**Beta diversity analyses**

* Removed samples with low number of ASVs and low counts of reads (92 samples used)
* Estimated the SPARCC correlations for the whole data set, then subset the resulting correlation matrix to the 161 *Planctomycetes* ASVs. Used the correlation matrix and the count tables to estimate PINA/TINA (Schmidt, Matias Rodrigues, and von Mering 2017). Only the TINA results are shown.
* Did a basic PCoA using the TINA index, and interpolated (Inverse distance weighted (Bailey and Gatrell 1995)) the PCoA PC1 in a map to make the data interpretation more intuitive in addition to the classical ordination plot.
* As already seen in the OSD2014 manuscript I am preparing, MEOW provinces together with the temperature are key for *Planctomycetes* as well
* Permanova and in Figure 1B and 1D

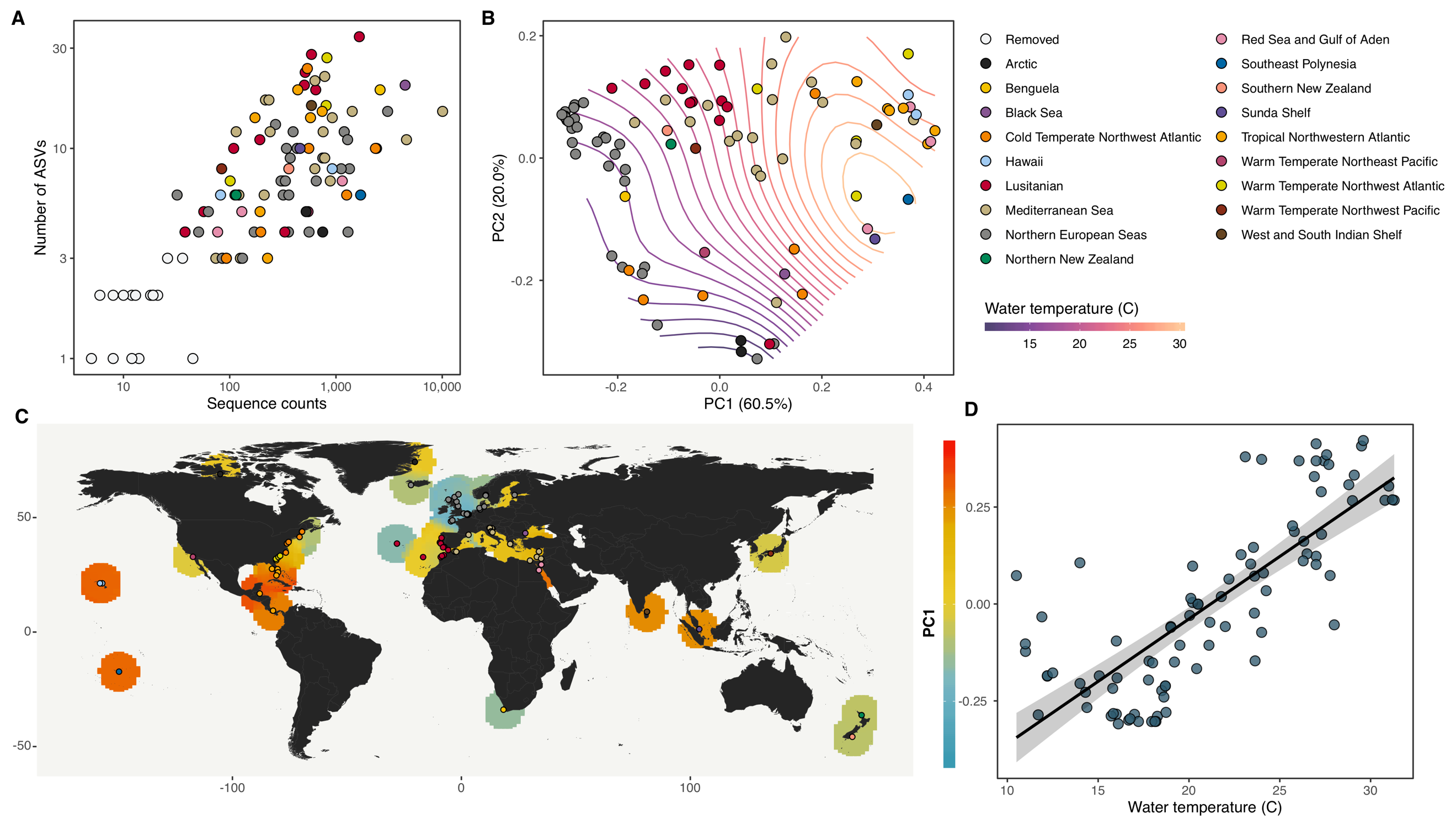


Figure 1 A) Samples used for the analyses. B) PCoA using the weighted TINA index index. C) Interpolation of the PCoA PC1 in a map, Similar colors indicate Planctomycetes communities that are more similar in overall community composition. C) Relationship between community similarity and temperature (r = 0.62; p < 0.001. Removed two arctic samples with outlier temperatures)

Table 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PERMANOVA | | | BETADISPER | |
|  | R2 | F | Pr(>F) | F | Pr(>F) |
| MEOW province | 0.666 | 8.681 | **0.001** | 1.604 | 0.191 |
| MEOW realm | 0.354 | 5.693 | **0.001** | 5.183 | **0.001** |
| Temperature | 0.498 | 21.659 | **0.001** | 1.761 | 0.129 |
| Longhurst code | 0.615 | 6.961 | **0.001** | 1.558 | 0.183 |
| Longhurst biome | 0.204 | 7.540 | **0.001** | 2.537 | 0.066 |

**Graph based analyses**

* Used the graph we inferred for the main OSD2014 paper. This graph was inferred with SPARCC (Friedman and Alm 2012) and filtered using a similar approach like we did in Chafee et al. 2018 and in Žure et al. 2017. Briefly, the main idea is to maintain the connectivity of the graph with the least number of edges. To accomplish this, we do a two-step process where the edges with small correlation values (we use the ) of the graph are iteratively removed until the graph gets disconnected; we repeat the same procedure using the p-values, iteratively removing large p-values until the graph gets disconnected (more than one component). On the trimmed graph we identify communities using the Louvain algorithm (Blondel, Guillaume, and Lambiotte 2008). Due to the non-deterministic nature of the clustering algorithm it was repeated 1000 times, and the membership of each node was determined by a majority voting approach. The Louvain communities in the whole network have a strong association to the MEOW ecoregions and temperature.
* Created first order ego networks centered on the 161 *Planctomycetes* ASVs. In Figure 2 you have the distribution of the edge weights between Planctomycetes and other phyla. In each case, but in the *Planctomycetes* ego network, we removed all edges between *Planctomycetes* ASVs and only kept the cross-phyla associations. Figure 2 shows the SPARCC correlation value distribution for each phylum.
* First exploration of the associations in each phylum, looking at the proportion of each phylum ASV in each sample. Samples in Figure 3 (only showing the top 10 phyla, at the end of the document there are all the rest) are sorted by the PCoA PC1 and PC2. Bar plot colors are the community membership of each ASVs. The nodes in the graphs in Figure 3 are the result of collapsing all nodes based on phylum and graph community information. The edges are the median SPARCC correlations of all the nodes that have been collapsed to their neighbors.
* Plots in Figure 3 show to which ASVs the *Planctomycetes* ASVs tends to be associated. You can see how certain patterns emerge in terms of the associations, suggesting that the associations are region specific and temperature dependent. In other words, *Planctomycetes* associations to other phyla are very specific depending where they are located. If you explore the graphs combination in Figure 3, you will see how the Phyla from the different graph communities show positive or negative SPARCC correlations in function of their biogeography, where temperature and MEOW province is a very important factor, not only on the Planctomycetes community structure but on their potential associations to other organisms. There are detailed tables in the folder I shared for a more in-depth exploration.

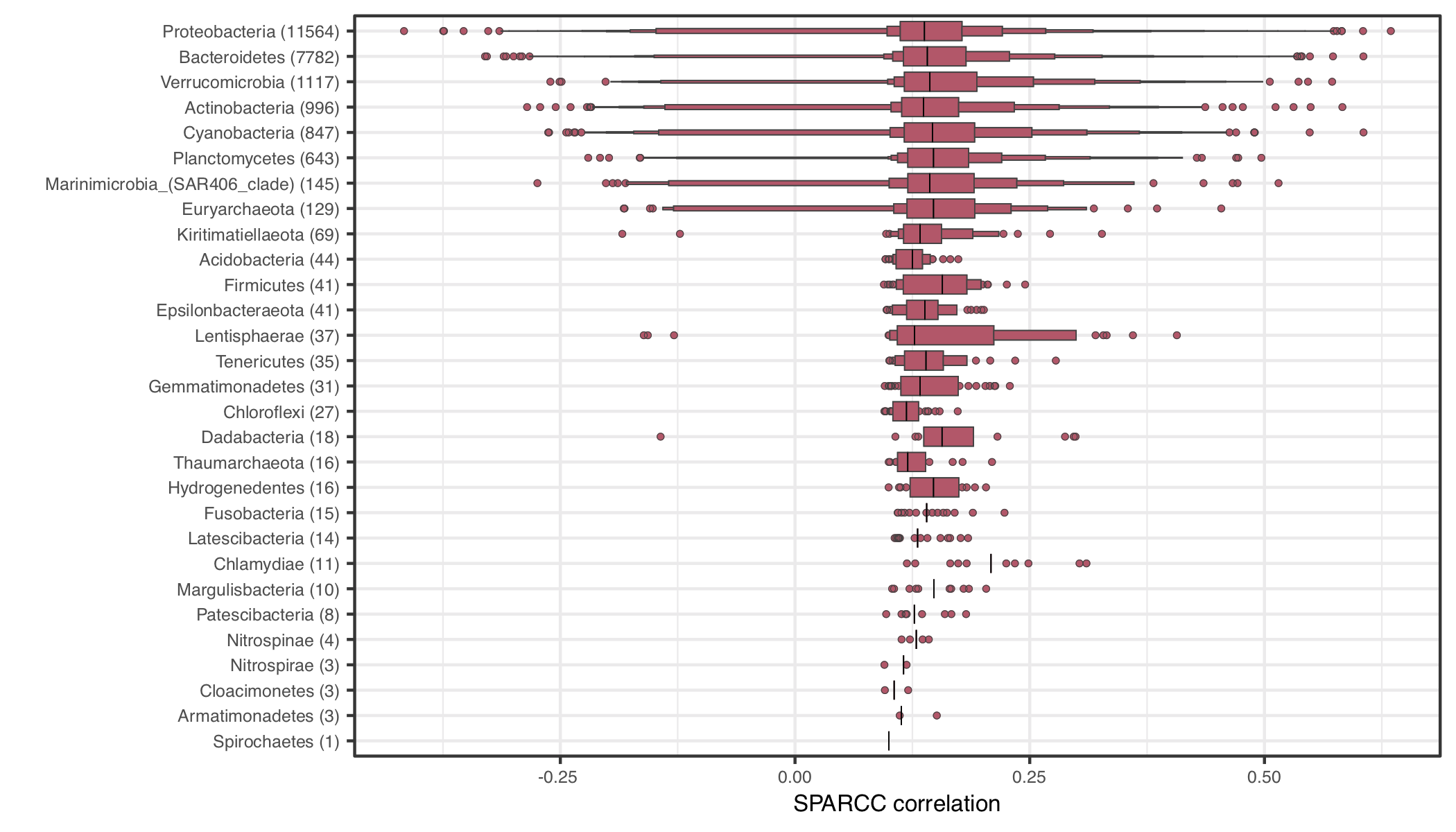


Figure 2 Letter-value plot showing the distribution of the SPARCC correlation values for phyla associated to the Planctomycetes ASVs. In parenthesis the number of edges in between Planctomycetes and the phylum ASVs

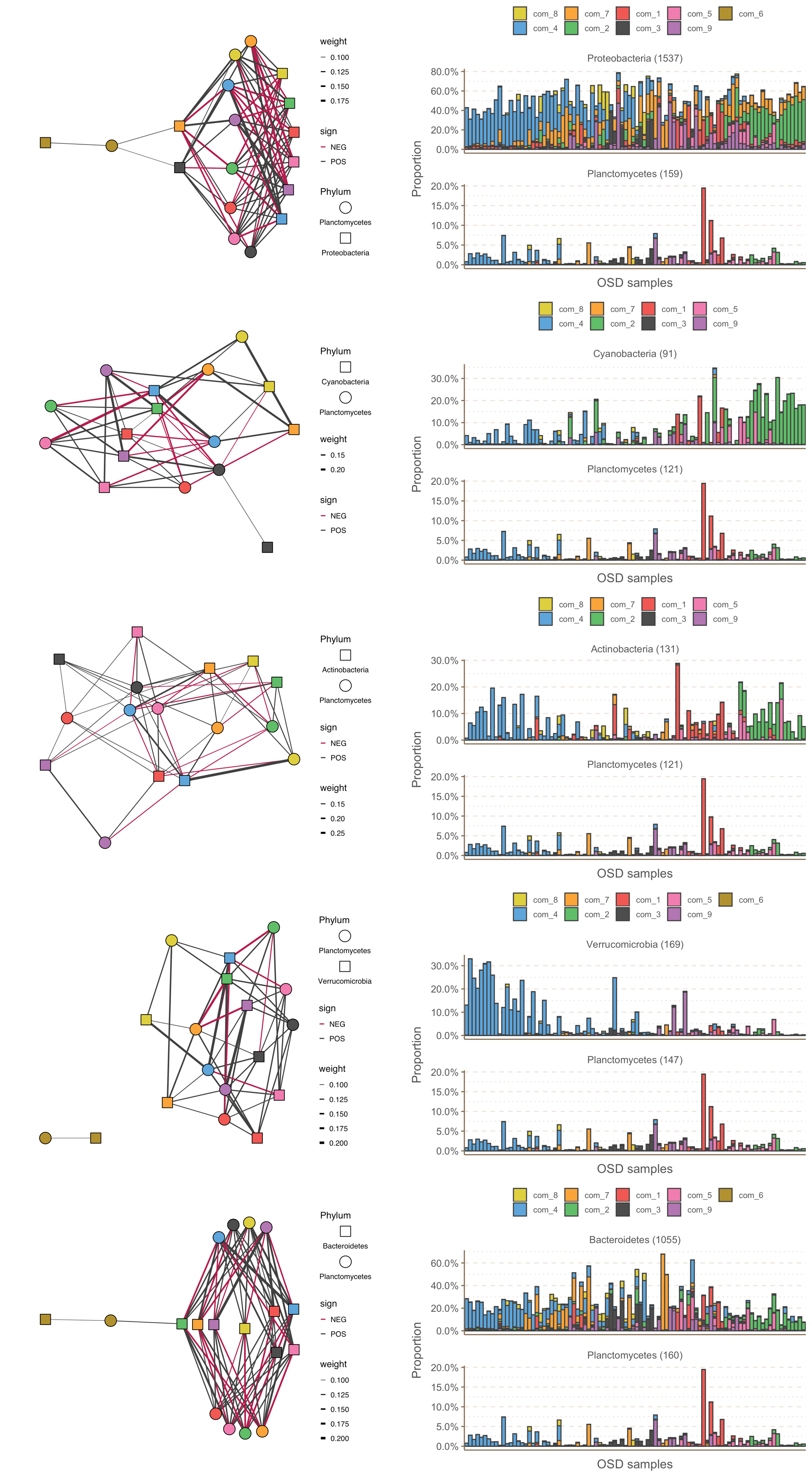


Figure 3 Ego networks of the 5 phyla with more associations with the Planctomycetes ASVs. Node colour correspond to the Louvain communities. Edge colours represent the sign of the correlation and the width the strength.

Bailey, Trevor C., and Anthony C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Vol. 413. Longman Scientific & Technical Essex.

Blondel, V. D., J. L. Guillaume, and R. Lambiotte. 2008. “Fast Unfolding of Communities in Large Networks.” *Journal of Statistical*. http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta.

Chafee, Meghan, Antonio Fernàndez-Guerra, Pier Luigi Buttigieg, Gunnar Gerdts, A. Murat Eren, Hanno Teeling, and Rudolf I. Amann. 2018. “Recurrent Patterns of Microdiversity in a Temperate Coastal Marine Environment.” *The ISME Journal* 12 (1): 237–52.

Friedman, Jonathan, and Eric J. Alm. 2012. “Inferring Correlation Networks from Genomic Survey Data.” *PLoS Computational Biology* 8 (9): e1002687.

Schmidt, Thomas Sebastian Benedikt, João Frederico Matias Rodrigues, and Christian von Mering. 2017. “A Family of Interaction-Adjusted Indices of Community Similarity.” *The ISME Journal* 11 (3): 791–807.

Žure, Marina, Antonio Fernandez-Guerra, Colin B. Munn, and Jens Harder. 2017. “Geographic Distribution at Subspecies Resolution Level: Closely Related Rhodopirellula Species in European Coastal Sediments.” *The ISME Journal* 11 (2): 478–89.

