

Oceanographic structure drives the assembly processes of microbial eukaryotic communities

Adam Monier, Jérôme Comte, Marcel Babin, Alexandre Forest, Atsushi Matsuoka, and Connie Lovejoy

Supplementary Information (SI)

SI Materials and Methods

Seawater sampling and oceanographic data analyses.

SI references.

SI Figures

Titles and legends to SI figures.

Figure S1: Temperature and beam attenuation profiles.

Figure S2: Unweighted Unifrac and OTU abundances.

Figure S3: Phylogenetic diversity.

Figure S4: Relative abundance heatmap.

Figure S5: Phylogenetic placements of Picozoa OTUs.

SI Tables

Table S1: Ancillary data.

Table S2: Sequence and diversity summary.

Table S3: Environmental variable statistical analyses.

Table S4: SCM/wSCM differentially abundant taxa.

SI Materials and Methods

Seawater sampling and oceanographic data analyses

Sampling took place aboard the Canadian research icebreaker CCGS *Amundsen* in August 2009, as part of the International Polar Year, project MALINA. At every station, a rosette system equipped with a conductivity-temperature-depth profiler (CTD, Seabird SBE-911+), a fluorometer (Seapoint), and a transmissometer (WetLabs C-Star) was deployed multiple times. Relative nitrate profiles were retrieved from the rosette mounted *in situ* ultraviolet spectrometer probe (ISUS; Satlantic). The rosette system was also equipped with a sensor (Biospherical Instruments) for downwelling photosynthetically active radiation (PAR; 400 to 700 nm, dynamic range from 1.4×10^{-5} to $0.5 \mu\text{E cm}^{-2} \text{ s}^{-1}$). A similar deck sensor measured surface PAR during the rosette deployments. The CTD data were calibrated and verified following the Unesco Technical Papers (Crease *et al.* 1988). Periodically, discrete water samples were collected directly from the Niskin bottles for salinity calibration using a Guildline Autosal salinometer. The fluorescence data from the fluorometer were calibrated against *in situ* chlorophyll *a* (Chl *a*) concentrations as described in Forest *et al.* (2013). The transmissivity signal was transformed into c_p using standard equations (Kirk 1994). Verified underwater PAR data were normalized to surface irradiance to compute the percentage of transmitted PAR. All CTD and sensor data were averaged over 1 m bins.

Nutrient samples were collected at 10 to 20 m intervals in the upper water column (Tremblay *et al.* 2008; Martin *et al.* 2010) directly from Niskin-type bottles. Nitrate [NO_3^-] and nitrite [NO_2^-] samples were dispensed into polyethylene tubes, immediately poisoned with mercuric chloride [$1 \mu\text{g ml}^{-1}$; (Kirkwood 1992)] and stored for subsequent laboratory analysis. [NO_3^-] and [NO_2^-] were measured using a Technicon AutoAnalyzer II following Tréguer and LeCorre (1975). Concentrations in the nanomolar range (detection limit of 3 nmoles l^{-1}) were obtained as in Raimbault *et al.* (1990). Ammonium concentrations [NH_4^+] were measured directly on board using the sensitive fluorescent method of Holmes *et al.* (1999) with a detection limit of 5 nmoles l^{-1} . The reproducibility of nutrient measurements between analyses was assessed using in-house standards compared with commercially available products (OSIL: <http://www.osil.co.uk/Products/SeawaterStandards>).

Cell abundances were determined using a flow cytometer FACS ARIA (Becton Dickinson, San Jose, CA) on-board the CCGS *Amundsen* as previously described for bacterial (Ortega-Retuerta *et al.*, 2012), pico- and nano-phytoplanktonic cells (Balzano *et al.*, 2012). Bacterial production was measured using ³H-leucine incorporation as in Ortega-Retuerta *et al.* (2012). All physical, biological and nutrient data are available at <http://malina.obs-vlfr.fr>.

SI References

- Balzano S, Gourvil P, Siano R, Chanoine M, Marie D, Lessard S, Sarno D, Vaultot D. (2012). Diversity of cultured photosynthetic flagellates in the northeast Pacific and Arctic Oceans in summer. *Biogeosciences* **11**:4553–4571.
- Crease J, Dauphinée TM, Grose PL, Lewis EL, Fofonoff NP, Plakhin EA, *et al.* (1988). The acquisition, calibration and analysis of CTD data. *UNESCO Tech. Pap. Mar. Sci* **54**:94.
- Forest A, Babin M, Stemmann L, Picheral M, Sampei M, Fortier L, *et al.* (2013). Ecosystem function and particle flux dynamics across the Mackenzie Shelf (Beaufort Sea, Arctic Ocean): an integrative analysis of spatial variability and biophysical forcings. *Biogeosciences* **10**:2833–2866.
- Harding T, Jungblut AD, Lovejoy C, Vincent WF. (2011). Microbes in high arctic snow and implications for the cold biosphere. *Appl Environ Microb* **77**:3234–3243.
- Holmes RM, Aminot A, K erouel R, Hooker BA, Peterson BJ. (1999). A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can J Fish Aquat Sci* **56**:1801–1808.
- Kirk JTO. (1994). Light and photosynthesis in aquatic ecosystems. Cambridge university press.
- Kirkwood DS. (1992). Stability of solutions of nutrient salts during storage. *Mar Chem* **38**:151–164.
- Martin J, Tremblay J , Gagnon J, Tremblay G, Lapoussi re A, Jose C, *et al.* (2010). Prevalence, structure and properties of subsurface chlorophyll maxima in Canadian Arctic waters. *Mar. Ecol. Prog. Ser.* **412**:69–84.
- Ortega-Retuerta E, Jeffrey WH, Babin MB, B langer S, Benner R, Marie D, Matsuoka A, Raimbault P, Joux F. (2012). Carbon fluxes in the Canadian Arctic: patterns and drivers of bacterial abundance, production and respiration on the Beaufort Sea margin. *Biogeosciences* **9**:3679–3692.

Raimbault P, Slawyk G, Coste B, Fry J. (1990). Feasibility of using an automated colorimetric procedure for the determination of seawater nitrate in the 0 to 100 nM range: Examples from field and culture. *Mar. Biol.* **104**:347–351.

Tremblay J-É, Simpson K, Martin J, Miller L, Gratton Y, Barber D, *et al.* (2008). Vertical stability and the annual dynamics of nutrients and chlorophyll fluorescence in the coastal, southeast Beaufort Sea. *J Geophys Res-Oceans* **113**.

Tréguer P, LeCorre P. (1975). Manuel d'analyses des sels nutritifs dans l'eau de mer (Utilisation de l'Autoanalyser II). *Université de Bretagne Occidentale, Laboratoire d'Océanographie chimique* **110**.

Titles and legends to SI figures

Figure S1. **A.** Temperature and **B.** beam attenuation (c_p) profiles. Measures were collected from Beaufort Sea CTD rosette casts from which microbial communities were retrieved. Distinct solid and dashed lines represent the six distinct sampling stations. Stations corresponding to weak SCM are colored in red.

Figure S2. **A.** Nonmetric multidimensional scaling (NMDS) ordination on unweighted UniFrac distances for the 24 Beaufort Sea microbial communities. Ellipses represent community clusters and 95% confidence intervals. Samples are represented by markers (depths) and numbers (stations). **B.** Procrustes rotations between NMDS ordinations computed on weighted and unweighted UniFrac distances (d_{w5000} and d_{w5000} , respectively). Arrows summarizing the rotations between ordinations are colored based on their corresponding community clusters. **C.** OTU rank abundance curves and, **D.** total sequence numbers for the 50 most abundant OTUs for each Beaufort Sea microbial community clusters.

Figure S3. **A.** Phylogenetic diversity (Faith's definition) measures for each Beaufort Sea microbial community clusters. Phylogenetic diversity was computed on datasets subsampled 1000 times at an even depth of 5000 sequences. Plotted are mean phylogenetic diversity and error bars are s.e.m. **B.** Relationship between sample depth and phylogenetic diversity. The line represents a least-squares linear regression. Dots are scaled based on corresponding OTU richness.

Figure S4. Relative abundance heatmap of the 25 most abundant taxa (genera, family or 'groups' depending on available taxonomic information) in the Beaufort Sea sequence datasets. Relative abundances were centered-scaled and corresponding z-score scale is displayed as a color gradient (green: abundant; red: rare). Taxonomic lineages of the 25 taxa are indicated by colored dots. Row dendrogram represents results from a hierarchical clustering of taxa relative abundances; columns were regrouped based on Beaufort Sea microbial community clusters.

Figure S5. Phylogenetic placements of the Picozoa OTUs. Representative sequences of Picozoa-like OTUs were mapped onto a Picozoa reference tree using RAxML evolutionary placement algorithm. The Picozoa maximum-likelihood reference tree was reconstructed based on an 18S rDNA gene sequence alignment. Picozoa reference sequences were clustered at 99% identity to remove redundancy prior to the phylogenetic reconstruction (number of sequences in a cluster is indicated in parentheses after the Genbank accession of the reference representative sequence). Bootstrap supports are represented by black dots when $\geq 70\%$. OTU identifiers are displayed in red; an asterisk after the OTU identifier denotes a phylogenetic placement with a likelihood ≥ 0.75 . Relative abundances of Picozoa OTUs across the Beaufort Sea community clusters are displayed on the right.

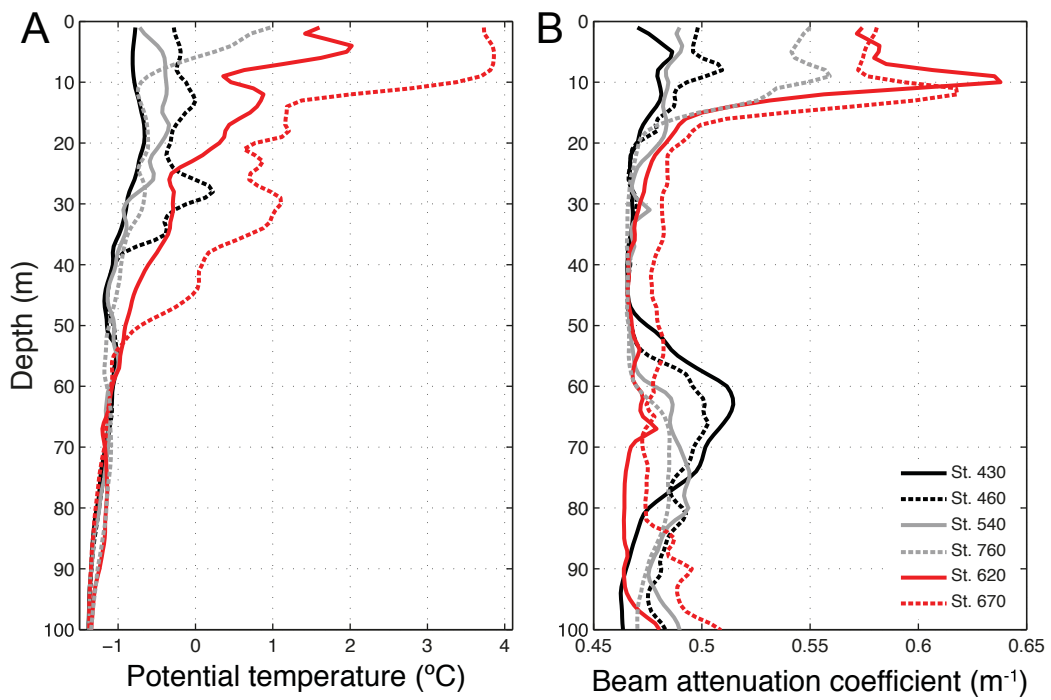


Figure S1

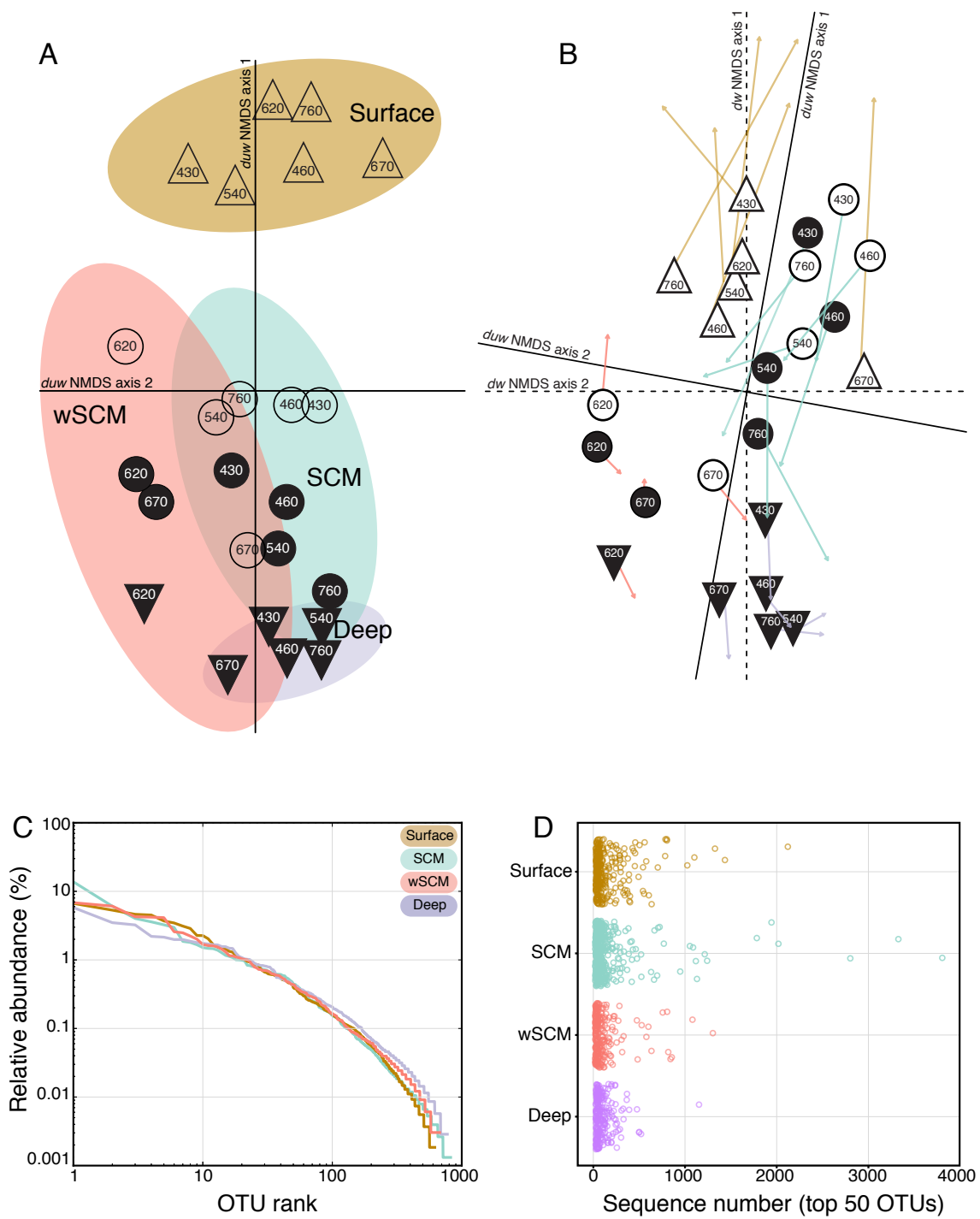


Figure S2

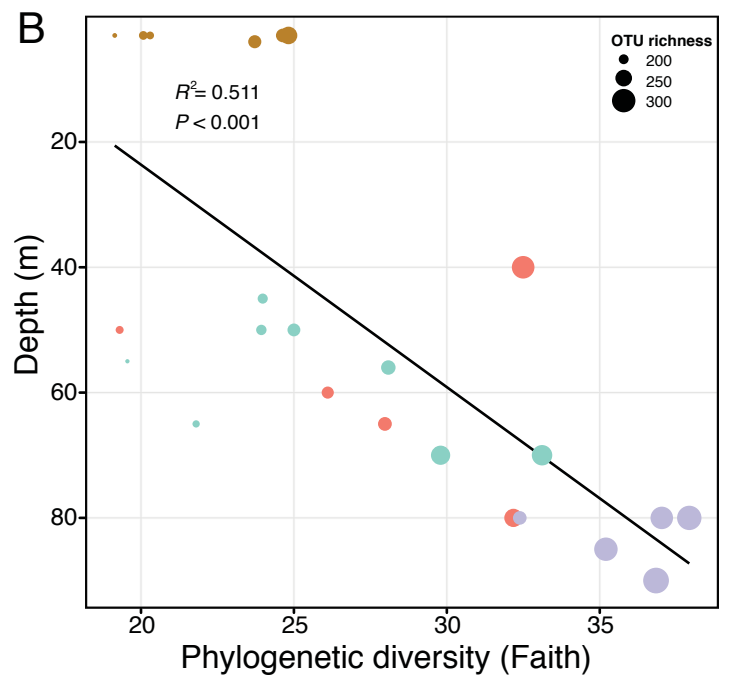
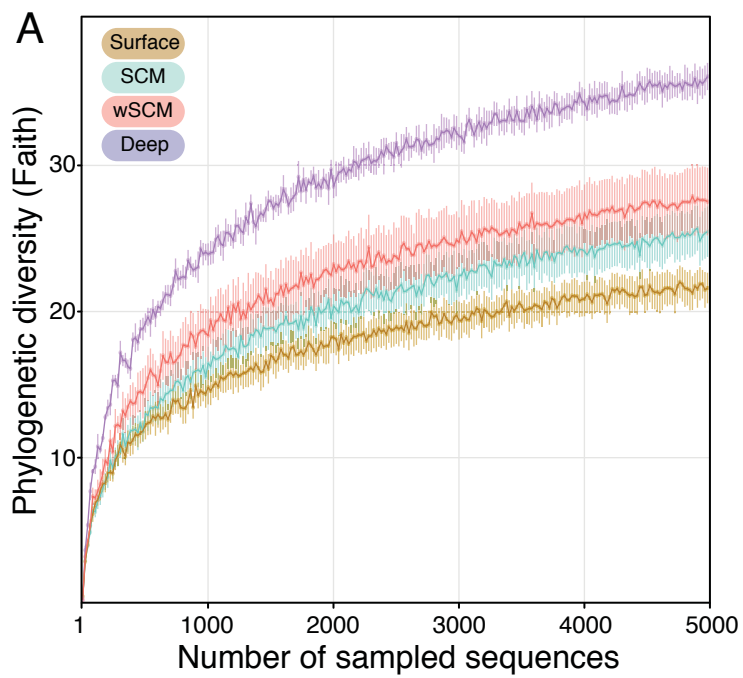


Figure S3

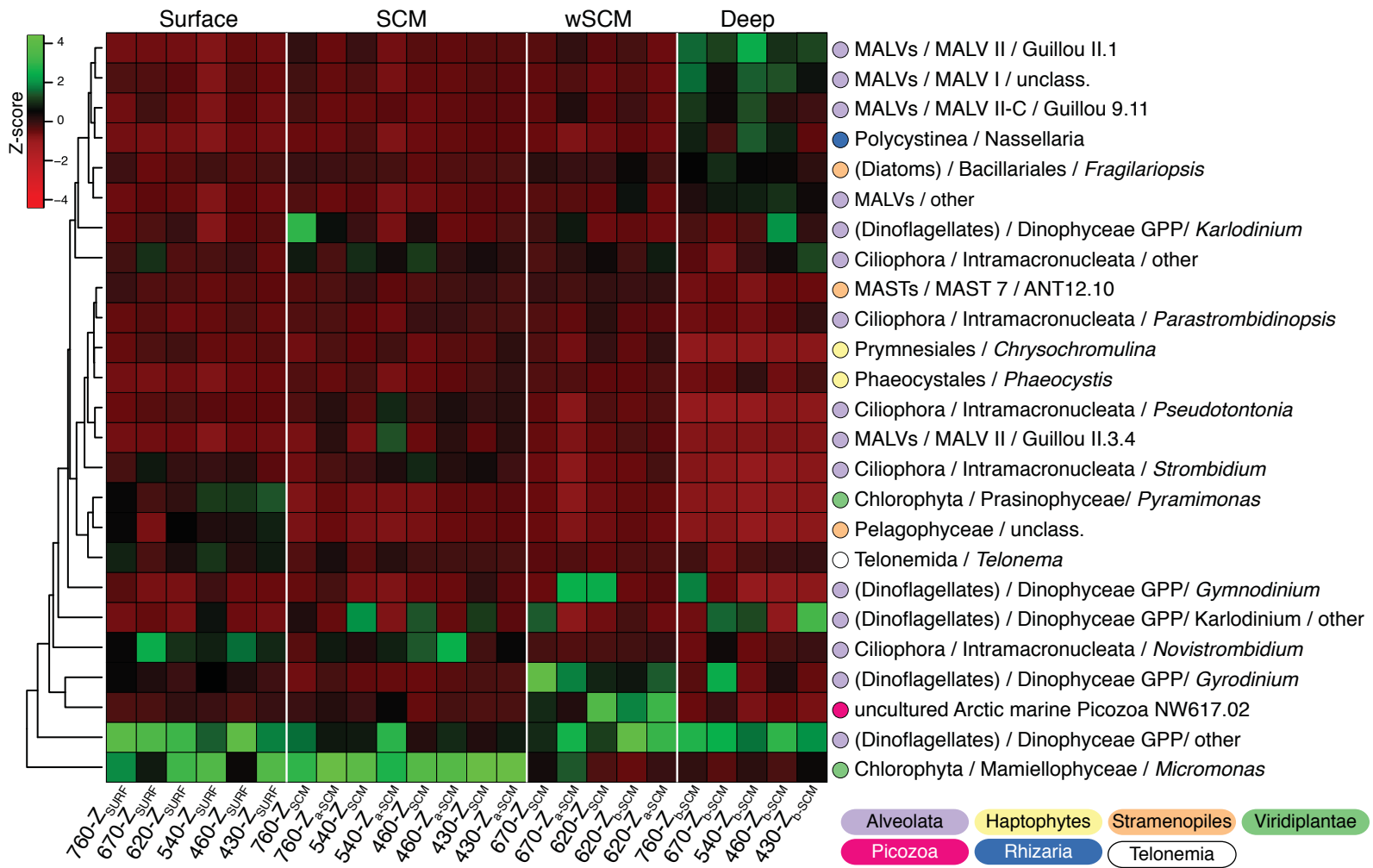


Figure S4

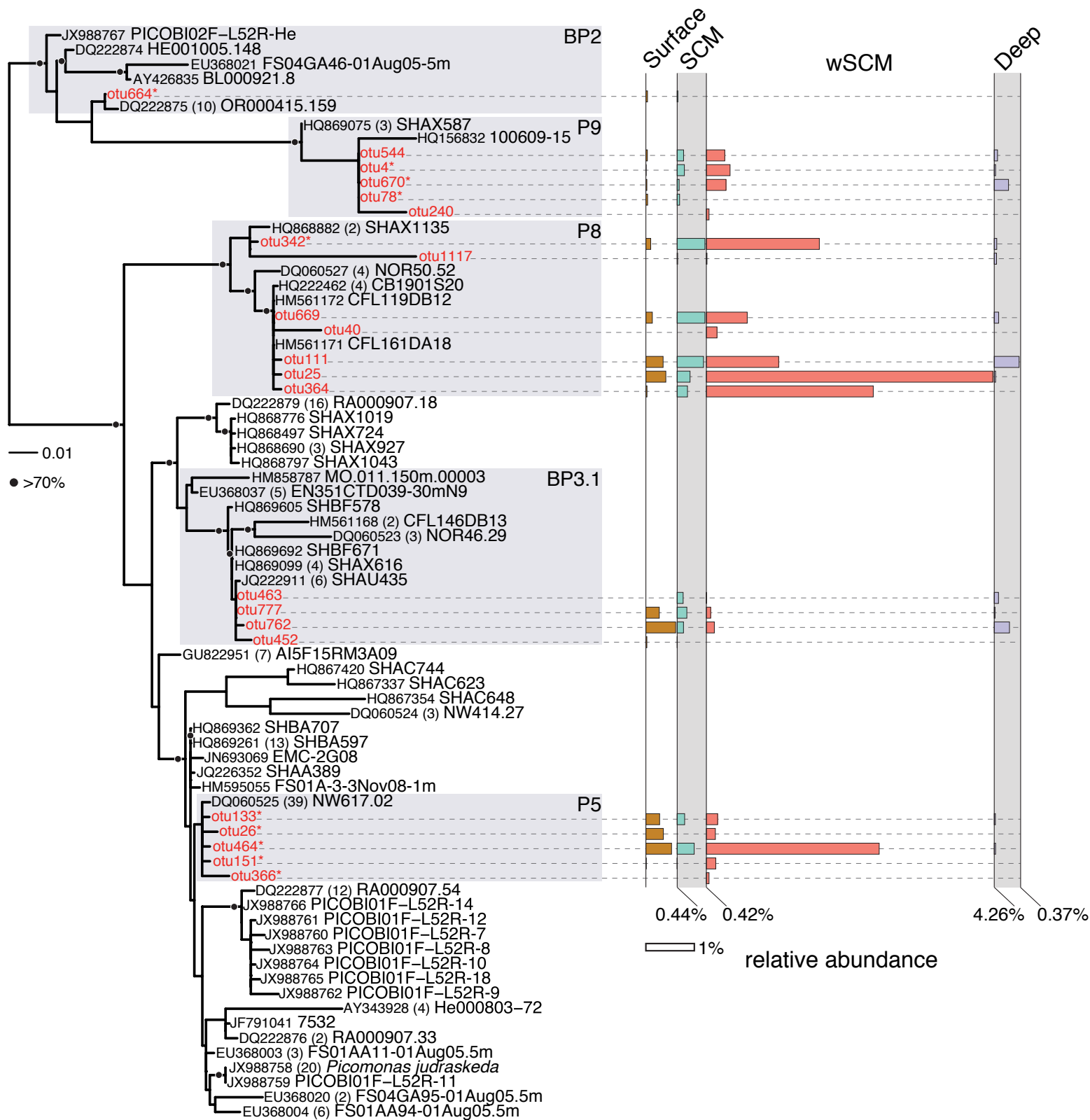


Figure S5

Table S1. Ancillary data for the Beaufort Sea sampling stations. The data, along with those presented in Figure 2, were used for BEST selection of environmental parameters. Additional ancillary data for the MALINA project are available at <http://malina.obs-vlfr.fr>.

Station	Date 2009	Z	Z (m)	O ₂ ^a	Phaeo small ^b	Phaeo large ^b	NO ₃ +NO ₂ ^c	NH ₄ ^c	PO ₄ ^c	SiOH ₄ ^c	POC ^c	DOC ^c	DON ^c	DOP ^c	PP ^d	BP ^e	Bacteria ^f	Nano ^f	Pico ^f	CDOM ^g
430	18/09	Z _{SURF}	3	8.63	0.01	0	0.01	0	0.52	2.89	1.76	60.99	5.57	0.19	0.52	3.07	238830	427	5171	0.07
		Z _{a-SCM}	55	8.88	0.33	0.05	2.93	0.02	0.99	8.99	3.11	n.a.	n.a.	n.a.	1.98	3.78	360052	1047	9854	0.08
		Z _{SCM}	65	8.21	0.37	0.04	6.80	0	1.30	14.52	2.38	65.76	6.16	0.10	0.45	3.06	447085	831	13057	0.10
		Z _{b-SCM}	80	7.22	0.03	0.02	12.89	0	1.79	27.88	n.a.	n.a.	n.a.	n.a.	n.a.	0.74	169411	682	974	0.10
460	19/08	Z _{SURF}	4	8.57	0.02	0	0.01	0	0.54	3.04	2.82	57.80	10.14	0.20	0.93	8.70	n.a.	n.a.	n.a.	0.08
		Z _{a-SCM}	45	9.26	0.03	0	0.02	0	0.75	4.04	1.18	59.44	12.24	0.22	0.42	3.85	n.a.	n.a.	n.a.	0.08
		Z _{SCM}	56	8.99	0.28	0.05	1.93	0.02	0.91	7.60	2.82	57.80	9.45	0.22	0.96	2.15	n.a.	n.a.	n.a.	0.09
		Z _{b-SCM}	80	7.32	0.06	0.03	11.50	0.01	1.66	25.21	0.59	63.67	9.47	0.27	n.a.	0.49	n.a.	n.a.	n.a.	0.10
540	17/08	Z _{SURF}	3	8.72	0.01	0.03	0.01	0	0.55	3.09	1.68	62.07	5.73	0.20	0.75	7.39	193087	490	2887	0.07
		Z _{a-SCM}	50	9.13	0.04	0	0.03	0	1.36	3.83	1.40	69.43	5.14	0.15	0.63	2.68	266374	509	2289	0.08
		Z _{SCM}	70	8.26	0.29	0.06	6.17	0.01	2.20	14.04	1.77	83.22	5.77	0.10	0.91	1.34	221987	915	2091	0.09
		Z _{b-SCM}	85	7.15	0.06	0.06	10.28	0.01	2.75	22.43	0.84	77.07	6.32	0.07	0.16	0.64	137003	146	368	0.10
620	11/08	Z _{SURF}	3	8.35	0	0	0	0.03	0.33	7.82	5.61	99.48	4.56	0.11	3.36	14.71	424847	522	4036	0.13
		Z _{a-SCM}	50	9.14	0.02	0.01	0.01	0	0.63	3.60	0.92	60.77	4.34	0.19	0.22	5.44	280552	253	67	0.08
		Z _{SCM}	65	8.90	0	0	2.01	0.16	0.91	8.90	1.15	51.40	4.21	0.14	0.40	4.91	297468	146	108	0.09
		Z _{b-SCM}	80	8.04	0	0.01	6.83	0.05	1.26	18.34	0.92	n.a.	n.a.	n.a.	0.07	2.70	237149	128	0	0.11
670	10/08	Z _{SURF}	3	8.11	0	0	0.01	0.02	0.41	7.76	5.96	94.43	4.43	0.10	2.07	36.95	874089	400	1902	0.14
		Z _{a-SCM}	40	9.28	0	0	0.04	0.02	0.65	3.46	n.a.	74.20	5.65	0.53	n.a.	10.18	399293	190	128	0.09
		Z _{SCM}	60	8.50	0	0.11	0.33	0.07	0.96	8.13	1.95	72.25	14.68	0.76	0.17	2.10	300520	1072	0	0.09
		Z _{b-SCM}	80	7.29	0.02	0.02	6.30	0.03	1.29	15.33	1.15	59.96	3.92	0.11	0.14	1.310	244390	433	1049	0.10
760	12/08	Z _{SURF}	3	8.62	0.01	0	0.01	0.01	0.39	4.86	3.55	108.40	15.83	0.56	1.70	13.87	395137	723	4333	0.11
		Z _{a-SCM}	50	9.22	0.03	0.01	0	0.01	0.68	3.38	1.83	57.57	4.20	0.09	0.11	3.27	366178	743	5000	0.08
		Z _{SCM}	70	8.90	0.18	0.01	2.93	0.01	1	8.43	1.37	62.56	4.66	0.08	0.34	1.22	306567	1174	4423	0.09
		Z _{b-SCM}	90	7.36	0.03	0.02	10.04	0.01	1.54	21.72	0.69	58.72	5.27	0.04	0.10	0.68	162854	406	0	0.10

Phaeo: phaeopigments; POC: particulate organic carbon; DOC: dissolved organic carbon; DON: dissolved organic nitrogen; DOP: dissolved organic phosphorus; PP: primary production; BP: bacterial production; CDOM: colored dissolved organic matter.

^a: $\mu\text{M kg}^{-1}$.

^b: in mg m^{-3} ; small: 0.7-5 μm , large: 5-20 μm .

^c: in μM .

^d: in $\text{mgC m}^{-3} 24\text{h}^{-1}$.

^e: in $\text{pmol leu l}^{-1} \text{h}^{-1}$.

^f: bacterial, nano- and picophytoplanktonic cell abundances in cell ml^{-1} (flow cytometry).

^g: *in situ* CDOM fluorescence in mg m^{-1} .

Table S2. Sequence, OTU and diversity information for the 24 Beaufort Sea microbial samples.

Station	Depth category	Raw reads	Quality reads ^a	Total OTUs ^{b,c}	PD ^c	Shannon ^c	Chao1 ^c	NRI
430	Z _{SURF}	13077	10495	163	19.14	4.98	204.17	2.38
	Z _{a-SCM}	12905	10609	159	19.55	4.60	167.71	1.31
	Z _{SCM}	11261	9381	181	21.80	4.60	212.05	1.56
	Z _{b-SCM}	11102	8405	278	32.32	6.41	319.00	2.96
460	Z _{SURF}	12600	9591	224	23.71	5.87	246.80	3.06
	Z _{a-SCM}	13723	11123	202	23.97	5.05	235.66	1.47
	Z _{SCM}	13186	10732	234	28.08	5.52	286.00	2.20
	Z _{b-SCM}	8356	6150	293	37.02	6.61	371.08	2.09
540	Z _{SURF}	12454	9900	192	20.07	5.58	244.00	3.32
	Z _{a-SCM}	11105	9138	222	24.99	5.70	260.78	1.91
	Z _{SCM}	12701	9985	269	29.79	5.79	345.66	1.83
	Z _{b-SCM}	10399	7544	300	35.20	6.67	350.83	1.91
620	Z _{SURF}	10128	7557	228	24.63	5.90	314.56	2.54
	Z _{a-SCM}	11323	7722	186	19.30	5.60	305.00	3.50
	Z _{SCM}	10271	6925	229	27.97	5.70	314.55	3.21
	Z _{b-SCM}	8449	5585	262	32.29	6.28	342.37	4.07
670	Z _{SURF}	12790	8348	257	24.82	6.45	289.46	3.47
	Z _{a-SCM}	9380	6859	294	32.49	6.24	338.38	3.75
	Z _{SCM}	9650	5816	217	26.10	5.30	255.02	3.19
	Z _{b-SCM}	9640	5799	307	37.92	6.71	377.03	3.25
760	Z _{SURF}	12130	8319	185	20.29	5.65	250.87	3.35
	Z _{a-SCM}	12009	6569	204	23.93	5.19	350.75	1.57
	Z _{SCM}	12465	8443	278	33.11	6.19	331.03	2.14
	Z _{b-SCM}	11156	6964	317	36.84	6.68	369.48	2.08

OTUs: operational taxonomic units; PD: phylogenetic diversity; NRI: net-relatedness index.

^a: after quality controls (de-noising, removal of short, low-quality and/or chimeric reads).

^b: at 98% sequence identity.

^c: computed on quality read datasets subsampled at an even depth of 5000 sequences.

Table S3. Results of environmental factor fitting on ordination and permutational analyses of variance on distance matrices (EFF and PVM, respectively).

Variable	EFF		PVM	
	<i>R</i> ²	<i>P</i>	<i>R</i> ²	<i>P</i>
SiOH ₄	0.575	0.003	0.370	< 0.001
Picoeuk ^a	0.549	0.001	0.321	< 0.001
NO ₃ +NO ₂	0.519	0.006	0.343	< 0.001
O ₂	0.477	0.009	0.341	< 0.001
NH ₄	0.338	0.026	0.062	0.204
Light	0.319	0.069	0.191	0.01
DOP	0.264	0.105	0.031	0.576
Bacteria ^a	0.188	0.235	0.096	0.122
Nanoeuk ^a	0.083	0.539	0.109	0.09

DOP: dissolved organic phosphorus; Picoeuk/Nanoeuk: picoeukaryotic and nanoeukaryotic phytoplankton cell counts, respectively.

^a: Cell counts via flow cytometry; data missing for station 460.

^b: DOP data missing for samples 430-Z_{aSCM}, 430-Z_{bSCM} and 620-Z_{bSCM}.

Table S4. Taxa differently abundant between SCM and wSCM communities ($P \leq 0.05$).

Eukaryotic clade	Lineage	Taxa	SCM relative abundance ^a	wSCM relative abundance ^a	<i>P</i> -value ^b
Alveolata	Ciliate (Intramacronucleata)	<i>Laboea</i>	0.62±3.5e ⁻³	0.02 ±2.7e ⁻⁵	0.011
		<i>Novistrombidium</i>	8.18±0.3	2.06±1.5e ⁻³	0.008
		<i>Pseudotontonia</i>	3.74±0.04	1.11±0.01	0.005
		<i>Tintinnopsis</i>	0.05±2.4e ⁻⁵	2.4e ⁻³ ±3.4e ⁻⁷	0.007
	Dinoflagellates (Dinophyceae GPP)	NOR26.29 group	0	0.39±2.2e ⁻³	0.042
		<i>G. helveticum</i> group	1.12±1.5e ⁻³	10.55±0.7	0.009
		<i>Lepidodinium</i>	1.29±0.01	0.15±3.1e ⁻⁴	0.013
	Ellobiopsidae	<i>Thalassomyces</i>	0	9.6e ⁻³ ±5.5e ⁻⁶	0.022
		IIB / Guillou II.27*	0.01±1.0e ⁻⁶	4.2e ⁻³ ±4.4e ⁻⁷	0.021
		MALVs	IIB / Guillou II.7.36*	0.98±1.6e ⁻³	0.49±6.3e ⁻⁴
Cryptophyta	Cryptomonadales	II / Guillou II.3.4*	3.20±0.09	0.61±2e ⁻³	0.029
		Cryptomonas	0	9.6e ⁻³ ±5.5e ⁻⁶	0.022
		Unclassified	CCMP2045	0.03±8.8e ⁻⁶	0
Picozoa	Picozoa	Arctic NW617.02	2.73±0.02	11.69±0.7	0.018
Rhizaria	Cercozoa	NOR46.14	6.6e ⁻³ ±8.1e ⁻⁷	0.08±4.7e ⁻⁵	0.012
		<i>Cryothecomonas</i>	0.07±3.2e ⁻⁵	0.01±4.0e ⁻⁶	0.011
		<i>Ebria</i>	0	0.02±2.6e ⁻⁶	0.0007
		<i>Pseudopirsonia</i>	0	0.01±4.3e ⁻⁶	0.022
Stramenopiles	Bolidophyceae	<i>Bolidomonas</i>	0.06±2.9e ⁻⁵	0.13±1.7e ⁻⁵	0.011
	Bacillariophyceae	<i>Craticula</i>	0	0.01±8.5e ⁻⁶	0.022
	Mediophyceae	<i>Chaetoceros</i>	0.04±4.4e ⁻⁵	0.29±5.3e ⁻⁴	0.014
		<i>Thalassiosira</i>	0.50±6.4e ⁻⁴	0.99±2.4e ⁻³	0.033
	MAST	MAST-1B	0.13±1.4e ⁻⁴	0.75±4.7e ⁻³	0.031
		MAST-2	0.08±3.1e ⁻⁵	0.03±4.5e ⁻⁶	0.044
MAST-3		0.03±2.9e ⁻⁵	0.24±4.0e ⁻⁴	0.017	
MAST-7		1.80±2.5e ⁻³	2.50±1.7e ⁻³	0.002	
MAST-8	0.20±4.9e ⁻⁴	1.15±3.1e ⁻³	0.001		
Telonemia	Telonemida	<i>Telonema</i>	2.57±0.01	1.34±2.7e ⁻³	0.034
Viridiplantae	Mamiellophyceae	<i>Micromonas</i> CCMP2099	24.89±0.95	9.44±2.1	0.03

^a : mean relative abundance ± standard deviation (%).

^b : significance at $P < 0.05$ (Metastats).

* : Guillou L., *et al.*, *Environmental Microbiology*, 10: 3349–3365. doi: 10.1111/j.1462-2920.2008.01731.x