Environmental Structuring of Mixoplankton Functional Types within Marine Protist Communities: A Global Analysis

Michaela E. Larsson[1*], Suzana G. Leles[2,3*], Aditee Mitra[4,5], Emile Faure[6], Daniel Vaulot[6,7], Luciana Santoferrera[8#]

*Equal contribution; #Corresponding author

[1]Aquatic Science Branch, Department of Water and Environmental Regulation, Western Australia, Australia. ORCID 0000-0003-4310-4582.

[2]Department of Marine Science, University of Texas at Austin, Port Aransas, TX, USA. ORCID 0000-0002-7819-5851.

[3]Department of Marine and Environmental Biology, University of Southern California, Los Angeles, CA, USA.

[4]School of Earth and Environmental Sciences, Cardiff University, Cardiff, UK. ORCID 0000-0001-5572-9331.

[5]ARK Dynamics, Walles, UK.

[6]Station Biologique UMR7144, Sorbonne Université and CNRS, Roscoff, France.

[7]Department of Biosciences, University of Oslo, Oslo, Norway. ORCID 0000-0002-0717-5685.

[8]Department of Biology, Hofstra University, Hempstead, NY, USA. ORCID 0000-0001-5587-2111.

Supplementary File 1

Content:

Table S1. Studies included in our V4 and V9 datasets, as downloaded from the metaPR^2 database (https://app.metapr2.org/).

Study	Marker	Ocean Region	Ecosystem	No. of Samples	No. of ASVs
Ocean Sampling Day - 2014 - V4 LGC	V4	Ocean survey	coastal	154	6816
Ocean Sampling Day - 2015 - V4	V4	Ocean survey	coastal	138	6250
Ocean Sampling Day - 2014 - V4 LW - HC	V4	Ocean survey	coastal	29	6109
Arctic - Beaufort Sea - MALINA cruise - 2009	V4	Arctic Ocean	oceanic	24	270
Arctic - Central Arctic Ocean - 2012	V4	Arctic Ocean	oceanic	8	182
Arctic - Nansen Basin - 2012	V4	Arctic Ocean	oceanic	17	311
Antarctic - Fildes Bay - 2013	V4	Southern Ocean	coastal	10	69
Antarctic - Fildes Bay - 2015	V4	Southern Ocean	coastal	123	729
Antarctic - Fildes Bay sorted - 2015	V4	Southern Ocean	coastal	60	281
Baltic Sea - Gulf of Finland - 2012-2013	V4	Baltic Sea	coastal	73	935
Atlantic - Oslo fjord - Norway - 2009-2011	V4	Atlantic Ocean	coastal	78	834
Roscoff - France - Time Series - 2009-2011	V4	English Channel	coastal	138	2485
Roscoff - France - Time Series - 2012-2016	V4	English Channel	coastal	225	3216
Malaspina - depth profiles - 2010-2011	V4	Ocean survey	oceanic	179	6073
Malaspina - surface - 2010-2011	V4	Ocean survey	oceanic	119	7357
Blanes - Spain - Time Series - 2004-2013	V4	Mediterranean Sea	coastal	288	7698
Arctic - Baffin Bay - 2013	V4	Arctic Ocean	oceanic	32	587
Arctic - White Sea - 2013-2015	V4	Arctic Ocean	oceanic	17	387
Arctic - ARK-XXVII/3 cruise - 2012	V4	Arctic Ocean	oceanic	45	996
Arctic Ocean - Survey - 2005-2011	V4	Arctic Ocean	oceanic	36	467
Arctic - Chukchi Sea - ICESCAPE - 2010	V4	Arctic Ocean	oceanic	23	257
Arctic - Nares Strait - 2014	V4	Arctic Ocean	oceanic	247	1723
Baltic Sea - Gdansk Gulf - 2012	V4	Baltic Sea	coastal	35	272
Baltic Sea - Gdansk Gulf - 2012 Hapto	V4	Baltic Sea	coastal	12	93
Bay of Naples - Italy - Time Series - 2011	V4	Mediterranean Sea	coastal	8	2334
Biomarks project - 2009	V4	coast of Europe	coastal	120	1156
Pacific - Mariana Trench - 2016 - 1	V4	Mariana Trench	oceanic	32	2804
Pacific - Mariana Trench - 2016 - 2	V4	Mariana Trench	oceanic	12	213
Arctic Ocean - 2012	V4	Arctic Ocean	oceanic	22	714
Amundsen Sea ASPIRE cruise - 2010-2011	V4	Southern Ocean, Amundsen Sea	oceanic	39	336
Arctic - Fram Strait - 2014	V4	Arctic Ocean, Fram Strait	oceanic	33	1567
Antarctic - Palmer Station LTER - 2014	V4	Western Antarctic Peninsula	oceanic	21	928
Tara Ocean - V4	V4	Ocean survey	oceanic	104	9515
Tara Arctic - V4	V4	Arctic Ocean	oceanic	28	1462
Arctic and Scotian Shelf - 2009-2011	V4	Arctic Ocean; Nova Scotia	oceanic	59	495
Arctic - Baffin Bay - 2008-2018	V4	Arctic Ocean, Baffin Bay	oceanic	79	1281
Southern Ocean - 2017	V4	Southern Ocean	oceanic	13	2333
Arctic - Fram observatory 2016	V4	Arctic Ocean, East Greenland	oceanic	96	3833
Antarctic Peninsula - 2012-2016	V4	Southern Ocean, WAP	oceanic	119	2951
Atlantic transect - cruise PS113 - 2018	V4	Atlantic Ocean	oceanic	129	6187
Pacific South West - NIWA - 2009-2018	V4	South West Pacific Ocean	oceanic	529	6751
Tara Ocean - V9	V9	Ocean survey	oceanic	1152	30016

Table S2. Number of samples and sampling locations included in each dataset, categorized by sequenced marker and seawater filtration strategy. Based on the strategies used in each study for seawater filtration, and considering if size fractionation was used, and which filter pore sizes were utilized, the metaPR2 database classifies samples into 'pico' (0.2-2 μm), 'pico-nano' (0.2-20 μm), 'nano' (2-20 μm), 'nano-micro' (2-200 μm), 'micro' (20-200 μm), and 'meso' (>200 μm) plankton size fractions; an additional 'total' category represents samples that were not fractionated.

Filtration Strategy	Number of samples		Number of sampling locations	
	V4	V9	V4	V9
Total	1376	150	672	102
Pico	899	227	221	169
Pico-nano	152	8	35	8
Nano	339	210	32	159
Nano-micro	299	18	36	16
Micro	64	221	7	168
Meso	0	227	0	175

Figure S1. Overview of the analytical workflow used in this study.

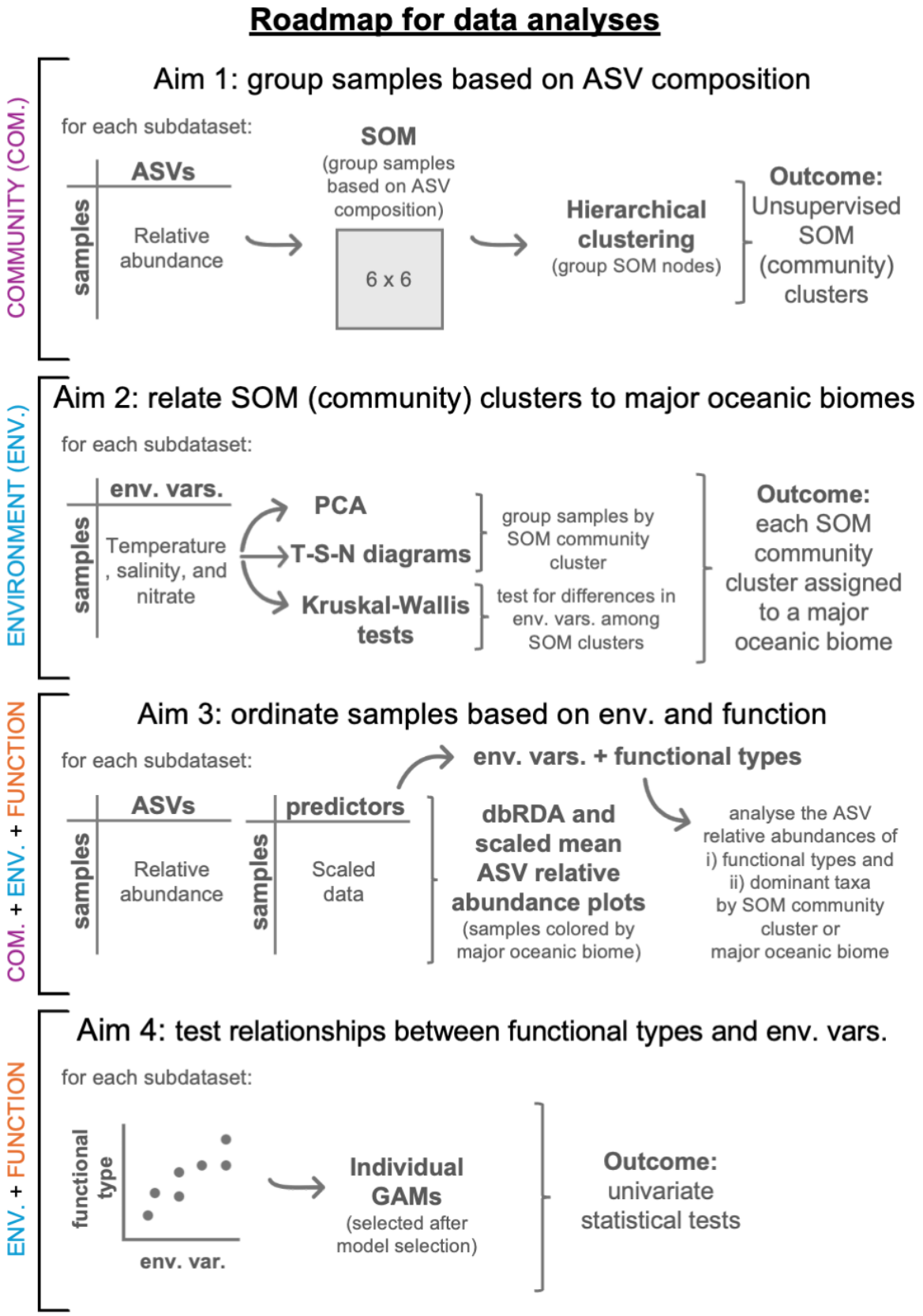

Figure S2. Mixoplankton were recovered in most metabarcoding studies regardless of the strategy used for seawater filtration, with relatively low ASV richness (A-B) but substantial relative abundance (C-D) compared to other functional types. Data are categorised based on filtration strategy (total or fractionated using filters of different pore sizes) as classified in the metaPR^2 database (see Methods for more details). The number of samples and sample locations included in each dataset are shown in Table S2. Filtration strategies are also detailed in Table S2.

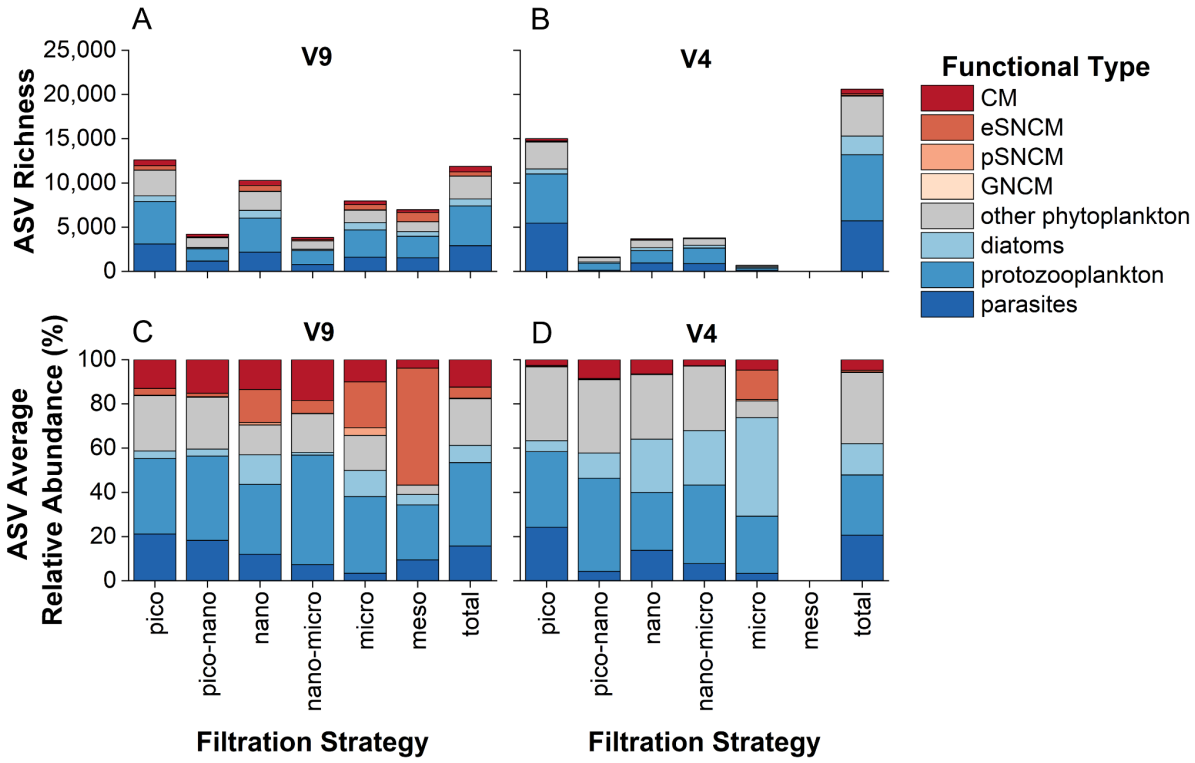

Figure S3. Spatial distribution of emergent protist communities from the unsupervised self-organizing map (SOM) analyses. Each subdataset yielded five independent SOM community clusters (as shown by colours) based on ASV composition. **A** Analysis based on the V9-micro subdataset (7,960 ASVs across 221 samples). The variability explained was 75%. The number of samples per community cluster is as follows: cluster 1 (n = 40), cluster 2 (n = 19), cluster 3 (n = 56), cluster 4 (n = 41), and cluster 5 (n = 65). **B** Analysis based on the V9-total subdataset (n = 11,867 ASVs across 150 samples). The variability explained was 78%. The number of samples per community cluster is as follows: cluster 6 = 43, cluster 7 = 11, cluster 8 = 38, cluster 9 = 29, and cluster 10 = 29. **C** Analysis based on the V4-total dataset (20,595 ASVs across 1,376 samples). The variability explained was 75%. The number of samples per community cluster is as follows: cluster 11 = 693, cluster 12 = 328, cluster 13 = 145, cluster 14 = 134, and cluster 15 = 76). Note that each SOM cluster was then categorized according to one of four major oceanic biomes (see Methods and Figs. 2 and 3), as follows: polar = clusters 12 and 15; subpolar/temperate = clusters 5, 8 and 11; temperate/subtropical = clusters 1, 3, 6, 7 and 13; subtropical/tropical = clusters 2, 4, 9, 10 and 14.

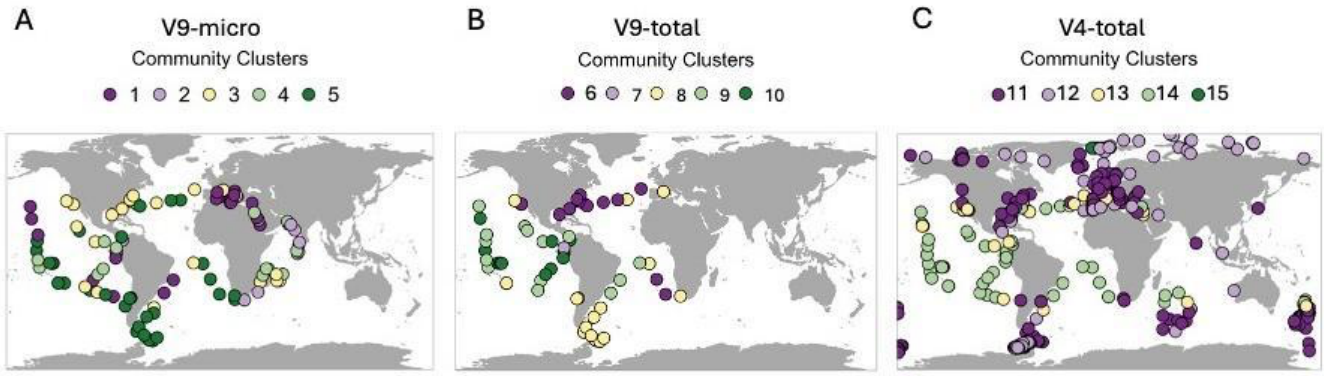

Figure S4. Environmental characterization of the emergent SOM community clusters shown in Fig. S2. Black dots indicate the median and colors indicate numeric density and distribution. **A** For the V9-micro subdataset, Kruskal-Wallis tests showed significant differences ($p < 0.0001$) in temperature ($\chi^2 = 64$), salinity ($\chi^2 = 38$), and nitrate concentrations ($\chi^2 = 35$) among the five community clusters. Dunn's tests indicated significant differences ($p < 0.05$) between clusters for temperature (all comparisons, except for cluster 1 versus 3 and 5, and cluster 2 versus 4), nitrate (cluster 4 versus 1, 2 and 3, and cluster 5 versus 1 and 2) and salinity (only cluster 1 versus 3, 4 and 5). **B** For the V9-total subdataset, Kruskal-Wallis tests showed significant differences ($p < 0.001$) in temperature ($\chi^2 = 76$) and nitrate concentrations ($\chi^2 = 20$) among the five community clusters, but not for salinity ($\chi^2 = 6$, $p = 0.16$). Dunn's tests indicated significant differences ($p < 0.05$) between clusters for temperature (all comparisons, except for cluster 6 versus 7 and 8, cluster 7 versus 8, 9 and 10, and cluster 9 versus 10) and nitrate (only between cluster 9 versus 10). **C** For the V4-total subdataset, Kruskal-Wallis tests showed significant differences ($p < 0.0001$) in temperature ($\chi^2 = 521$), salinity ($\chi^2 = 395$), and nitrate concentrations ($\chi^2 = 328$) among the five community clusters. Dunn's tests indicated significant differences ($p < 0.05$) between clusters for temperature (all comparisons, except for cluster 12 versus 15), salinity and nitrate (all comparisons, except for cluster 11 versus 15, and cluster 13 versus 14 for both environmental variables).

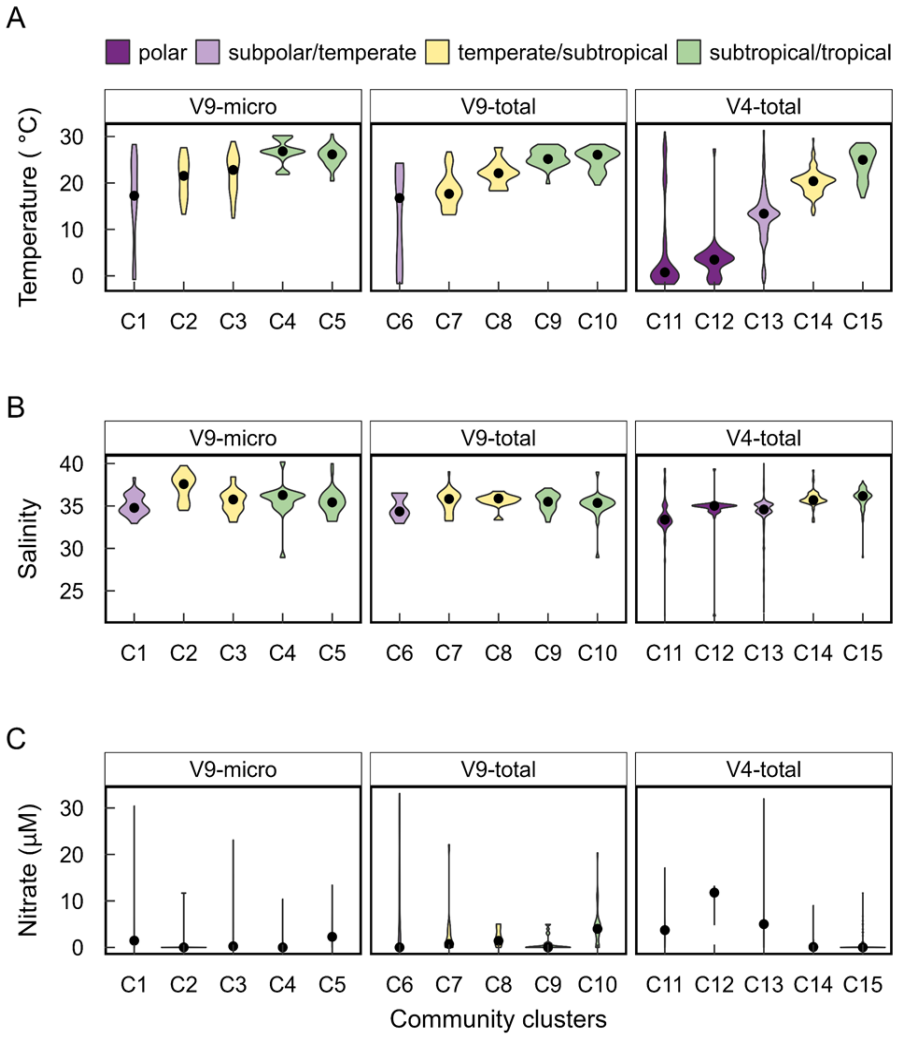

Figure S5. Scaled mean ASV relative abundance of each functional type indicating over- (positive values) or under- (negative values) representation within the SOM community clusters shown in Fig. S2. **A** V9-micro, **B** V9-total and **C** V4-total subdatasets. Note that community cluster colors do not match Fig. S2, but instead represents the major oceanic biomes shown in Fig. 2A.

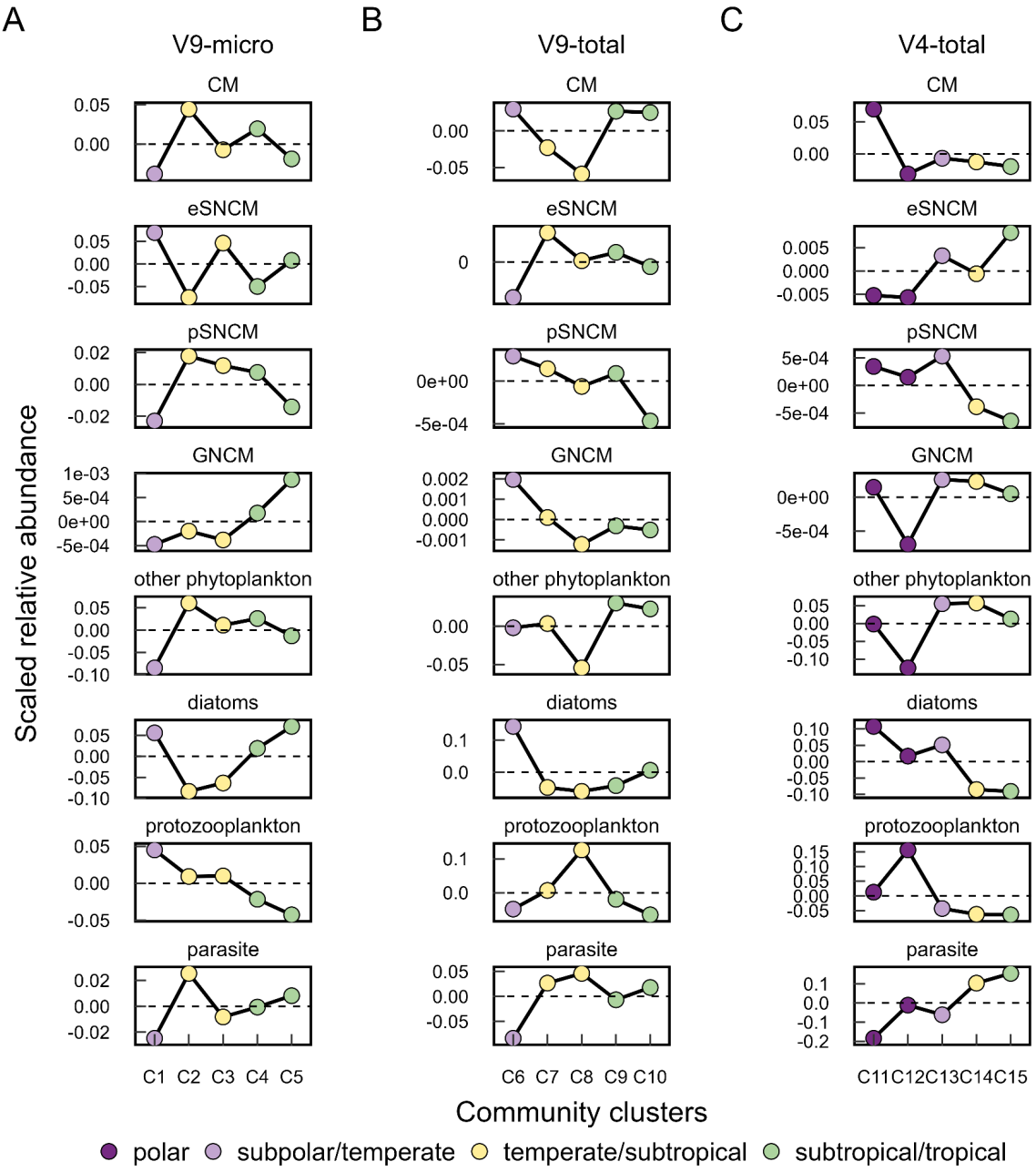

Figure S6. Relationships between functional types and environmental variables. **A** V9-micro, **B** V9-total and **C** V4-total subdatasets. Solid black lines with a 95% confidence interval represent general additive model fits. For V9-micro, all fits were significant ($p < 0.05$), except for GNCM versus temperature, GNCM, protozooplankton versus salinity, and protozooplankton versus nitrate. For V9-total, all fits were significant ($p < 0.05$), except for CM and protozooplankton versus temperature, GNCM and other phytoplankton versus salinity, and protozooplankton versus nitrate. For V4-total, all fits were significant ($p < 0.05$), except for pSNCM versus nitrate.

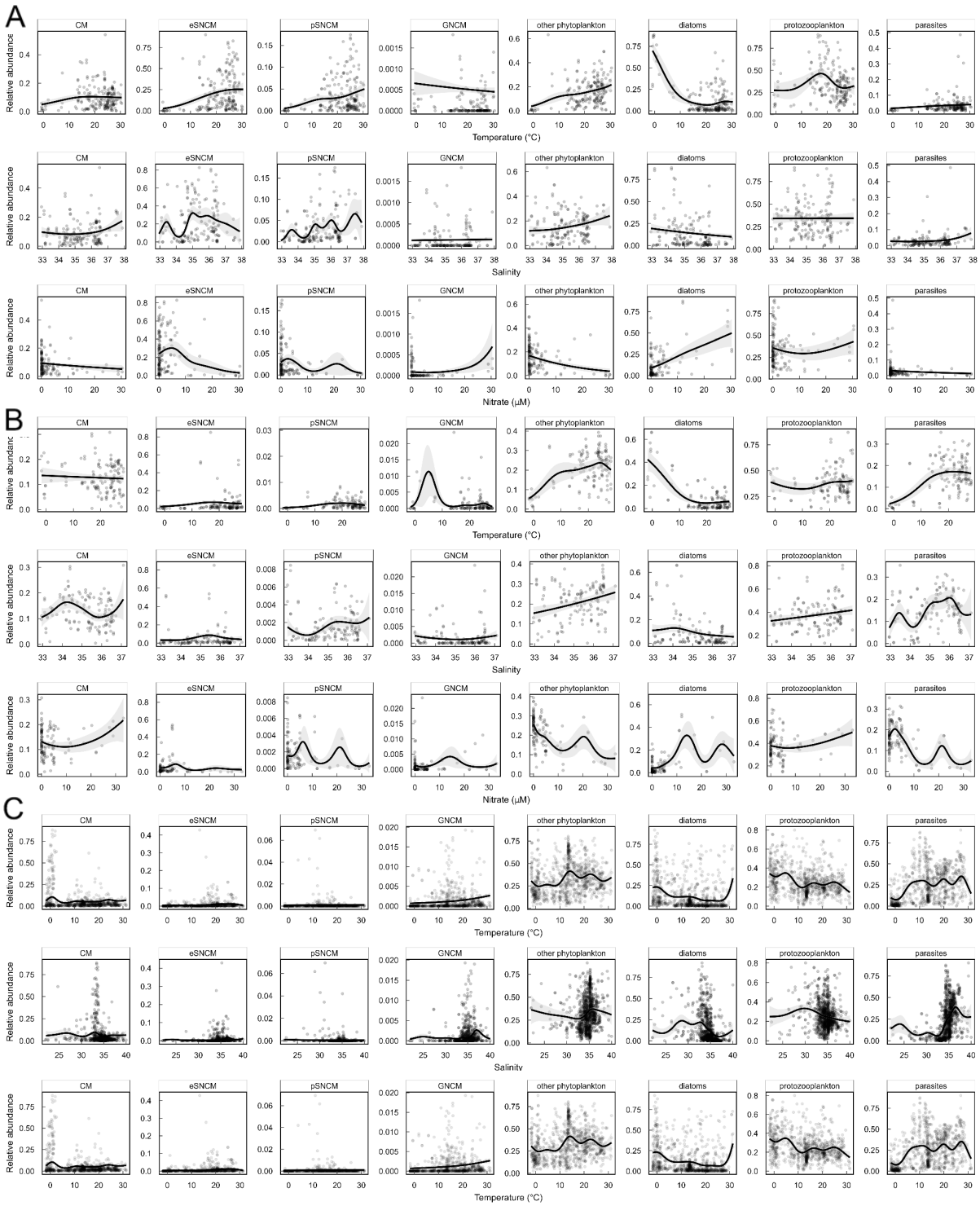

Figure S7. Principal component analysis (PCA) based on ASV composition was used to compare the communities across the three subdatasets. **A** V9-micro, V9-total and V4-total datasets. **B** V9-micro and V9-total only.