

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

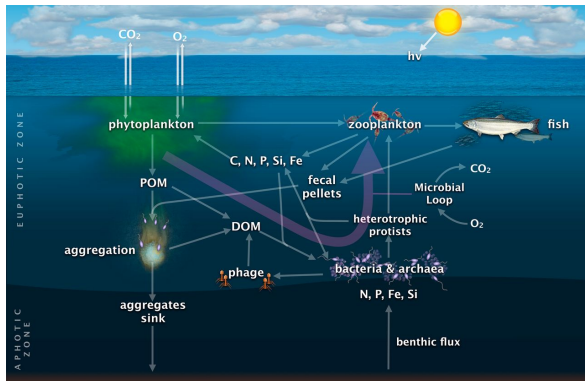
Nils Giordano, Samuel Chaffron

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

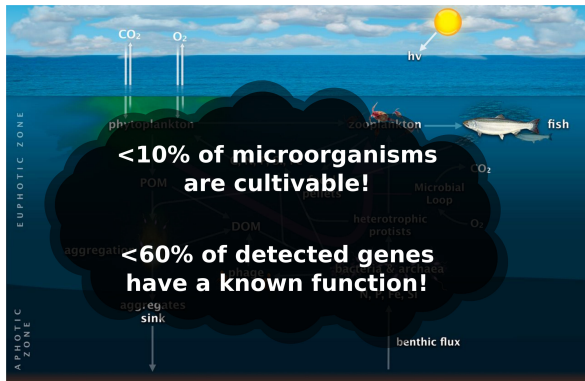
RCAM'19
September, 30th 2019 (Paris)



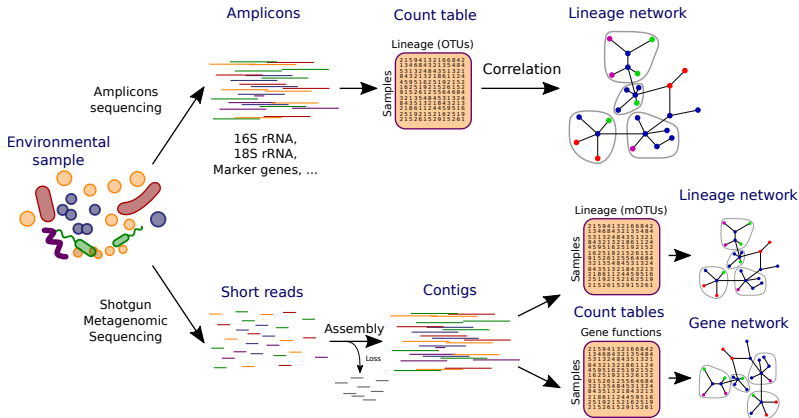
MARINE MICROBIAL COMMUNITIES PLAY CRUCIAL ECOLOGICAL ROLES



MARINE MICROBIAL COMMUNITIES PLAY CRUCIAL ECOLOGICAL ROLES

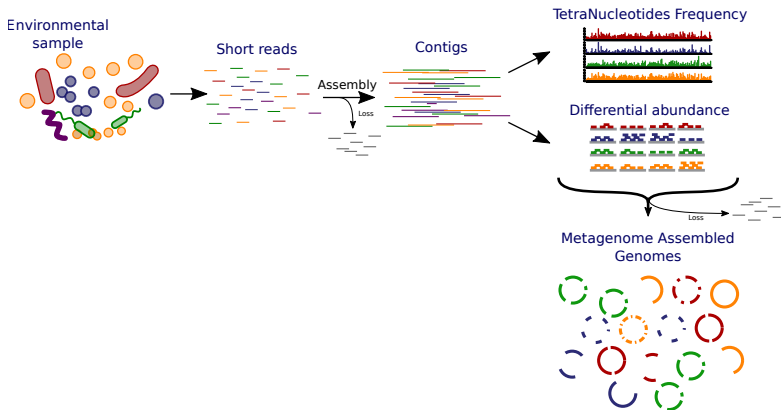


CULTURE-INDEPENDENT ENVIRONMENTAL GENOMICS



No link between taxonomy and functions?

HIGH QUALITY GENOMES CAN BE ASSEMBLED FROM METAGENOMES



LITERATURE IS FILLING WITH NEW METAGENOME ASSEMBLED GENOMES (MAGs)

2014

nature
biotechnology

Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes

H Bjørn Nielsen
Shinichi Sunaga
Eric Pelletier^{1,2}
Jean-Michel Batt
Fouad Boumezz
Torben Hansen^{1,3}
Pierre Léonard^{3,4}
Oluf Pedersen^{1,3}
Sebastian Tims²
Thomas Sicherit

Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life

Donovan H. Parks
Paul N. Evans, Phi

2017

nature
microbiology

2017/18

Data Descriptor: The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans

SCIENTIFIC DATA

Benjamin J. Tully¹

Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes

nature
microbiology

2018

Tom O. Delmont^{1*}, Christopher Quince²
Michael S. Rapp⁴, Sandra L. MacLellan¹

Metabolic Diversity within the Globally Abundant Marine Group II *Euryarchaea* Drives Ecological Patterns

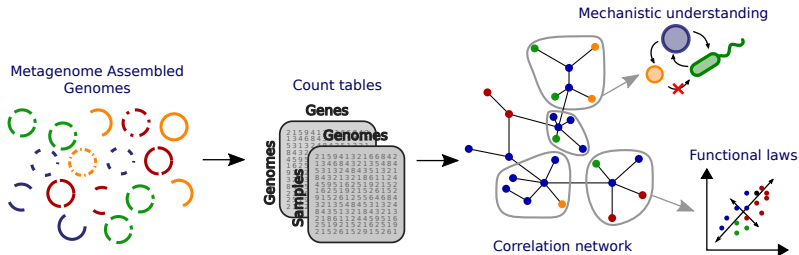
Benjamin J Tully

2018



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

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

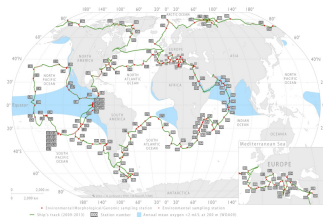
7658*Dereplication: 95% ANI over $\geq$ 60% of the genome (tool: dRep)

METAGENOMES

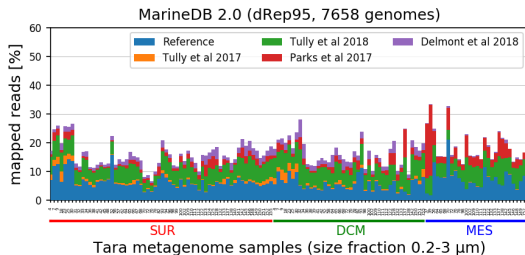
Tara expeditions dataset (2009-2013)

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

*ribo-depleted MetaRNA (ribosomal RNA chemically removed)



COVERAGE ACROSS TARA SAMPLES



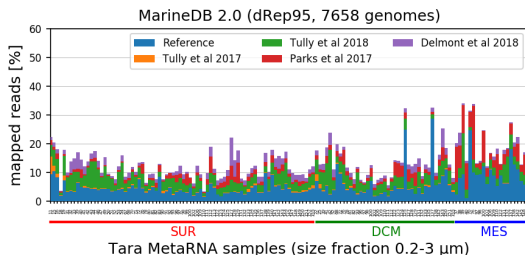
MetaDNA

Total: 16.6%

SUR: 15.8%

DCM: 16.5%

MES: 18.5%



MetaRNA

Total: 13.5%

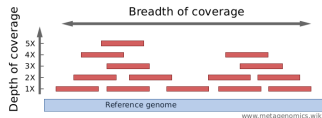
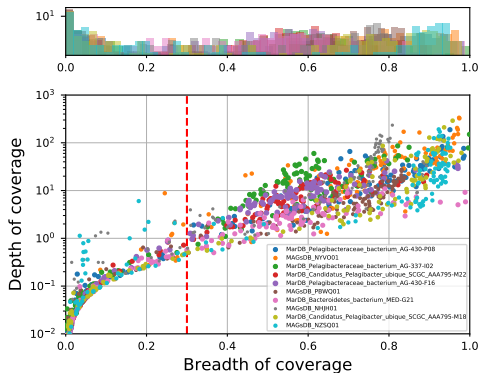
SUR: 11.8%

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

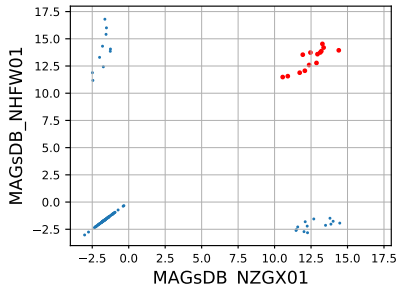
- Problem with zeros $\rightarrow$ pseudocount

Gloor, G. B. et al, Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 8, (2017)

CO-ABUNDANCE NETWORK

CORRELATION

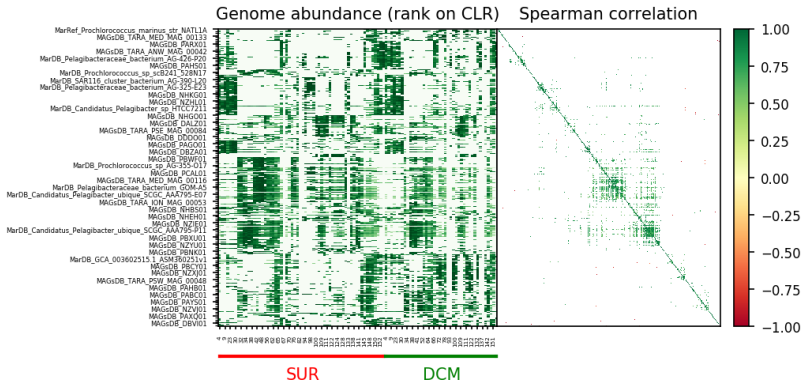
- Spearman correlation on **'co-occurring' datapoints only**
→ Constraint: co-occurrence in at least 10% of samples (N=11)



- Benjamini-Hochberg FDR correction ($\text{FDR} \leq 0.01$)

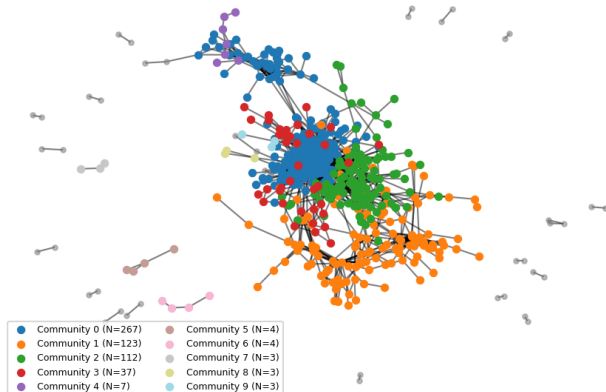
CO-ABUNDANCE NETWORK

RESULTS



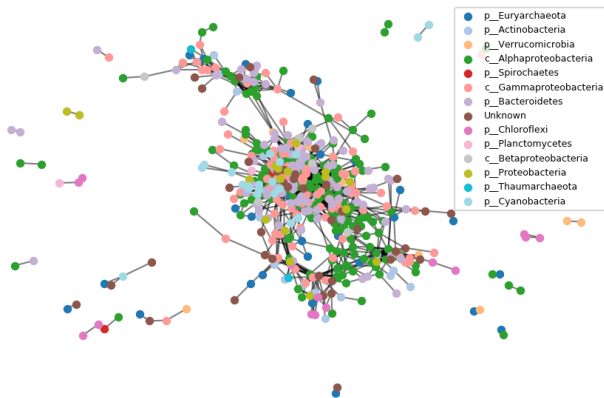
623 genomes with at least 1 significant correlation.

COMMUNITIES OF CO-ABUNDANT MAGS



Clauset-Newman-Moore greedy modularity

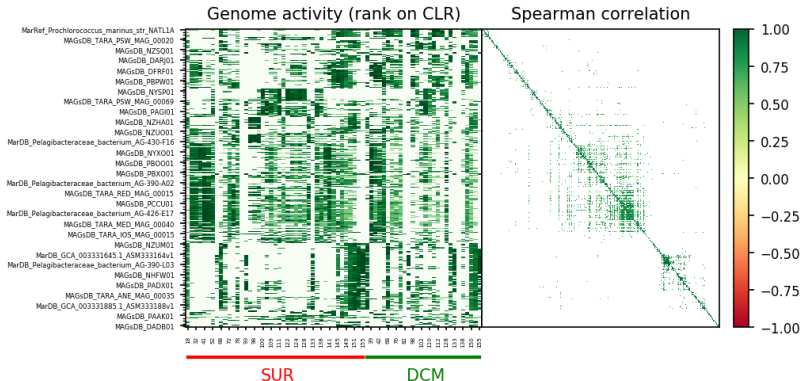
COMMUNITIES OF CO-ABUNDANT MAGs



Taxonomy assortativity: 0.032

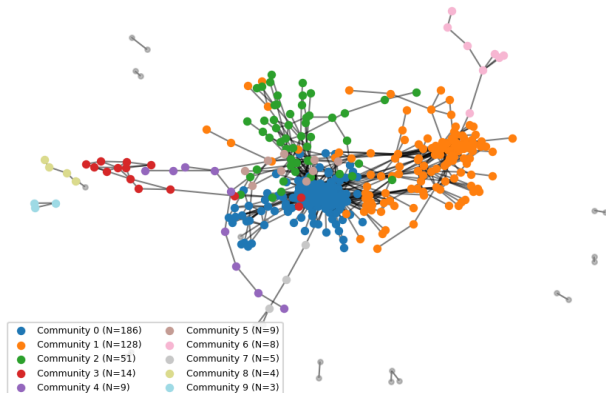
CO-ACTIVITY NETWORK

RESULTS



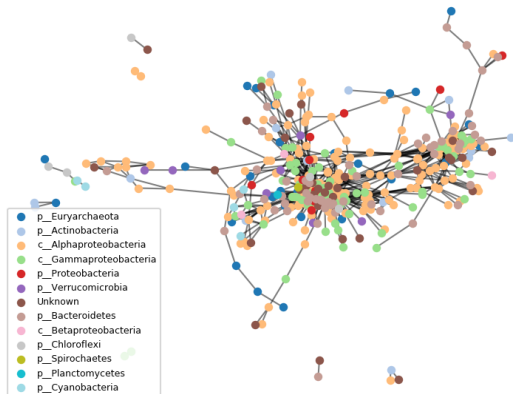
438 genomes with at least 1 significant correlation.

COMMUNITIES OF CO-ACTIVE MAGs



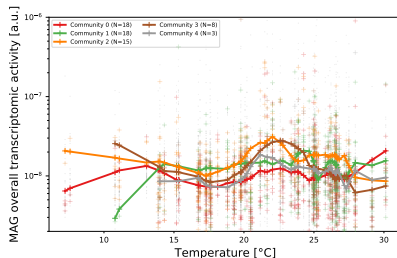
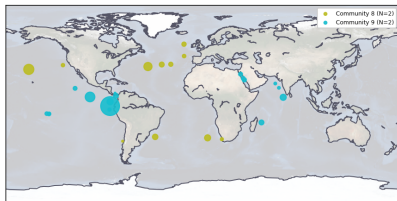
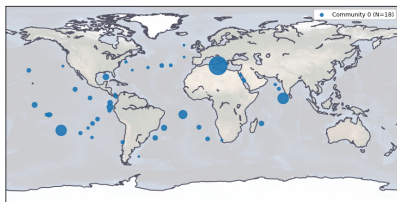
Clauset-Newman-Moore greedy modularity

COMMUNITIES OF CO-ACTIVE MAGs



Taxonomy assortativity: 0.039

LINKING COMMUNITIES TO ENVIRONMENTAL PARAMETERS

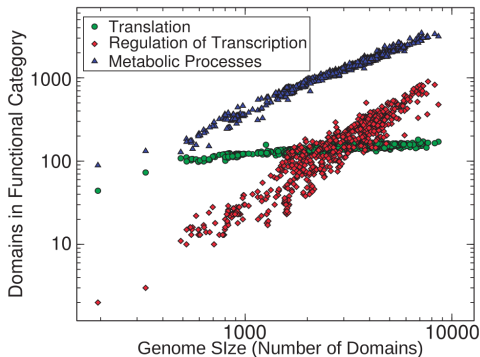


- Global/local communities
- Ecological niches

In progress...

MAGs VERSUS LAB-CULTIVATED GENOMES?

LAB-CULTIVATED GENOMES FOLLOW SCALING LAWS



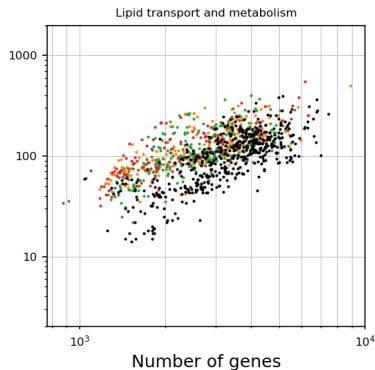
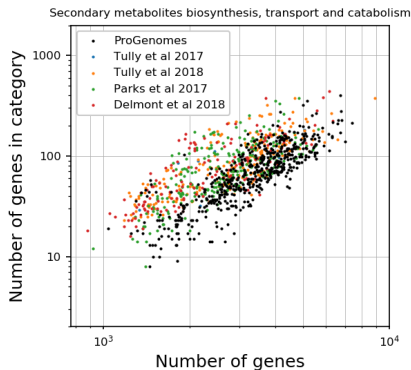
Functional scaling laws:

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

Annotation of MAGs
with Prodigal + EggNOG

MAGs VERSUS LAB-CULTIVATED GENOMES?

SCALING LAWS ON HIGH QUALITY MAGs

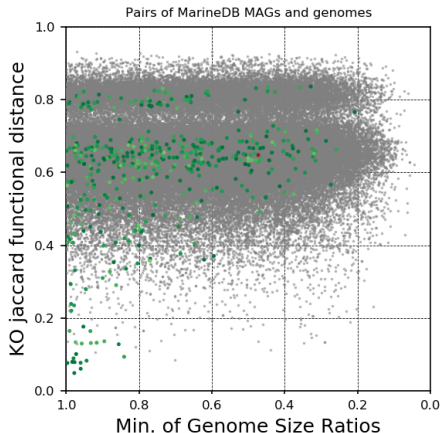


Cross-feeding widespread in non-cultivable organisms?

COMBINING FUNCTIONAL DISTANCE AND CO-ACTIVITY NETWORK

Jaccard distance:

		K00001	K00002	K00003	K20112	K20113	K20114	
Genome 1	[1, 1, 0, ..., 1, 1, 1]	1	1	0	...	1	1	1
Genome 2	[1, 0, 0, ..., 1, 0, 0]	1	0	0	...	1	0	0

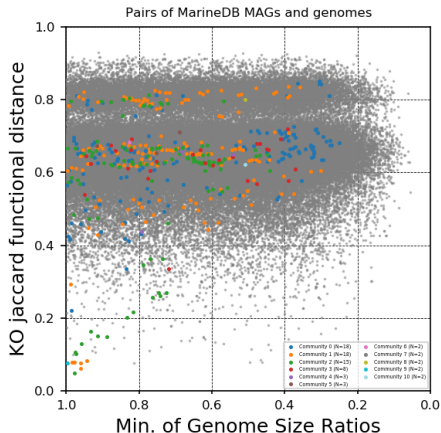


Most predicted interactions are between functionally distant organisms!

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(mmunity)-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 Churcheward
- Damien Eveillard
- Marko Budinich Abarca
- Tara Oceans consortium

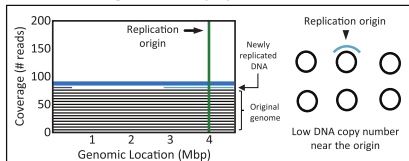


Funding



WHAT ABOUT CO-GROWTH? (REPLICATION)

Non-dividing bacterial population

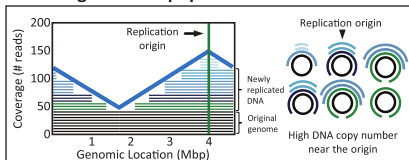


$$\frac{Ori_{cov}}{Ter_{cov}} \geq 1$$

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

Growing bacterial population



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 $\text{MAPQ} \geq 20$, $\text{Identity} \geq 95\%$)
- Pysam (coverage estimation on whole genomes and annotated CDS)
- Data analysis on Jupyter Lab (Pandas, Matplotlib, Networkx)