Functional diversity of ocean microbiomes.

Dustin Michels

Nov 2017

Abstract

No abstract yet…

Introduction

There are important differences

When comparing communities of macroorganismal organisms, one might compute metrics of species diversity. In microbial communities however, species are harder to define, in part due to the frequency with which horizontal gene transfer occurs. Some have suggested shifting focus away from the notion of species towards distributions of genes, when evaluating microbial communities. [1].

Furthermore, while geographic isolation is considered a major factor in determining the distinctiveness of macroorganismal populations, microbiology has long entertained the hypothesis that 'everything is everywhere: but the environment selects'– articulated by Lourens G. M. Baas Becking in 1934 [1]. If taxonomy is

Thus, one question in microbiology is what can

Various microbiomes

"Functional redundancy"

has been observed in many environments

The idea t– remained a widespread and well-regarded feature of microbiological throughout the 20th century.

Various authors have found that differences in taxonomic composition reveal less about the particular physicochemically of an ecosystem than differences in functional composition.[2].

In this paper, we use metagenomics first to characterize taxonomic and functional diversity across eleven disparate Tara Ocean samples, and evaluate evidence of functional redundancy. Secondly, we will investigate environmental drivers of taxonomic and functional diversity, and especially whether environmental or geographic factors play a more important role. It is expected that functional redundancy will observable, and that environmental factors will prove a better differentiator of both functional and taxonomic diversity than geographic location.

In their 2015 paper "Structure and function of the global ocean microbiome", Sunagawa et al., analyzed metagenomic data from 243 *Tara Oceans* samples from around the world, in part to characterize the oceans taxonomic and functional diversity, and consider factors that could be driving taxonomic and functional stratification.

- temp and diversity

    "

Differences in metabolic function among organisms are thought to underlie much of this variation as a result of selection for specific metabolic pathways based on physicochemical conditions ("metabolic niche effects" or "environmental filtering").

However, other factors such as biotic interactions (5-7), limits to spatial dispersal (4), and neutral demo- graphic drift (8) could also affect community composition"[2].

which factors drive strafaction of taxonomic and functional groups

"However, the recent claim that geographical barriers do not exist for any micro-organism (Finlay 2002) has renewed the search for examples of 'microbial marsupials' (Fenchel 2003)." [3]

"In general, environmental conditions strongly predicted the functional profiles of microbial communities but weakly predicted the taxonomic composition"[2]

"most environmental variables either correlated more strongly with relative functional group abundances than with OTU proportions within functional groups

" In particular, an organism's metabolic potential appears to be the main trait selected for by environmental conditions across the global ocean.""

One paper that addresses these questions is "Structure and fucntion of global ocean microbiome" by Sunagawa et al. [4].

"The bulk of global biogeochemical fluxes is driven by a core set of metabolic pathways that evolved in response to past geochemical condi- tions (1). Through time, these pathways have spread across microbial clades that compete within metabolic niches, resulting in an enormous microbial diversity characterized by high functional redundancy"[2]

This project is inspired by the paper "Structure and function of the global ocean microbiome" by Sunagawa et al. [4], as well as the paper "An obesity-associated gut microbiome with increased capacity for energy harvest" by Turnbaugh et al. [**???**]. The first paper seeks to characterize staxonomic differences and differences in gene function between ocean microbiomes, using Tara Ocean samples from around the world. They seek to identify which factors could best explain that variation. The second paper, by Turnbaugh et al., at one point describes the differences between the distal gut microbiomes of mice using functional annotations of the various microbiomes, and visualizes the result using heat maps.

My goal in this paper was to employ some of the tactics of the Turnbaugh et al paper– understanding the functional differences between metagenomes using heat maps– to the domain of the Sunagawa et al. paper, characterizing functional differences between ocean regions.

Questions: * Which are the dominant functional and taxonomic groups? *

One thing to note: I am interested in the challenge of ensuring reproducibility in bioinformatics projects, and have attempted to structure this investigation for maximum reproducibility. Namely,

- All files are being stored (at various stages) using a public GitHub repository (See: https://github.com/dustinmichels/biol338-final-project/tree/master). The code associated with this draft is tagged as "draft01" and be retrieved in its entirety from GitHub.
- I am attempting to carry out most stages of data collection, analysis, and presentation, using scripts that should be viable to re-run on other machines at other times to generate the same results. Even this report is being written in plain-text (Markdown) and converted into a report using a tool called Pandoc!
- Finally, I am taking care to document the software I am using, and which version I am using, and how I am using it.

Materials & Methods

Data

I have

I have used the eleven Tara Ocean samples that our class has used throughout the term, but I have downloaded Gene Ontology (GO) files for each dataset, which detail the number of reads from that dataset that mapped to each of about 2000 different functional categories, pertaining to molecular function, cellular components, or biological processes.

For each dataset, I normalized the 'read count' into a 'read proportion' (the number of reads that mapped to a given functional category out of all the reads.) I then selected the top 30 most abundant functional categories, for each dataset. I ended up dropping three of these functional categories from my analysis, which had missing data for one or more samples. That is to say, I wound up plotting the *union* of the 30 most abundant functional groups from each sample, which is 27 functional groups.

I created three heat maps, one for all eleven samples, one grouped by region, and one grouped by zone.

I then created three "cluster maps." A cluster map is similar to a heat map, except it clusters data hierarchically. It calculates pairwise Euclidean distance between observations in n-dimensional space, and groups closely related attributes together. It rearranges data accordingly, and even draws tree structures resembling phylogenies to indicate groupings. This method produced pretty interesting results, but I will need to further investigate exactly how this works to ensure the results are meaningful.

Figure 1: Map of sample sites.

The exact steps I took to do this analysis can be seen and reproduced on-line at http://bit.ly/2zEMitP.

Functional Annotations

EMBL-EBI GO annotations. Gene Ontology (GO) terms derived from InterPro matches to your sample.

Taxonomy

Mothur (version 1.38.1)

Notes on exact commands are available at: https://github.com/dustinmichels/biol338-final-project/tree/master/data/taxonomy

classified against to SILVA database

Further Analysis and Plotting

11 Samples. Downloaded Complete GO annotation

EBI Metagenomics

https://www.ebi.ac.uk

Results

Here are the heat maps and cluster maps that I generated.

Heat map for all samples, and for samples grouped by region, and for samples by grouped by zone. The top 30 most abundant functional groups are on the vertical axis, and the proportion of reads that mapped to that functional group is shown by the heat map color.

Cluster map for all samples

Cluster map, grouped by region

Cluster map, grouped by zone

Heat / Cluster

- Functional diversity smoother than taxonomic

Joint

- Functional diversity correlated with

The fact that depth captures diversity better than other metadata may suggest that there is an aspect to "depth" not captured by the other variables.

OCEAN PAPER:

They suggest that temperature and light have stronger effects on functional trait composition than nutrients or salinity.

Taxonomic compositions do not show a clear separation by regional origin

Global ocean paper suggests "increase in taxonomic and functional" richness with depth.

The idea that the

Discussion

The cluster maps seem to suggest that deep chlorophyll maximum layer and the surface water later are more closely related to each other, in terms of prevalence of functional gene categories, prevalence than either is to the mesopelagic zone

Similarly, the cluster maps would suggest that the North Atlantic and South Atlantic are the most functionally similar, followed by the North Pacific, followed by the Souther Ocean, and finally the Arabian Sea.

More analysis underway!

They stated members of Proteobacteria, including Alphaproteobacteria and Gammaproteobacteria, as well as Cyanobacteria, Deferribacteres, and Thaumarchaeota.

For me, temperature shadowed by depth.

Functional redundancy. (including Cyanobacteria, De)

"This has implications for the interpretation of differences in community structure across envi- ronments and time. Differences in taxonomic composition that do not affect functional com- position may have little relevance to ecosystem biochemistry; conversely, physicochemically sim- ilar environments could host taxonomically dis- tinct communities (26). Functional (rather than purely taxonomic) descriptions of microbial com- munities should therefore constitute the baseline for microbial biogeography, particularly across transects where geochemical gradients shape microbial niche distribution"[2]

Further Research

- Look at relatedness of species? (more related across depths or regions?)

References

1. O'Malley MA. "Everything is everywhere: But the environment selects": Ubiquitous distribution and ecological determinism in microbial biogeography. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences. 2008;39: 314–325. doi:10.1016/j.shpsc.2008.06.005

2. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. Science. 2016;353: 1272–1277. doi:10.1126/science.aaf4507

3. Whitaker RJ. Allopatric origins of microbial species. Philosophical Transactions of the Royal Society B: Biological Sciences. 2006;361: 1975–1984. doi:10.1098/rstb.2006.1927

4. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science. 2015;348: 1261359. doi:10.1126/science.1261359

Appendix

Table 1: For each sample, a 'Reads encoding 5S rRNA' file and a 'Complete GO annotation' file was downloaded from the given link.

Run ID	Description	Link
ERR599104	Southern Ocean (DCM)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS491095/runs/ERR599104/results/versions/2.0
ERR599090	Southern Ocean (SURF)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS491044/runs/ERR599090/results/versions/2.0
ERR599008	Southern Ocean (MESO)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS491110/runs/ERR599008/results/versions/2.0
ERR598948	South Pacific (DCM)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS492699/runs/ERR598948/results/versions/2.0
ERR598992	South Pacific (SURF)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS492642/runs/ERR598992/results/versions/2.0
ERR598999	South Pacific (MESO)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS492680/runs/ERR598999/results/versions/2.0
ERR598995	North Pacific (DCM)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS493340/runs/ERR598995/results/versions/2.0
ERR598980	North Pacific (MESO)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS493372/runs/ERR598980/results/versions/2.0
ERR599142	North Pacific (SURF)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS493300/runs/ERR599142/results/versions/2.0
ERR599078	North Atlantic (SURF)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS494579/runs/ERR599078/results/versions/2.0
ERR599031	Arabian Sea (MESO)	https://www.ebi.ac.uk/metagenomics/projects/ERP001736/samples/ERS488769/runs/ERR599031/results/versions/2.0