Accompanying Material

The following material provides additional information on [1]:

1. Network construction

2. Discussion of false negative interactions

3. Discussion of putative false positive interactions

4. Acantharia and *Phaeocystis*

The registry of samples used in this work is available at <http://doi.pangaea.de/10.1594/PANGAEA.840721>. Sample associated environmental data is archived in [http://doi.pangaea.de/10.1594/PANGAEA.840718](http://doi.pangaea.de/10.1594/PANGAEA.840718%0B).

This article is part of a Special Issue on Ocean plankton published in Science. Complementary information on sampling, OTU definition annotations can be found in [2-5], http://ocean-microbiome.embl.de/companion.html and Eukaryotic fractions: <http://taraoceans.sb-roscoff.fr/EukDiv/> .

The Tara Oceans nucleotide sequences are available at the European Nucleotide Archive (ENA) under projects PRJEB402 and PRJEB6610. Taxonomic and trophic mode annotation of eukaryotic amplicons is archived at <http://doi.pangaea.de/10.1594/PANGAEA.843022> and <http://doi.pangaea.de/10.1594/PANGAEA.843018>.

## 1. Network construction

Taxon-taxon co-occurrence networks were constructed following the procedure outlined in [6], with Spearman and Kullback-Leibler dissimilarity as measures of similarity. Briefly, for each similarity measure, a distribution of all pair-wise scores was computed. Given these distributions, initial thresholds were selected such that the initial network contained 10,000 to 20,000 positive and negative edges, supported by each measure. For each measure and each edge, 1,000 renormalized permutation and bootstrap scores were generated, following the ReBoot routine, which mitigates compositional bias [6] (the renormalization step was omitted for Kullback-Leibler, which is robust to compositional bias). The measure-specific p-value was then computed as the probability of the null value (represented by the mean of the null distribution) under a Gauss curve generated from the mean and standard deviation of the bootstrap distribution. Edges with original edge scores outside the 95% confidence interval provided by the bootstrap distribution were discarded. Since a one-sided test was carried out, p-values close to one were considered indicative of mutual exclusion and were converted into low p-values by subtraction from one. P-values were then merged using Brown's method [7] and corrected for multiple testing using Benjamini-Hochberg's procedure [8]. Edges with a p-value above 0.05 were discarded. The procedure is implemented in CoNet (http://systemsbiology.vub.ac.be/conet/).

## 2. Discussion of false negative interactions

From the 91 known parasitic and mutualistic interactions that could be mapped to the TARA data set, 18 were correctly predicted, whereas 73 known interactions were missed. The majority of these false negatives (68) either contains OTUs that do not occur in enough samples to carry out co-occurrence analysis and were therefore discarded from the input matrices (48 cases) or they do not pass the initial thresholds for Spearman correlation and Kullback Leibler dissimilarity.

Two examples of the latter type are plotted in Figure 1 and Figure 2; both cases represent known endosymbioses in plankton, where the host (the radiolaria *Collozoum* and the dinoflagellate *Noctiluca*) contain intracellular microalgae. The photosymbiosis between *Collozoum* and *Scrippsiella* (Dinophyceae, recently renamed as *Brandtodinium*) is known to be obligatory for the host, but may be facultative for Scrippsiella [4] (for instance, this alga can grow in host-free culture media). Facultative photosymbiosis for the symbiont is common in another Radiolarian group, Acantharia, where symbiotic *Phaeocystis* cells (haptophytes) are found in high abundance outside their host cells [5].

The association between *Noctiluca* and *Pedinomonas* has only been observed in the Indian Ocean and characterized only based on the morphology [6]. It may not have been predicted from the TARA data because (i) it is limited to one single oceanic region, (ii) the endosymbiont also exists in a free-living form, (iii) the symbiont may not correspond to the genus *Pedinomonas* (the 18S rDNA sequence has not been obtained to confirm the morphology-based identification), (iv) the host is found in different oceanic regions without intracellular symbionts [7].

In general, photosymbioses are less specific than parasitic interactions, which can be a reason why we detected more parasitic interactions. For instance, the symbiont species can change depending on the oceanic region [8], thereby reducing co-variance. Not surprisingly, only 11% of the true positives are mutualistic interactions as compared to 28% of the 25 false negatives present in the matrices. All of the remaining true positive and false negative interactions are parasitic.

Macintosh HD:Users:u0097353:Documents:Documents_Karoline:BSB_Lab:Results:TARA-indirect-disentangling3:TruePositives:Round4:PlotsTargetedFN:603f2dca0a9af55fc13a5966c26016bc_4077dd0358107a8bb7d2213700329d83_taxonmatrix_30_32.txt.pdf

Figure 1. Abundance profiles of Collozoum and Scrippsiella, which have been reported to engage in photosymbiosis.

Macintosh HD:Users:u0097353:Documents:Documents_Karoline:BSB_Lab:Results:TARA-indirect-disentangling3:TruePositives:Round4:PlotsTargetedFN:eaadfd24b88f874ab847d9e130dd7e9e_717f93cd1ef051b7166839f93d345ef9_taxonmatrix_1.txt.pdf

Figure 2. Abundance profiles of Noctiluca and Pedinomonas, which have been reported to engage in photosymbiosis.

## 3. Discussion of putative false positive interactions

A part of the false positive edges represents indirect interactions. Although we filter indirect taxon edges that are likely due to environmental factors, not all environmental factors that may induce indirect taxon edges, e.g. trace elements, were measured.

Organisms can be affected indirectly not only by environmental factors, but also by other organisms. We decided not to attempt to filter indirect taxon edges driven by other taxa, for the following reason. Indirect edge detection implies the detection of causal relationships in the data. When environmental factors are involved, unraveling causality is easier, because we can make the assumption that taxon abundances are changed by environmental factors, but not the other way round (although organisms change their environment, the time scale is a different one). This assumption reduces the number of possible causal relationships that can occur between two taxa and an environmental factor: either the environmental factor directly modifies the taxon abundances, thereby introducing an indirect taxon edge, or the change in abundance of one taxon in response to the environmental factor affects the other taxon. We simplify by removing the taxon edge whenever the interaction information of a taxon-environment triplet indicates redundancy.

The causal dependencies among taxon edges are potentially far more complicated; no simplifying assumptions can be made. It is our opinion that current causality detection methods are not reliable enough to be employed without further support and we consequently did not filter indirect taxon edges that do not involve an environmental factor.

## 4. Acantharia and *Phaeocystis*

Despite the issues mentioned above inherent to photosymbiosis in plankton, we detected a number of positive correlations between barcode pairs that correspond to mutualistic symbioses in plankton, such as the one between Acantharia (Radiolaria) and the endosymbiotic microalga *Phaeocystis* (haptophyte). Photosymbiotic relationships between members of Acantharia and *Phaeocystis* are widely distributed and common in the surface oligotrophic waters [8]. In particular, we found an interaction between a V9 barcode of *Phaeocystis* and a V9 barcode of Acantharia, which has been assigned at 97% to the genus *Hexaconus.* Additional phylogenetic analyses showed that the V9 of Acanthariabelongs to clade E (all acantharian species, including *Hexaconus* sp. from this clade live in symbiosis with *Phaeocystis*), confirming that this is a symbiotic acantharia (Figure 3),

Macintosh HD:Users:u0097353:Documents:Documents_Karoline:Publications:TARA-paper:Revision_round2:Acantharea_Phaeocystis:V9_Acantharia_hexaconus_100215-BioNJ_tree.pdf

Figure 3. Distance tree (BioNJ [9]) with the V9 sequences of Acantharia (including the V9 of the pair, named "query" that groups within the clade E, close to subclade E3).

## Literature

1. Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, et al.: **Determinants of community structure in the global plankton interactome**. *Science* 2015.

2. de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E: **Eukaryotic plankton diversity in the sunlit ocean**. *Science* 2014, **Submitted**.

3. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G: **Structure and function of the global ocean microbiome.** *Science* 2014, **Submitted**.

4. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A: **Global patterns and ecological drivers of ocean viral communities in the Tara Oceans Virome dataset**. *Science* 2014, **submitted**.

5. Pesant S, Not F, Picheral M, Kandels‐Lewis S, Le Bescot N, Sciandra A: **Open science resources for the discovery and analysis of Tara Oceans Data**. *Sci. Data* 2015, **2**.

6. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C: **Microbial co-occurrence relationships in the human microbiome**. *PLoS Comput Biol* 2012, **8**:e1002606.

7. Brown MB: **A Method for Combining Non-Independent, One-Sided Tests of Significance**. *Biometrics* 1975, **31**:987-992

8. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing**. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995, **57**:289-300.