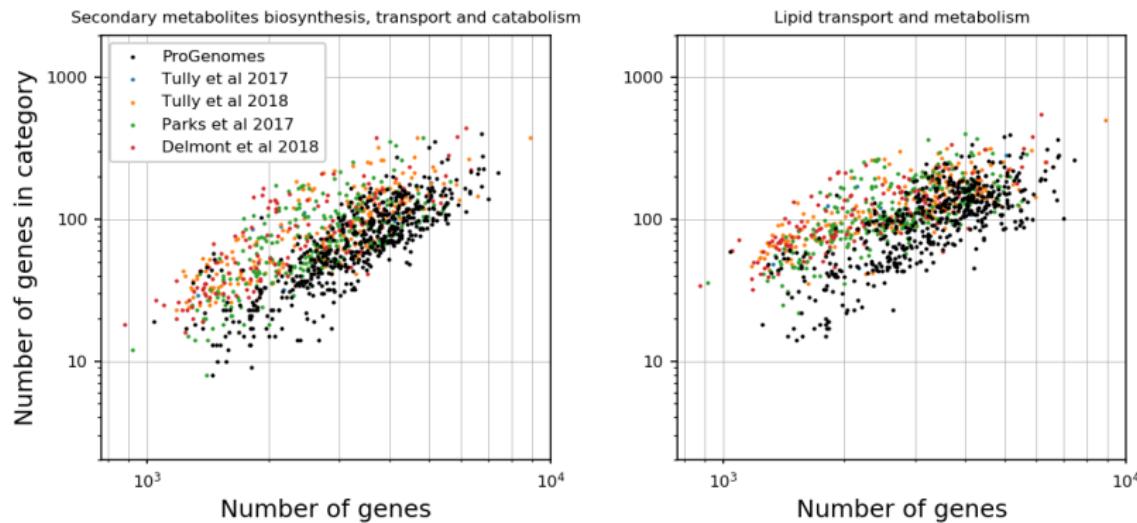
Functional scaling laws:

$$n_c \propto n_{tot}^{\alpha_c}$$

Annotation of MAGs
with Prodigal + EggNOG

MAGs VERSUS LAB-CULTIVATED GENOMES?

SCALING LAWS ON HIGH QUALITY MAGS



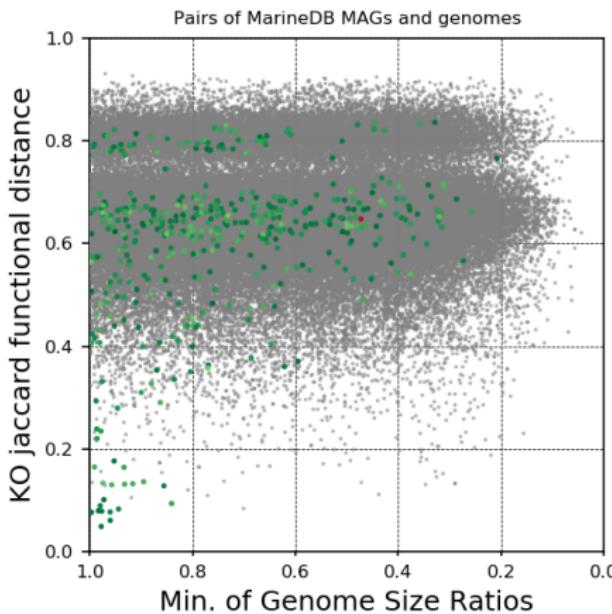
Cross-feeding widespread in non-cultivable organisms?

COMBINING FUNCTIONAL DISTANCE AND CO-ACTIVITY NETWORK

Jaccard distance:

K00001 K00002 K00003 K20112 K20113 K20114

Genome 1 [1, 1, 0, ..., 1, 1, 1]
Genome 2 [1, 0, 0, ..., 1, 0, 0]



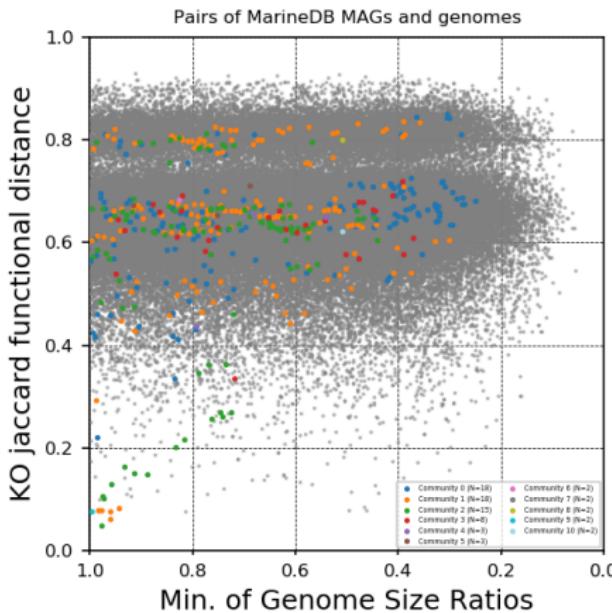
Most predicted interactions are between functionally distant organisms!

COMBINING FUNCTIONAL DISTANCE AND CO-ACTIVITY NETWORK

Jaccard distance:

K00001 K00002 K00003 K20112 K20113 K20114

Genome 1 [1, 1, 0, ..., 1, 1, 1]
Genome 2 [1, 0, 0, ..., 1, 0, 0]



Most predicted interactions are between functionally distant organisms!

WRAPPING UP

- Marine microbial diversity is largely unknown...
- ... but "environmental" genomes can be predicted...
- ... and putative communities can be inferred from co-trait networks (e.g. transcriptomic activity)

Perspective

- Co(community)-metabolic modeling to infer interactions (secondary metabolites cross-feeding)
- Analysis of larger size fractions (aggregates) and interactions with Eukaryotes?
- Co-replication network based on differential coverage (Korem et al, Science 2015)

THANK YOU FOR YOUR ATTENTION

Collaborators

- Samuel Chaffron
- Benjamin Churceward
- Damien Eveillard
- Marko Budinich Abarca
- Tara Oceans consortium



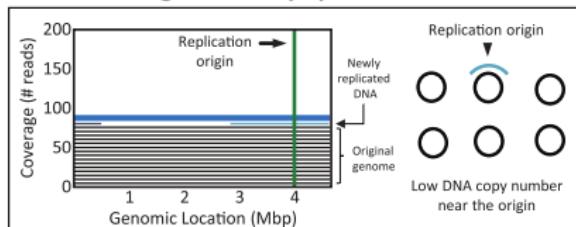
Funding

UNIVERSITE
BRETAGNE
LOIRE

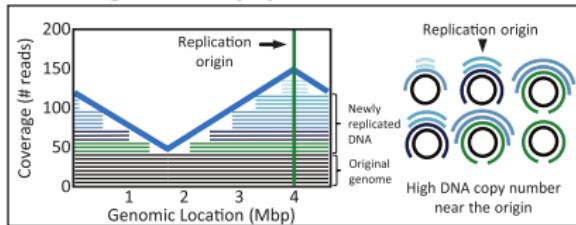


WHAT ABOUT CO-GROWTH? (REPLICATION)

Non-dividing bacterial population



Growing bacterial population



$$\frac{\text{Ori}_{\text{cov}}}{\text{Ter}_{\text{cov}}} \geqslant 1$$

A growing interest...

- Korem et al, Science 2015
- Brown et al, Nature Biotechnology 2016
- Emiola et al, Nature Communications 2018
- Gao et Li, Nature Methods 2018

PIPELINE OUTLINE

- Workflow management system: Snakemake
- Download and group genomes
- CheckM (taxonomy, completeness, contamination)
- Prodigal + eggNOG mapper (CDS prediction and annotation)
- dRep 95%
- Bowtie2 (mapping of MetaDNA and MetaRNA Tara samples)
- Pysam (read filtering MAPQ ≥ 20 , Identity $\geq 95\%$)
- Pysam (coverage estimation on whole genomes and annotated CDS)
- Data analysis on Jupyter Lab (Pandas, Matplotlib, Networkx)