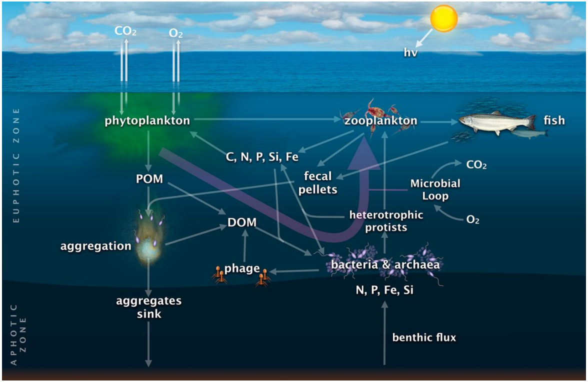
Co-activity networks reveal the structure of planktonic symbiosis in the global ocean

Nils Giordano, Samuel Chaffron

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

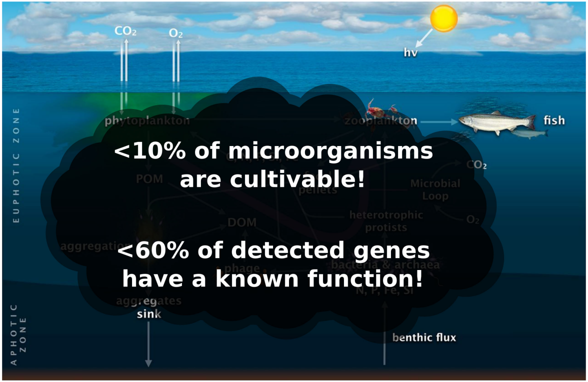
GDR-GE'19
October, 8-10th 2019 (La Rochelle)

MARINE MICROBIAL COMMUNITIES PLAY CRUCIAL ECOLOGICAL ROLES



Worden *et al*, Science 2015

MARINE MICROBIAL COMMUNITIES PLAY CRUCIAL ECOLOGICAL ROLES

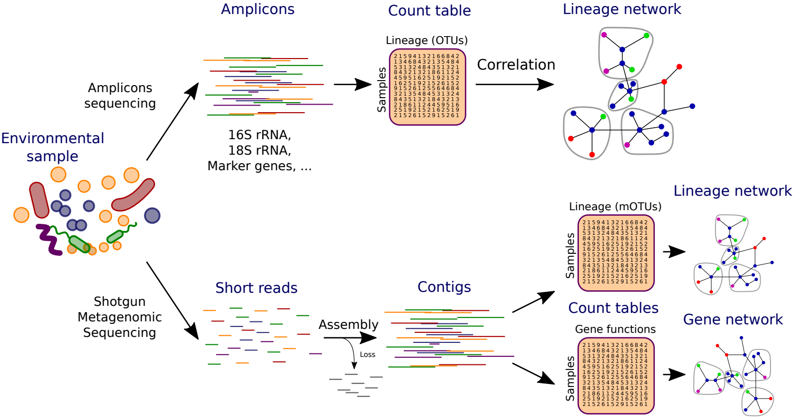


<10% of microorganisms are cultivable!

<60% of detected genes have a known function!

Worden *et al*, Science 2015

CULTURE-INDEPENDENT ENVIRONMENTAL GENOMICS



Environmental sample

Amplicons sequencing

Amplicons

16S rRNA,
18S rRNA,
Marker genes, ...

Count table

Lineage (OTUs)

Samples

Correlation

Lineage network

Shotgun Metagenomic Sequencing

Short reads

Assembly

Loss

Contigs

Lineage (mOTUs)

Samples

Lineage network

Count tables

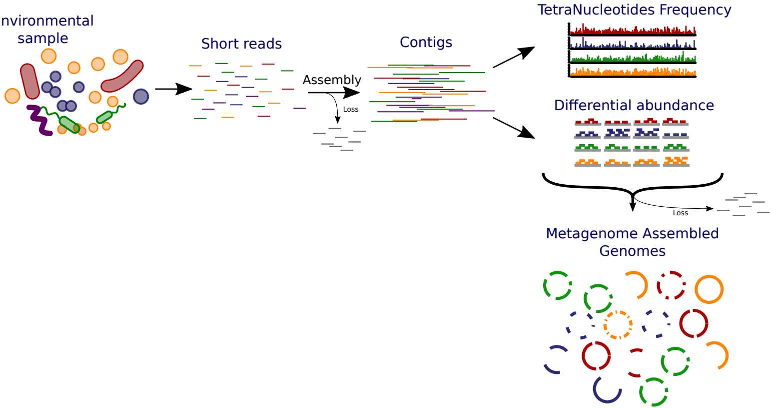
Gene functions

Samples

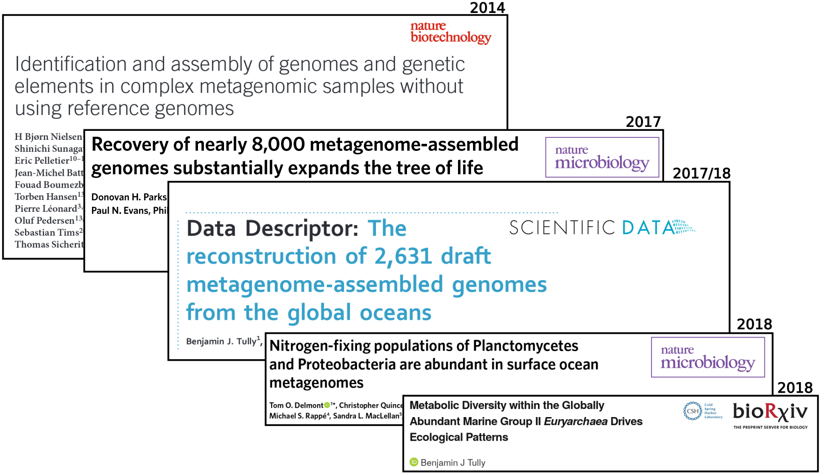
Gene network

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[10–
Jean-Michel Batt
Fouad Boumezb
Torben Hansen[13
Pierre Léonard[3,
Oluf Pedersen[13,
Sebastian Tims[2
Thomas Sicherit

2017

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

nature microbiology

Donovan H. Parks
Paul N. Evans, Phi

2017/18

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

SCIENTIFIC DATA

Benjamin J. Tully[1],

2018

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

nature microbiology

Tom O. Delmont[1]*, Christopher Quince
Michael S. Rappé[4], Sandra L. MacLellan[5

2018

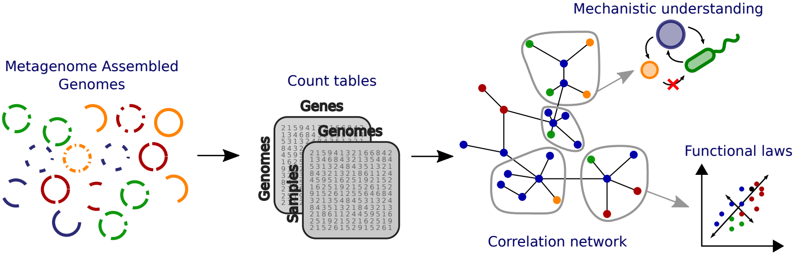
Metabolic Diversity within the Globally Abundant Marine Group II *Euryarchaea* Drives Ecological Patterns

CSH Cold Spring Harbor Laboratory

bioRχiv THE PREPRINT SERVER FOR BIOLOGY

Benjamin J Tully

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?

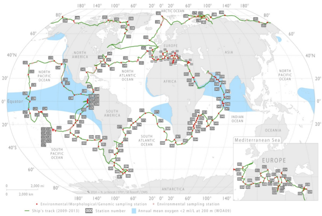
DATASETS

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 ⩾ 60% of the genome (tool: dRep)

METAGENOMES

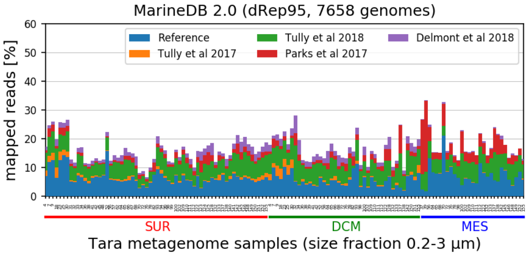


Tara expeditions dataset (2009-2013)

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

*ribo-depleted MetaRNA (ribosomal RNA chemicaly removed)

COVERAGE ACROSS TARA SAMPLES



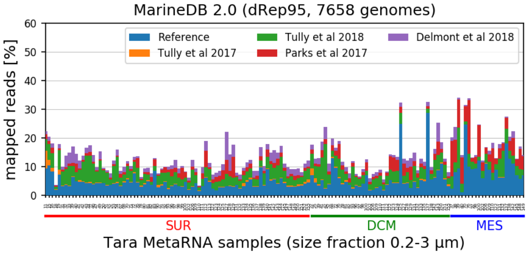
MetaDNA

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-ACTIVITY NETWORK

NORMALIZATION

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

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

- Center 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 $G(x) = \sqrt[D]{x_1 \cdot \ldots \cdot x_D}$

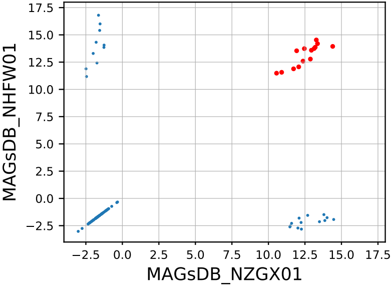
- Problem with zeros → pseudocount

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

CO-ACTIVITY NETWORK

CORRELATION

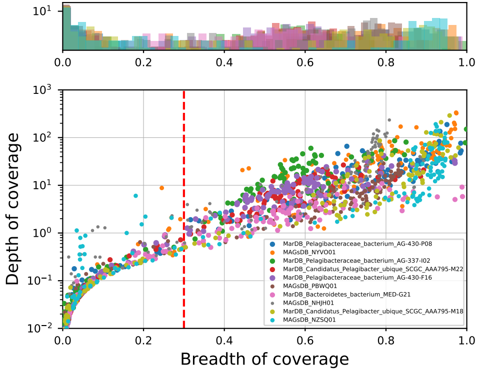
- Spearman correlation **on 'co-occuring' datapoints only**
  → Constraint: co-occurence in at least 10% of samples (N=11)

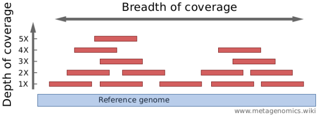


- Benjamini-Hochberg FDR correction (FDR $\leqslant 0.01$)

CO-ACTIVITY NETWORK

HOW TO DEFINE OCCURENCE OF A GENOME?

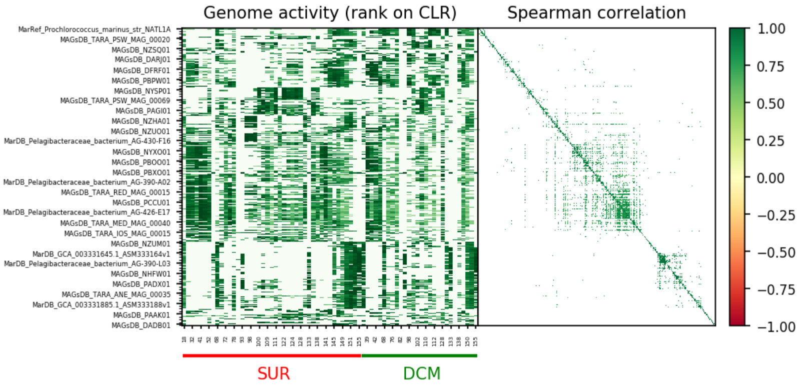




- 30% threshold on coverage breadth (horizontal coverage)

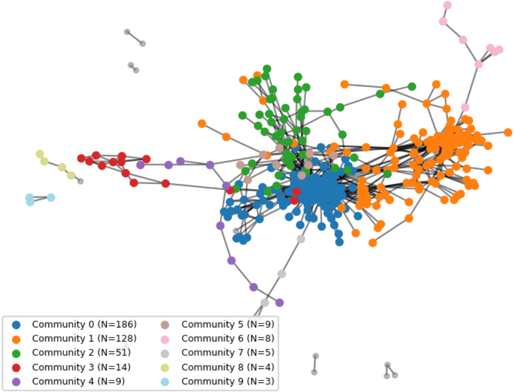
CO-ACTIVITY NETWORK

RESULTS



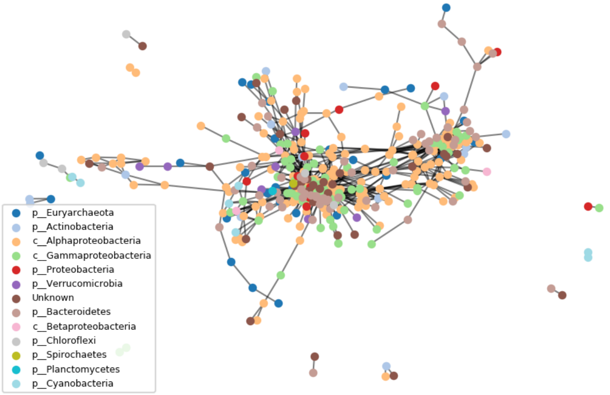
438 genomes with at least 1 significant correlation.

COMMUNITIES OF CO-ACTIVE MAGS



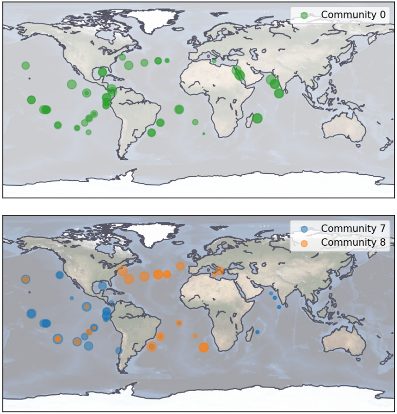
Clauset-Newman-Moore greedy modularity

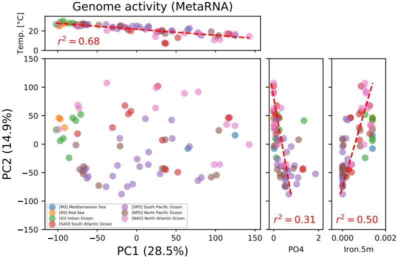
COMMUNITIES OF CO-ACTIVE MAGS



Taxonomy assortativity: 0.039

LINKING COMMUNITIES TO ENVIRONMENTAL PARAMETERS


Genome activity (MetaRNA)

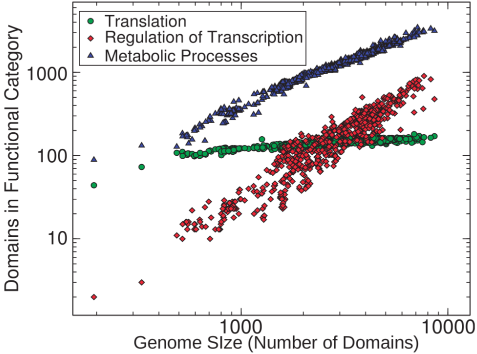


- Global/local communities
- Communities structured by environment

In progress...

MAGS VERSUS LAB-CULTIVATED GENOMES?

LAB-CULTIVATED GENOMES FOLLOW SCALING LAWS



Functional scaling laws:

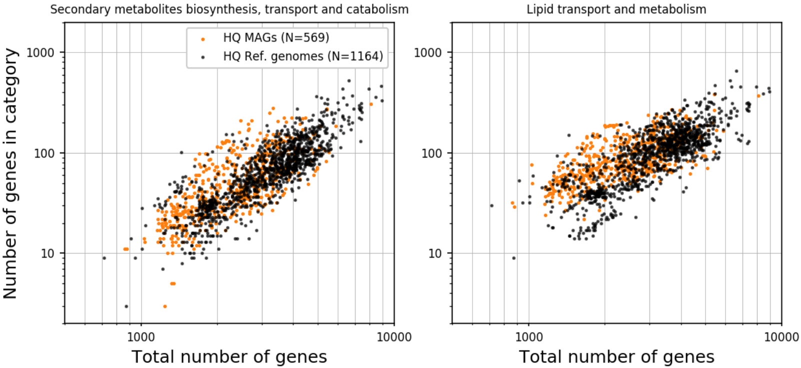
$$n_c \propto n_{tot}^{\alpha_c}$$

Annotation of MAGs with Prodigal + EggNOG

Grilli *et al*, Nucleic Acid Research 2012

MAGS VERSUS LAB-CULTIVATED GENOMES?

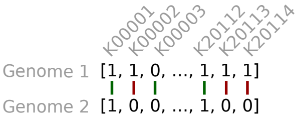
SCALING LAWS ON HIGH QUALITY MAGS

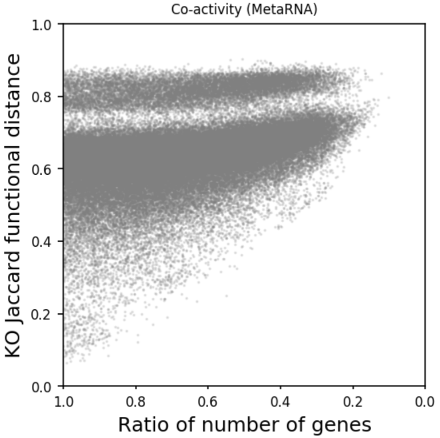


Cross-feeding widespread in non-cultivable organisms?

COMBINING FUNCTIONAL DISTANCE AND CO-ACTIVITY NETWORK

Jaccard distance:

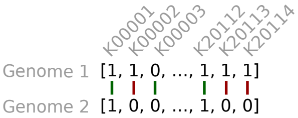


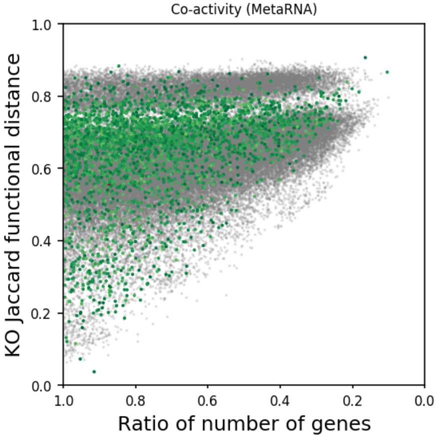


Most predicted interactions are between functionally distant organisms!

COMBINING FUNCTIONAL DISTANCE AND CO-ACTIVITY NETWORK

Jaccard distance:





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 infered 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

Fondation tara océan explore and share

Plateforme BiRD Nantes

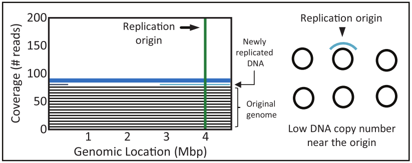
Funding

UNIVERSITE BRETAGNE LOIRE

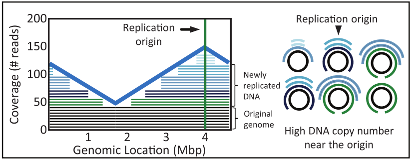
RECHERCHE FORMATION INNOVATION en Pays de la Loire

WHAT ABOUT CO-GROWTH? (REPLICATION)

Non-dividing bacterial population



Growing bacterial population



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

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