Supplementary Information

Water mass-driven multiple ecological effects determine the biodiversity and assembly processes of microbial flagellates' communities in subtropic-tropic marginal seas in China

Xin Guo[1], Qiang Liu[1], Xiaoqing Lin[1], Xinyi Zheng[1], Cheng Huang[1], Mengwen Pang[2], Lingfeng Huang[1 *]

[1]Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen, 361102, China
[2]Department of Ocean Science, Hong Kong University of Science and Technology, Hong Kong, 999077, China

***Corresponding author**
Lingfeng Huang. Phone (Office): +86592-2188455, Fax (Office): +86 592-2188455, Phone (Home): +8613906053195, E-mail: huanglf@xmu.edu.cn

The supplementary information includes:

- **30 pages**
- **Supplementary methods (pages 2–7)**
- **5 supplementary tables (pages 8–18)**
- **7 supplementary figures (pages 19–30)**

Supplementary Methods

A. Simplification of the sampling seasons

In this study, samples were collected in three seasons from three sea areas, which are the spring samples from the East China Sea (ECS) and the Taiwan Strait (TS), the summer samples from the TS and the South China Sea (SCS), and the autumn samples from the ECS. Notably, samples from the ECS and the TS were all taken in year 2019, while samples from the SCS were in year 2017. Thus, to reveal the spatiotemporal difference of the environmental variations and community compositions of samples, principal component analysis (PCA) and non-metric multidimensional scaling (NMDS) analysis were conducted, respectively.

PCA analysis revealed the similar environmental conditions in the autumn samples from the ECS and summer samples from the TS and the SCS (Fig. S1a). This result was corresponding with the short sampling time interval of three sea areas, i.e., the samples of the TS and the SCS was in the end of summer and the samples of the ECS was at the beginning of autumn. It appeared that the flagellate communities were also similar in the ECS's autumn samples and the TS's summer samples, which had large difference with the spring samples (Fig. S1b (i, iii)). However, we found that the communities in the SCS's summer sample were different from those in other samples (Fig. S1b (ii, iv)). This was probably because of their large geographical distance (Fig. 2a) and long sampling time interval (year 2017 and 2019, respectively) that cause the difference in communities.

Based on the results above, we combined the ECS's autumn samples with other summer samples to form a coherence in the spatial distribution and thus to better compare their spatiotemporal differences. In this case, we simplified the seasons to just two, spring and the transitional phase from summer to autumn (summer-autumn).

B. Environmental abiotic and biotic factors

Water temperature (Temp), salinity (Sal), depth, fluorescence (Fluo), and the dissolved oxygen (DO) concentration were recorded *in situ* with a SeaBird conductivity-temperature-depth (CTD) profiler (SBE917plus, SeaBird Corp., USA). Samples (100 mL) for dissolved inorganic nutrients, including nitrate (NO_3-N), nitrite (NO_2-N), ammonia nitrogen (NH_4-N), active phosphate (PO_4-P), and active silicate (SiO_3-Si) were filtered through Whatman GF/F filters and measured with a nutrient analyzer (QUATTRO, Seal, Germany) using the manufacturer's standard procedures.

Samples (200–1 ,000 mL) for chlorophyll *a* (Chl *a*) were pre-filtered through a 200 μm mesh sieve to remove large zooplankton and debris. Filtered water was then passed through 0.7 μm GF/F filters (Whatman, England), with the filters then wrapped in aluminum foil and stored at -20°C, in the darkness. Chl *a* was then extracted by 10 mL of 90% acetone at 0°C for 20 h in the dark and measured using a Trilogy Fluorometer (Turner Designs, Trilogy Module: CHL-A NA). The abundance of heterotrophic bacteria (HB), cyanobacteria (Cyan) and pico-sized eukaryotic algae (pico-Euk) were run with a FACSAria flow cytometer (BD Corp., Franklin lake, NJ, USA)(Chiang et al., 2014). Samples (500–1,000 mL) for microzooplankton (micro-Zoo) were filtered through 100 μm nylon mesh and then fixed with Lugol's solution to a final concentration of 1.5%. Samples were mixed thoroughly and stood still for 48 h before siphoning

the supernatant to concentrate to a final volume of 50 mL(Utermöhl, 1958). The abundance of micro-Zoo was measured using FlowCAM with a 10 × magnification microscope objective lens and FC100 flow cell and counted through VisualSpreadsheet software(Poulton, 2016).

The sampling of nano-sized flagellate (NF) cells followed the protocols with the steps of fixing, filtering, staining, and mounting(Granda & Anadon Alvarez, 2008; Huang et al., 2008; Lin, Huang, Zhu, & Jia, 2013). NF were then grouped according to their trophic status and identified according to their fluorescence under epifluorescence microscopy (Leica DM 4500B) at 1000 × magnification: non-pigmented heterotrophic nano-sized flagellates (HNF) and pigmented nano-sized flagellates (PNF). In short, cells which that appeared blue under UV light were the biological cells, while PNFs were distinguished from HNFs by the presence of red autofluorescence under a blue excitation light. At least 40 fields of view with more than 100 cells (PNF plus HNF) were examined for each filter to obtain reliable estimates of abundance.

C. DNA extraction, high-throughput sequencing, and sequencing data processing

DNA was extracted from cells collected onto filters that were cut into pieces using the PowerWater® DNA Isolation Kit (Qiagen, USA) according to the manufacturer's instructions. DNA extracts were eluted in 100 μL elution buffers provided in the kit, quantified using a Nanodrop ND-2000 Spectrophotometer (Thermo Scientific), and verified on a 1.5% agarose gel. DNA extracts were kept at -20°C until processing. The hypervariable V4 region (ca. 400 bp) of the eukaryotic 18S rRNA gene was amplified using the paired primers 3NDF (5'-GGCAAGTCTGGTGCCAG-3')(Cavalier-Smith, Lewis, Chao, Oates, & Bass, 2009) and V4_euk_R2 (5'-ACGGTATCTRATCRTCTTCG-3')(Brate et al., 2010) referred the PCR program of Xu et al(Xu et al., 2018). The amplification mix (50 μL) contained: 25 μL 2 × *EasyTaq*® PCR SuperMix (TransGen Biotech, Beijing, China) polymerase, 20 μL DEPC water (TransGen Biotech, Beijing, China), 1 μL of each primer (10 μM, Sangon Biotech, Shanghai, China), and 1 μL genomic DNA as template. Cycling conditions were an initial denaturation step at 95°C for 2 min; 35 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 45 s; and a final 10-min extension at 72°C.

Forward and reverse primers were tagged with 2 bp links and 8 bp barcodes to allow pooling multiple samples in one run of sequencing and later differentiation of different samples(Kozich, Westcott, Baxter, Highlander, & Schloss, 2013). Five individual PCR reactions were run per sample to reduce the amplification error. Used 1.5% agarose gel electrophoresis to detect the PCR amplified product and the gel of the target gene fragment with the length of 400 bp was cut. The retained product was then purified with a Fast Agarose Gel DNA Recovery Kit (BioTeke Corp., Wuxi, China), followed by pooling of individual amplicon libraries in equimolar concentrations. The mixture of products was sequenced by a commercial sequencing company (Novogene Corp., Ltd., Beijing, China) with a paired-end 2 × 250 bp sequencing run with V3 chemistry on an Illumina Hiseq 2500 platform (Illumina, Inc., San Diego, CA, USA).

The sequencing data for pico- and nano-sized samples were processed separately, according to the standard operating procedure of processing sequencing data generated from llumina's MiSeq platform using paired end reads (https://www.mothur.org.\/wiki/454_SOP, 23th

Sept., 2019). Raw sequencing reads were in total of 9,705,020 and 11,965,629, ranging 4,301 to 456,075 and 3,140 to 423,599 per sample for pico- and nano- sized datasets, respectively (Table S2). Following removal of low-quality reads, rare reads, potential chimeras, reads that were not assigned as protists (including unknown), singletons and doubletons, there were 2,421,967 and 3,235,828 sequences left, ranging 1,083 to 62,840 reads and 1,107 to 89,027 reads per sample for pico- and nano- sized datasets, respectively (Table S2). Specifically, demultiplexing processes for the raw sequencing data were firstly conducted according to the following parameters: sequence length remains between 400 and 444 bp; number of ambiguous bases is 0; maximum homopolymer length is 6. Quality-checked reads with only one or two sequences (i.e., rare reads) were then removed to avoid sequencing errors. This step removed a large part of the sequencing reads (Table S2). The remaining reads were aligned against the aligned SILVA 132 database(Quast et al., 2013) (https://www.arb-silva.de/, 6th Jul., 2019). The bulk of sequences that started at position 13,135 and ended at position 22,581 were extracted. These alignments were then trimmed using "vertical = T" and "trump =." options to ensure that reads with the same primer set were aligned to the same exact region. A further screening step (pre.cluster) was applied to decrease sequencing noise by clustering reads that differ by 5 bp (diffs = 5). Chimeras were detected on individual samples using the UCHIME algorithm(Edgar, Haas, Clemente, Quince, & Knight, 2011) in a de novo setup and subsequently removed. For each sample, clean reads were dereplicated. An operational taxonomic unit (OTU) table was constructed by clustering high-quality reads at a 3% genetic distance, based on the furthest neighbor cluster method(Logares et al., 2014). OTUs were then taxonomically classified using the RDP Classifier with a naïve Bayesian approach against a Protist Ribosomal Reference (PR^2) database(Guillou et al., 2013) at an 80% confidence level. Distant OTUs with an e-value > 10^{-100} (below ~ 80% similarity) were considered as 'unknown' and were removed. Taxa that are not affiliated with protists (e.g., Bacteria, Archaea, Nucleomorphs, and Metazoa) were removed from the dataset before downstream analysis to avoid distortion of the relative abundance of DNA sequences of microbial eukaryotes. Singletons, doubletons (i.e., OTUs with only one or two sequences), and OTUs present in a single sample were discarded before the downstream analyses as potential sequencing errors.

Normalization was conducted to enable comparison between samples in different sequencing coverage depths (Fig. S2a (i, ii)), based on the lowest sequence count (1,083 sequences for pico-sized samples and 1,107 sequences for nano-sized samples) from each dataset (Table S2, Fig. S2a (iii, iv)). Samples with low sequencing depths (< 5,000) account for 28.74% and 15.95% of the pico- and nano-sized samples, respectively (Fig. S2d). Thus, the subsampling threshold was set at a low level (~1,000) to keep as many samples as possible. A mantel test based on Bray-Curtis distances and using Spearman's correlation coefficient was then conducted to test the relationship between original OTUs and subsampled OTUs. The result indicated that this subsampling set represented very well the original pool of sequences (R = 0.84 for pico-sized and R = 0.75 for nano-sized, $P < 0.0001$, Fig. S2e). Overall, we think the subsampled sequencing depth is reasonable and representative for the subsequent analysis.

To focus on our target organism, we then only selected OTUs affiliated to the microbial flagellates according to the literature (Table S3)(Adl et al., 2019; Archibald et al.; Throndsen,

1997). Taxa that are not affiliated with flagellates (e.g., bacteria, archaea, metazoa, diatoms, ciliates, radiolarians, and fungi) were removed from the datasets. Finally, two OTU tables with a total of 131,606 and 125,487 microbial flagellate sequences clustered into 1,489 and 1,340 flagellate OTUs were formed separately for pico- and nano- datasets (Table S4). For the taxonomic compositions, OTUs belonging to the lower (i.e., genus) to higher (i.e., phylum) taxonomic level were identified into different lineages and the undetermined or unclassified OTUs from the given lineages were assigned as "others". The lineages were classified mainly based on Adl et al. (2019)(Adl et al., 2019).

D. Water masses classification in the studied area

In this study, we focused on the temperature-salinity data that were measured by conductivity-temperature-depth (CTD) profiler (SBE917plus, SeaBird Corp., WA, USA) from all the stations during several cruises and used them to classify the water masses in spring and summer-autumn. There was total 277 stations, i.e., 104 in spring, and 173 in summer-autumn. Samples from ECS and TS were collected at two water depths (2 m beneath the sea surface representing the surface layer and 3–4 m above the seabed representing the bottom layer). Samples from the SCS were collected at three layers (surface, deep chlorophyll maximum (DCM), and the bottom of the photic zone, referred to as Bottom, at a depth of ~200 m). Thus, there were a total of 450 samples used to determine the water masses in the ECS and the TS during spring and summer-autumn in 2019 (Fig. S4), while total 155 samples were used in water mass classification in the SCS in summer, 2017 (Fig. S5).

The horizontal distribution of temperature and salinity in different layers in the studied area in spring and summer-autumn were shown in Fig. S3 and Fig. S5a. The water masses in spring and summer-autumn were determined on the basis of the temperature and salinity data using the fuzzy cluster analysis method recorded in Zhu et al. (2019)(Zhu et al., 2019). Furthermore, a temperature-salinity similarity number (TSSN) were defined to quantify the water sample by clearly specifying the boundaries between the water masses, thus to present a more realistic distribution of water masses. (Zhu et al., 2019). The temperature-salinity diagrams were also conducted for the analysis of water masses (Fig. S4c, d, Fig. S5b). Overall, there were 7 and 10 water masses identified in spring and summer-autumn in 2019, respectively (Fig. S4a, b). Other three water masses were defined in the SCS in summer, 2017 (Fig. 5c).

It is notable that water mass classification of the SCS had some limitations and problems. As the northern SCS is located near the Luzon Strait with variable hydrological conditions, the water masses in this area are very complex and difficult to identify(Zhu et al., 2019). Because of the limitation of the cruise, the sample size used for water mass analysis in this area were not sufficient (Fig. S5). Thus, the continuity and accuracy of the distribution of the water masses in the SCS were still open to question. Therefore, we stated that the result of the water mass classification in the SCS was only for reference in this study. According to the previous study, we identified three water masses in the SCS in summer, 2017, i.e., South China Sea surface water (Ss, less than 50 m), South China Sea subsurface water (Su, 100–250 m), and Pacific Ocean subsurface water (Pu, 150–250 m)(Zhu et al., 2019). Because the distributions of water masses usually show great interannual and seasonal dynamics, especially the surface water, we further did principal component analysis (PCA) to display the difference between the SCS water

masses in summer in year 2017 and in year 2019 (identified as "Ss", "Sm", and "Si", shown in Fig. S2b). The result confirmed the distinction of the three water masses in the SCS in 2017 (Fig. S5c). Therefore, the water masses in the SCS in summer in 2017 were named as “Ss-2017”, “Su-2017”, and “Pu-2017”.

Supplementary references

Adl, S. M., Bass, D., Lane, C. E., Lukes, J., Schoch, C. L., Smirnov, A., . . . Zhang, Q. (2019). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology, 66*(1), 4-119.https://doi.org/10.1111/jeu.12691

Archibald, J. M., Simpson, A. G. B., Slamovits, C. H., Margulis, L., Melkonian, M., Chapman, D. J., & Corliss, J. O. (Eds.). (2017). *Handbook of the Protists*: Springer, Cham.

Brate, J., Logares, R., Berney, C., Ree, D. K., Klaveness, D., Jakobsen, K. S., & Shalchian-Tabrizi, K. (2010). Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *The ISME journal, 4*(9), 1144-1153.https://doi.org/10.1038/ismej.2010.39

Cavalier-Smith, T., Lewis, R., Chao, E. E., Oates, B., & Bass, D. (2009). Helkesimastix marina n. sp.(Cercozoa: Sainouroidea superfam. n.) a gliding zooflagellate of novel ultrastructure and unusual ciliary behaviour. *Protist, 160*(3), 452-479.https://doi.org/10.1016/j.protis.2009.03.003

Chiang, K.-P., Tsai, A.-Y., Tsai, P.-J., Gong, G.-C., Huang, B.-Q., & Tsai, S.-F. (2014). The influence of nanoflagellates on the spatial variety of picoplankton and the carbon flow of the microbial food web in the oligotrophic subtropical pelagic continental shelf ecosystem. *Continental Shelf Research, 80*, 57-66.https://doi.org/10.1016/j.csr.2014.02.019

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*(16), 2194-2200.https://doi.org/10.1093/bioinformatics/btr381

Granda, A. P., & Anadon Alvarez, R. (2008). The annual cycle of nanoflagellates in the Central Cantabrian Sea (Bay of Biscay). *Journal of Marine Systems, 72*(1-4), 298-308.https://doi.org/10.1016/j.jmarsys.2007.09.009

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., . . . Christen, R. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research, 41*(D1), D597-D604.https://doi.org/10.1093/nar/gks1160

Huang, B. Q., Lan, W. L., Cao, Z. R., Dai, M. H., Huang, L. F., Jiao, N. Z., & Hong, H. S. (2008). Spatial and temporal distribution of nanoflagellates in the northern South China Sea. *Hydrobiologia, 605*(1), 143-157.https://doi.org/10.1007/s10750-008-9330-3

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology, 79*(17), 5112-5120.https://doi.org/10.1128/AEM.01043-13

Lin, S., Huang, L., Zhu, Z., & Jia, X. (2013). Changes in size and trophic structure of the nanoflagellate assemblage in response to a spring phytoplankton bloom in the central Yellow Sea. *Deep-Sea Research II, 97*, 93-100.https://doi.org/10.1016/j.dsr2.2013.05.017

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., . . . Massana, R. (2014). Patterns of rare and abundant marine microbial eukaryotes. *Current Biology, 24*(8), 813-

821.https://doi.org/10.1016/j.cub.2014.02.050

Poulton, N. J. (2016). FlowCam: quantification and classification of phytoplankton by imaging flow cytometry. In N. S. Barteneva & I. A. Vorobjev (Eds.), *Imaging Flow Cytometry: Methods and Protocols, Methods in Molecular Biology* (2016/07/28 ed., Vol. 1389, pp. 237-247). New York: Humana Press.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research, 41*(D1), D590-D596.https://doi.org/10.1093/nar/gks1219

Throndsen, J. (1997). The Planktonic Marine Flagellates. In C. R. Tomas (Ed.), *Identifying Marine Phytoplankton* (pp. 591-729): Academic Press.

Utermöhl, H. (1958). The improvement of quantitative phytoplankton methodology. *SIL Communications: Internationale Vereinigung für Theoretische und Angewandte Limnologie: Mitteilungen, 1953-1996, 9*(1), 1-38.https://doi.org/10.1080/05384680.1958.11904091

Xu, Z. M., Wang, M., Wu, W. X., Li, Y. F., Liu, Q., Han, Y. Y., . . . Liu, H. B. (2018). Vertical Distribution of Microbial Eukaryotes From Surface to the Hadal Zone of the Mariana Trench. *Frontiers in microbiology, 9*, 2023.https://doi.org/10.3389/fmicb.2018.02023

Zhu, J., Zheng, Q. A., Hu, J. Y., Lin, H. Y., Chen, D. W., Chen, Z. Z., . . . Kong, H. (2019). Classification and 3-D distribution of upper layer water masses in the northern South China Sea. *Acta Oceanologica Sinica, 38*(4), 126-135.https://doi.org/10.1007/s13131-019-1418-2

228 **Table S1 The sample size of two size-fraction groups of microbial flagellate**
229 **communities for each space and time.**

Number of samples		pico-sized	nano-sized	Total
Spring	**ECS**	37	37	74
	TS	17	27	44
Summer-autumn	**ECS**	29	31	60
	TS	34	40	74
	SCS	50	28	78
Total		167	163	330

ECS: East China Sea; TS: Taiwan Strait; SCS: South China Sea.

230

Table S2 General information of the number of sequencing reads of samples after each step of sequencing data processing.

Steps of sequencing data processing	Number of sequencing reads							
	Pico (N = 167) [a]				Nano (N = 163) [a]			
	Minimum	Maximum	Average	Total	Minimum	Maximum	Average	Total
Raw sequencing reads	4 301	456 075	58 114	9 705 020	3 140	423 599	73 409	11 965 629
Demultiplexing and quality check	3 974	355 809	50 626	8 454 562	2 951	289 669	67 113	10 939 408
Removal of rare reads [b]	1 365	68 224	17 771	2 967 739	1 399	114 224	26 029	4 242 701
Alignment and trimming	1 362	64 584	17 467	2 916 983	1 390	113 826	25 364	4 134 285
Pre-clustering	1 362	64 584	17 467	2 916 983	1 390	113 826	25 364	4 134 285
Chimera removal	1 319	63 037	17 064	2 849 730	1 372	113 155	24 835	4 048 095
Clean reads affiliated with protists	1 083	62 841	14 504	2 422 227	1 107	89 033	19 854	3 236 261
OTUs clustering	1 083	62 841	14 504	2 422 227	1 107	89 033	19 854	3 236 261
Removal of singletons and doubletons [c]	1 083 [d]	62 840	14 503	2 421 967	1 107 [d]	89 027	19 852	3 235828
Normalization	1 083	1 083	1 083	180 861	1 107	1 107	1 107	180 441

a. N: the number of samples collected in each dataset;
b. Rare reads: reads with only one or two sequences;
c. Singletons and doubletons: OTUs with only one or two representative reads;
d. The subsampled threshold was set based on the lowest sequence count of the samples.

233 **Table S3 The selected microbial flagellate taxa that detected in this study.**

Taxa	Taxonomic level	Ref.
Eukaryota; Alveolata; Dinophyta; Syndiniales; MALV-I	Order	[1, 2]
Eukaryota; Alveolata; Dinophyta; Syndiniales; MALV-II	Order	[1, 2]
Eukaryota; Alveolata; Dinophyta; Syndiniales; MALV-III	Order	[1, 2]
Eukaryota; Alveolata; Dinophyta; Syndiniales; MALV-IV	Order	[1, 2]
Eukaryota; Alveolata; Dinophyta; Syndiniales; MALV-V	Order	[1, 2]
Eukaryota; Alveolata; Dinophyta; Dinophyceae	Class	[1–3]
Eukaryota; Apusozoa; Apusomonadidae	Phylum	[4]
Eukaryota; Archaeplastida; Chlorophyta; Mamiellophyceae; Mamiellales; Bathycoccaceae; *Bathycoccus*	Genus	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Chlorodendrophyceae; Chlorodendrales	Order	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Mamiellophyceae; Mamiellales; Mamiellaceae; *Mantoniella*	Genus	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Pedinophyceae; Marsupiomonadales; Marsupiomonadaceae; *Marsupiomonas*	Genus	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Mamiellophyceae; Mamiellales; Mamiellaceae; *Micromonas*	Genus	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Trebouxiophyceae	Class	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Prasino-Clade	Class	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Pyramimonadales; Pyramimonadales_X; Pyramimonadales_XX; *Pyramimonas*	Genus	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Chlorophyceae; Chlamydomonadales; Chlamydomonadales_X; *Chlamydomonas*	Genus	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Chlorophyceae; Sphaeropleales; Sphaeropleales_X; *Mychonastes*	Genus	[3, 5]
Eukaryota; Archaeplastida; Chlorophyta; Pyramimonadales; Pyramimonadales_X; Pyramimonadales_XX; *Pterosperma*	Genus	[3, 5]
Eukaryota; Hacrobia; Cryptophyta	Phylum	[2, 3, 6]
Eukaryota; Hacrobia; Haptophyta	Phylum	[2, 7]
Eukaryota; Hacrobia; Katablepharidophyta	Phylum	[2, 8]
Eukaryota; Hacrobia; Picozoa	Phylum	[2, 3]
Eukaryota; Hacrobia; Telonemia	Phylum	[2, 3]
Eukaryota; Opisthokonta; Choanoflagellida	Phylum	[9, 10]
Eukaryota; Rhizaria; Cercozoa	Phylum	[2, 11]
Eukaryota; Stramenopiles; Ochrophyta; Chrysophyceae	Class	[12]
Eukaryota; Stramenopiles; Stramenopiles_X; Labyrinthulea	Class	[13]
Eukaryota; Stramenopiles; Stramenopiles_X; MAST	Class	[13]
Eukaryota; Stramenopiles; Stramenopiles_X; MOCH	Class	[2]
Eukaryota; Stramenopiles; Stramenopiles_X; Oomycota	Class	[2]

Eukaryota; Stramenopiles; Ochrophyta; Pelagophyceae	Class	[2]
Eukaryota; Stramenopiles; Stramenopiles_X; Pirsonia_Clade	Class	[2]
Eukaryota; Stramenopiles; Ochrophyta; Raphidophyceae	Class	[3, 14]

References:

[1] Saldarriaga, J. F., and F. J. R. Taylor. 2017. Dinoflagellata, p. 1-54. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_22-1.

[2] Adl, S. M. and others 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J Eukaryot Microbiol 66: 4-119. doi:10.1111/jeu.12691.

[3] Throndsen, J. 1997. The Planktonic Marine Flagellates, p. 591-729. *In* C. R. Tomas [eds.], Identifying Marine Phytoplankton. Academic Press. doi:10.1016/B978-012693018-4/50007-0.

[4] Heiss, A. A., M. W. Brown, and A. G. B. Simpson. 2016. Apusomonadida, p. 1-27. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_15-1.

[5] Leliaert, F., D. R. Smith, H. Moreau, M. D. Herron, H. Verbruggen, C. F. Delwiche, and O. De Clerck. 2012. Phylogeny and Molecular Evolution of the Green Algae. Critical Reviews in Plant Sciences 31: 1-46. doi:10.1080/07352689.2011.615705.

[6] Hoef-Emden, K., and J. M. Archibald. 2016. Cryptophyta (Cryptomonads), p. 1-41. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_35-1.

[7] Eikrem, W. and others 2017. Haptophyta, p. 1-61. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_38-2.

[8] Bennett, R. M., D. Honda, G. W. Beakes, and M. Thines. 2017. Labyrinthulomycota, p. 1-36. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_25-1.

[9] Richter, D. J., and F. Nitsche. 2016. Choanoflagellatea, p. 1-19. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_5-1.

[10] King, N. 2005. Choanoflagellates. Current Biology 15: R113-R114. doi:10.1016/j.cub.2005.02.004.

[11] Ota, S., W. Eikrem, and B. Edvardsen. 2012. Ultrastructure and Molecular Phylogeny of Thaumatomonads (Cercozoa) with Emphasis on Thaumatomastix salina from Oslofjorden, Norway. Protist 163: 560-573. doi:10.1016/j.protis.2011.10.007.

[12] Kristiansen, J., and P. Škaloud. 2016. Chrysophyta, p. 1-38. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_43-1.

[13] Hamamoto, Y., and D. Honda. 2019. Nutritional intake of Aplanochytrium (Labyrinthulea, Stramenopiles) from living diatoms revealed by culture experiments suggesting the new prey-predator interactions in the grazing food web of the marine ecosystem. PLoS One 14: e0208941. doi:10.1371/journal.pone.0208941.

[14] Massana, R., J. del Campo, M. E. Sieracki, S. Audic, and R. Logares. 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. ISME Journal 8: 854-866. doi:10.1038/ismej.2013.204.

[15] Horiguchi, T. 2016. Raphidophyceae (Raphidophyta), p. 1-26. *In* J. M. Archibald et al. [eds.], Handbook of the Protists. Springer, Cham. doi:10.1007/978-3-319-32669-6_37-1.

234

Table S4 General descriptions of the richness and diversity of protist and microbial flagellate communities in the regional area.

Communities	Sobs [c]	Reads	Chao 1[d]	Shannon [d]	Simpson [d]
Pico (N = 167)					
Total protists	3 227	180 861	331	3.06	0.20
After removing OTUs only present in a single sample [a]	2 208 (68.42%)	179 669 (99.34%)	308	3.04	0.20
Microbial flagellates [b]	1 489 (67.44%)	131 606 (73.25%)	220	2.81	0.23
Nano (N = 163)					
Total protists	3 019	180 441	280	3.12	0.15
After removing OTUs only present in a single sample [a]	1 937 (64.16%)	179 083 (99.25%)	256	3.09	0.15
Microbial flagellates [b]	1 340 (69.18%)	125 487 (70.07%)	191	2.93	0.17

N, the number of samples collected in each dataset;

a. The percentages in parentheses of the Sobs and the reads are the percentages of the protists after excluding the OTUs that only appear in one sample;

b. The percentages in parentheses of the Sobs and the reads are the percentages of whole microbial flagellates in the protists after excluding the OTUs that only appear in one sample;

c. Sobs represents the observed OTU numbers;

d. The average value of each diversity index among all the samples.

238 **Table S5 Environmental parameters of water masses in spring and summer-autumn throughout study (mean values).**

Mean values	Spring					Summer-autumn								
	Coastal	Mixed			Oceanic	Coastal		Mixed		Oceanic				
	Mc	Cm	Es	Ss	Ks	Mc	Zd	Cm	Es	Km	Sm	Ss-2017	Su-2017	Pu-2017
Temp (℃)	17.80	20.20	19.50	24.54	21.65	26.25	30.54	26.43	26.10	17.44	21.21	28.03	19.92	19.83
Sal (‰)	30.53	33.48	33.38	34.43	34.59	30.82	31.79	33.90	33.76	34.50	34.62	33.69	34.56	34.72
DO (mg.L^{-1})	8.16	6.80	6.49	6.17	6.03	6.45	7.14	6.22	5.45	5.27	4.81	6.83	4.88	5.27
Fluo (mg.m^{-3})	1.06	1.49	1.26	0.21	0.16	86.26	2.70	8.30	133.98	189.10	7.91	0.04	0.32	0.01
SiO_3-Si (μmol.L^{-1})	14.08	6.58	9.91	1.57	7.84	9.39	2.43	4.77	7.93	14.36	7.88	2.68	9.15	6.91
PO_4-P (μmol.L^{-1})	0.40	0.13	0.57	0.01	0.70	0.16	0.09	0.07	0.24	0.70	0.31	0.30	0.61	0.41
NH_4-N (μmol.L^{-1})	1.47	1.59	1.16	1.17	0.90	2.73	5.28	4.37	1.72	0.53	5.07	-	-	-
NO_3-N (μmol.L^{-1})	8.30	2.07	5.63	0.13	4.38	3.45	0.47	0.98	2.36	5.64	3.62	2.24	7.70	6.02
NO_2-N (μmol.L^{-1})	1.14	0.57	0.43	0.02	0.35	0.30	0.09	0.25	0.22	0.12	0.28	0.06	0.06	0.15
Chl *a* (μg.L^{-1})	2.27	1.22	0.57	0.27	0.11	3.28	0.80	2.25	0.76	0.24	2.63	-	-	0.22
HB (×10^6 mL^{-1})	1.98	0.55	1.85	0.26	1.31	3.36	0.27	0.37	2.15	1.01	0.31	0.70	0.51	0.09
Cyan (×10^4 mL^{-1})	3.95	1.06	9.95	4.59	8.01	20.31	7.78	4.79	4.71	0.67	1.32	1.50	2.46	-
pico-Euk (×10^3 mL^{-1})	16.33	8.10	2.97	3.25	1.87	16.86	2.60	6.80	3.94	0.82	4.36	-	-	-
micro-Zoo (mL^{-1})	7.89	13.50	4.77	2.62	4.51	29.03	8.22	22.14	19.30	6.70	13.64	-	-	-
HNF (×10^3 mL^{-1})	1.37	1.56	1.31	0.97	0.88	1.78	1.14	1.12	1.37	0.89	0.94	0.95	0.88	1.40
PNF (×10^3 mL^{-1})	2.81	0.85	1.37	0.83	0.38	2.59	2.59	1.86	1.14	0.23	1.04	1.06	0.71	0.43
NF (×10^3 mL^{-1})	4.19	2.41	2.68	1.80	1.26	4.37	3.74	2.98	2.51	1.11	1.97	2.01	1.54	1.83

Refer to Fig. 1 and Fig. 2 for abbreviations of the water masses. Water masses with less than 3 samples were not included in this analysis. The Mc and Zd were considered as coastal waters, the Cm, Es, and Ss were considered as mixed waters, while the Ks, Km, Sm, Ss-2017, Su-2017, and Pu-2017 were assorted to oceanic waters.

Environmental codes are as follows: Temp, temperature; Sal, salinity; DO, dissolved oxygen; Fluo, fluorescence; SiO_3-Si, Silicate; PO_4-P, dissolved inorganic phosphorus; NH_4-N, ammonia nitrogen; NO_3-N, nitrate; NO_2-N, nitrite; Chl *a*, chlorophyll *a*; HB, heterotrophic bacteria; Cyan, cyanobacteria; pico-Euk, pico-sized eukaryotic algae; micro-Zoo, microzooplankton; HNF, heterotrophic nanoflagellate; PNF, pigmented nanoflagellate, NF, nanoflagellate.

239

Table S6 Adonis analysis of different comparisons of environmental variables (including abiotic and biotic factors) based on different dimensions, including sampling seasons (spring *vs* summer-autumn), water masses, marine areas (ECS *vs* TS *vs* SCS), distance from the shore (coastal *vs* mixed *vs* oceanic waters), and layers (Surface *vs* Bottom).

Adonis (R^2)	Seasons	Water masses	Marine areas	Distance from the shore [d]	Layers [e]
[All factors]					
All samples	0.076***	0.230***	0.143***	0.029*	0.153***
All samples (-SCS) [a]	0.144***	0.255***	0.108***	0.056**	0.147***
Spring	-	0.311***	0.081**	0.169***	0.168***
Summer-autumn (-SCS) [a]	-	0.285***	0.272***	0.086	0.168***
Summer-autumn	-	0.384***	0.383***	0.052*	0.198***
[Abiotic factors] [b]					
All samples	0.087	0.309***	0.143***	0.025+	0.150***
All samples (-SCS) [a]	0.163	0.284***	0.122***	0.052**	0.125***
Spring		0.399***	0.112**	0.240***	0.125***
Summer-autumn (-SCS) [a]		0.309***	0.320***	0.088*	0.186***
Summer-autumn		0.386***	0.393***	0.042+	0.228***
[Biotic factors][c]					
All samples	0.032**	0.194***	0.164***	0.044**	0.113***
All samples (-SCS) [a]	0.065***	0.175***	0.102***	0.053***	0.120***
Spring		0.218***	0.083***	0.088**	0.143***
Summer-autumn (-SCS) [a]		0.190***	0.102***	0.082**	0.115***
Summer-autumn		0.262***	0.176***	0.114***	0.115***

a. All samples without samples from the SCS in summer;

b. Abiotic factors include temperature, salinity, DO, fluorescence, SiO_3-Si, PO_4-P, NO_3-N, and NO_2-N;

c. Biotic factors include Chl *a*, heterotrophic bacteria, cyanobacteria, and pico-eukaryotic algae;

d. The Mc and Zd were considered as coastal waters, the Cm, Es, and Ss were considered as mixed waters, while the Ks, Km, Sm, Ss-2017, Su-2017, and Pu-2017 were assorted to oceanic waters;

e. Samples from DCM in the SCS were removed and the comparisons were only computed between surface and bottom.

Significance: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; $^+$, $p < 0.1$.

Table S7 ANOSIM results of different comparisons based on sampling seasons (spring *vs* summer-autumn), water masses, marine areas (ECS *vs* TS *vs* SCS), distance from the shore (coastal *vs* mixed *vs* oceanic waters), and layers (Surface *vs* Bottom) among microbial flagellate communities based on Bray-Curtis distance throughout study.

ANOSIM (Global R)	Seasons	Water masses	Marine areas	Distance from the shore [b]	Layers [c]
Pico					
All samples	0.029	0.094***	0.159***	0.096***	0.108***
All samples (-SCS) [a]	0.145***	0.095**	0.089***	0.004	0.084***
Spring	-	0.227***	0.125*	0.034	0.036
Summer-autumn (-SCS) [a]	-	0.081*	0.173***	0.029	0.159***
Summer-autumn	-	0.275***	0.434***	0.273**	0.177***
Nano					
All samples	0.137***	0.145***	0.210***	0.125***	0.110***
All samples (-SCS) [a]	0.025*	0.101**	0.072***	0.049	0.000
Spring	-	0.332***	0.180**	0.137*	0.041^{+}
Summer-autumn (-SCS) [a]	-	0.163**	0.040^{+}	0.087	0.137***
Summer-autumn	-	0.410***	0.347***	0.212***	0.161***

a. All samples without samples from the SCS in summer;

b. The Mc and Zd were considered as coastal waters, the Cm, Es, and Ss were considered as mixed waters, while the Ks, Km, Sm, Ss-2017, Su-2017, and Pu-2017 were assorted to oceanic waters;

c. Samples from DCM in the SCS were removed and the comparisons were only computed between surface and bottom.

Significance: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; +, $p < 0.1$.

Table S8 Simple and partial mantel tests demonstrate Spearman's rank correlations of Euclidean distance of water mass (WM), environmental variability (E), and geographic distance (S) with the variation of microbial flagellate alpha diversity (Euclidean distance of samples based on Chao1 and Shannon indices).

Correlation (r)	Simple-Mantel tests			Partial-mantel tests		
	WM [c]	E[d]	S[e]	WM-E&S [f]	E-WM&S [g]	S-WM&E [h]
Pico						
Spring	0.149*	0.170**	0.216***	-0.012	0.144*	0.209***
Summer-autumn (-SCS) [b]	0.007	0.061	0.207**	-0.009	0.043	0.204***
Summer-autumn[a]	-0.024	0.069*	0.184***	-0.027	0.056*	0.179***
Nano						
Spring	0.210**	0.238***	0.262**	0.093^{+}	0.179**	0.219**
Summer-autumn (-SCS) [b]	0.072^{+}	0.073*	0.104*	0.024	0.031	0.093^{+}
Summer-autumn[a]	0.098*	0.116*	0.101*	0.077*	0.107*	0.103*

a. Samples from all three marine areas (ECS, TS, and SCS) in summer-autumn;

b. Samples from the SCS in summer were removed;

c. Euclidean distance of the factor of water masses (WM) represented by temperature and salinity between sampling sites;

d. Euclidean distance of environmental factors (E) represented by all abiotic and biotic factors except temperature and salinity between sampling sites;

e. Euclidean distance of all spatial factors (S) between sampling sites;

f. Correlation between alpha diversity variation and water mass controlled by E and S;

g. Correlation between alpha diversity and environmental factors controlled by WM and S;

h. Correlation between alpha diversity variation and spatial factors controlled by WM and E.

Significance: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; $^{+}$, $p < 0.1$.

Table S9 Simple and partial mantel tests showed the effects of water mass (WM), environmental factors (E), and spatial factors (S) on the beta diversity (Bray-Curtis distance) of pico-sized and nano-sized microbial flagellate communities in different water layers in the geographical scale.

Correlation (r)	Simple-mantel tests			Partial-mantel tests		
	WM [b]	E [c]	S [d]	WM-E&S [e]	E-WM&S [f]	S-WM&E [g]
Pico						
Spring-Surface	0.457***	0.292***	0.137	0.364***	0.227**	0.094
Summer-autumn (-SCS) [a]-Surface	0.256***	0.335***	0.086	0.241**	0.323***	0.043
Summer-autumn-Surface	0.217***	0.129**	0.278***	0.204***	0.081^{+}	0.261***
Spring-Bottom	0.399**	0.316**	0.360***	0.271*	0.240*	0.305**
Summer-autumn (-SCS)-Bottom	0.311***	0.239***	0.193**	0.216**	0.155*	0.155*
Summer-autumn-Bottom	-0.005	0.101*	0.370***	-0.042	0.064^{+}	0.363***
Nano						
Spring-Surface	0.632***	0.404***	0.264***	0.547***	0.330***	0.191**
Summer-autumn (-SCS)-Surface	0.062	0.085^{+}	0.278**	0.053	0.024	0.268**
Summer-autumn-Surface	0.130*	0.011	0.087	0.128*	0.000	0.086
Spring-Bottom	0.491***	0.287**	0.534***	0.345**	0.095	0.454***
Summer-autumn (-SCS)-Bottom	0.142^{+}	0.269***	0.384***	0.043	0.140*	0.352***
Summer-autumn-Bottom	0.004	0.048	0.224**	-0.016	0.054	0.226***

a. Samples from the SCS in summer were removed;

b. Euclidean distance of the factor of water masses (WM) represented by temperature and salinity between sampling sites;

c. Euclidean distance of environmental factors (E) represented by all abiotic and biotic factors except temperature and salinity between sampling sites;

d. Euclidean distance of all spatial factors (S) between sampling sites;

e. Correlation between flagellate beta diversity and water mass controlled by E and S;

f. Correlation between flagellate beta diversity and environmental factors controlled by WM and S;

g. Correlation between flagellate beta diversity and spatial factors controlled by WM and E.

Significance: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; $^{+}$, $p < 0.1$.

Fig. S1 Spatiotemporal difference of environmental variables (a) and pico- and nano-sized microbial flagellate communities (b) displayed by principal component analysis (PCA) and non-metric multidimensional scaling (NMDS) analysis, respectively.

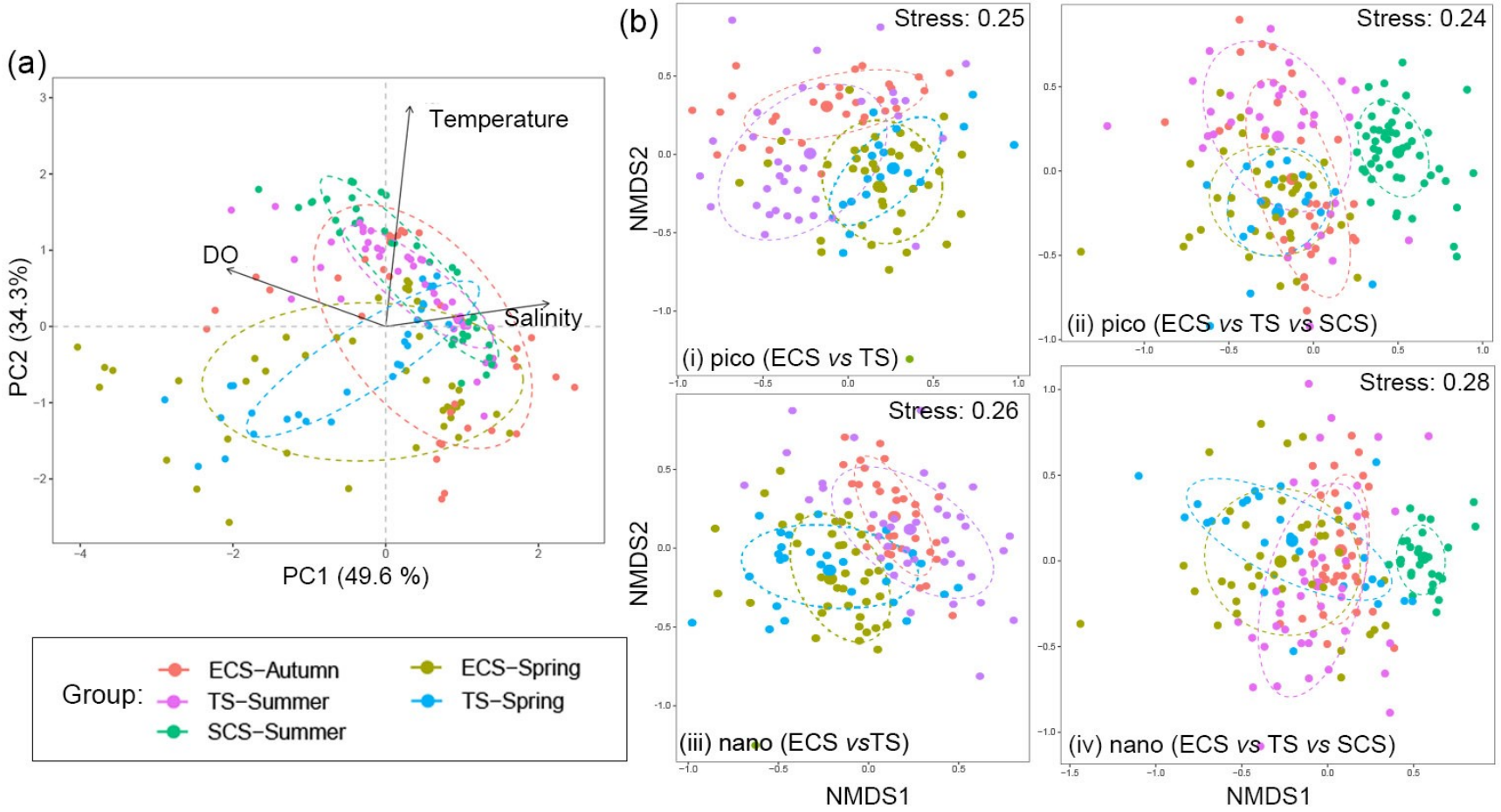

Fig. S2 Rarefaction curves (a), species accumulation curves (b), and Good's Coverage index (c) before and after sample standardizing for pico- and nano-datasets in this study; numbers of sequencing reads (d) of each sample before subsampling for pico- and nano- datasets; linear relationships (e) between the Bray-Curtis distance of original OTUs and subsampled OTUs for pico- and nano-datasets

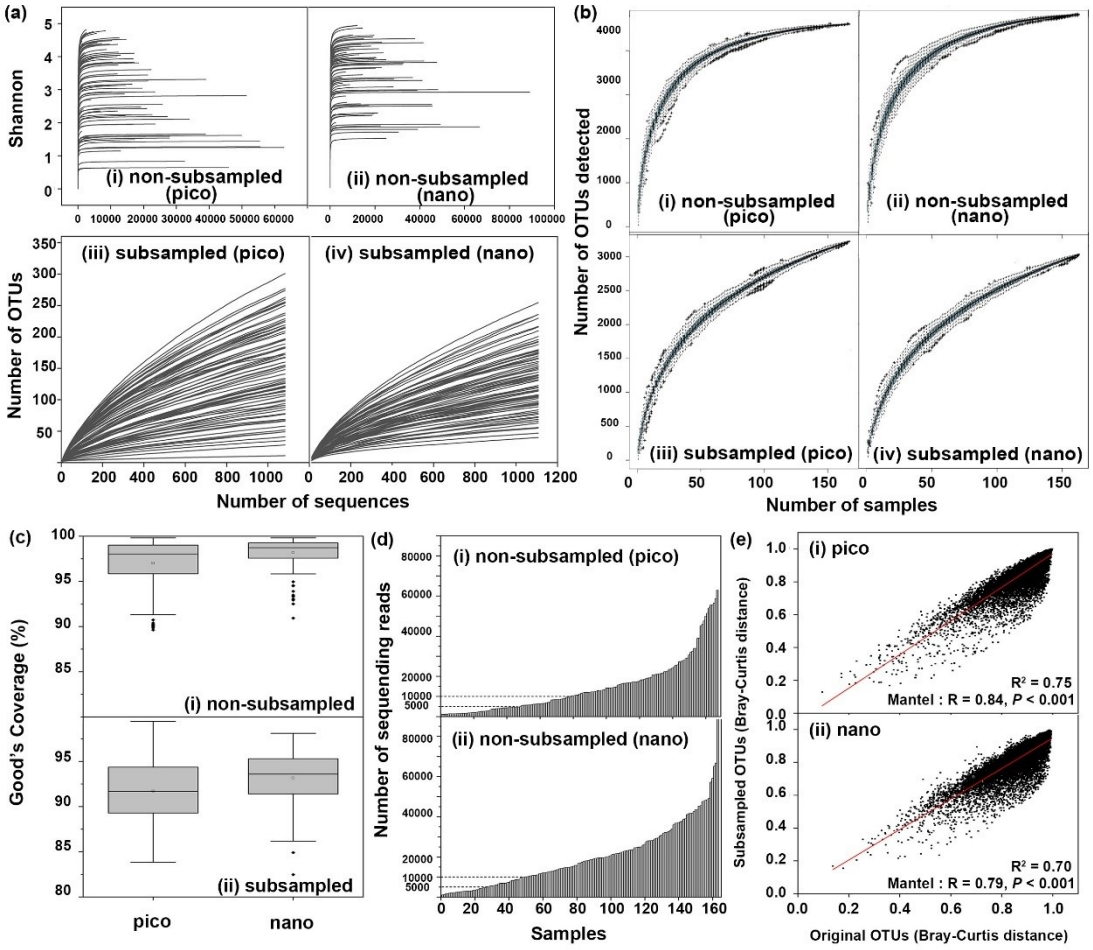

Fig. S3 Spatial distribution of temperature [shading in °C, (a, c)] and salinity [shading in PSU, (b, d)] in the East China Sea (ECS) and Taiwan Strait (TS) during spring (a, b) and summer-autumn (c, d) in 2019.

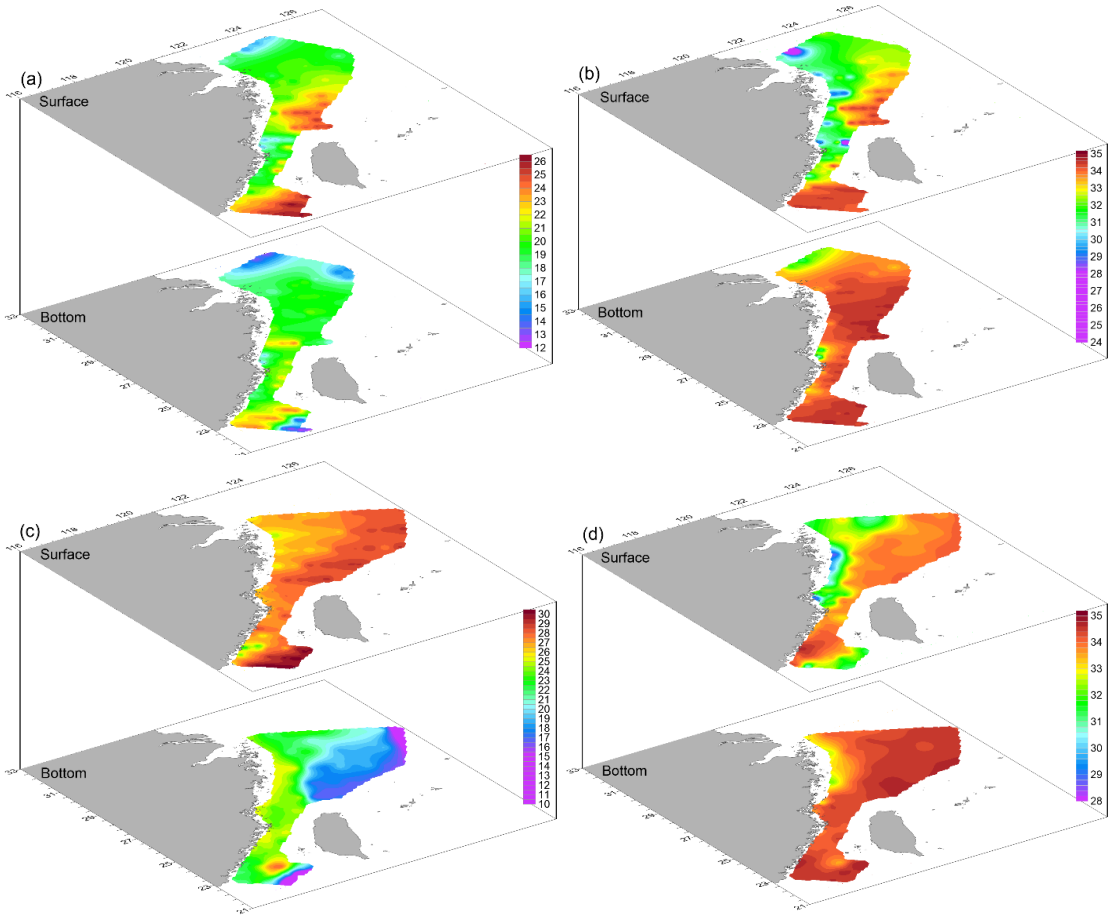

Fig. S4 Water mass classification (a, b) and their temperature-salinity diagrams (c, d) in the studied area of the ECS and TS in spring (a, c) and summer-autumn (b, d) in 2019. Hollow dots in the map represent all sampling spots during the cruises that had conductivity-temperature-depth (CTD) data used to classify the water mass. Light green: Min-Zhe coastal water (**Mc**); dark green: Zhujiang River diluted water (**Zd**); blue: East China Sea surface water (**Es**); orange: Kuroshio surface water (**Ks**); rose red: Kuroshio surface-subsurface mixed water (**Km**); purple: Kuroshio subsurface water (**Ku**); yellow: Coastal mixed water (**Cm**); pink: South China Sea surface water (**Ss**); red: South China Sea surface-subsurface mixed water (**Sm**); and black: South China Sea subsurface-intermediate mixed water (**Si**).

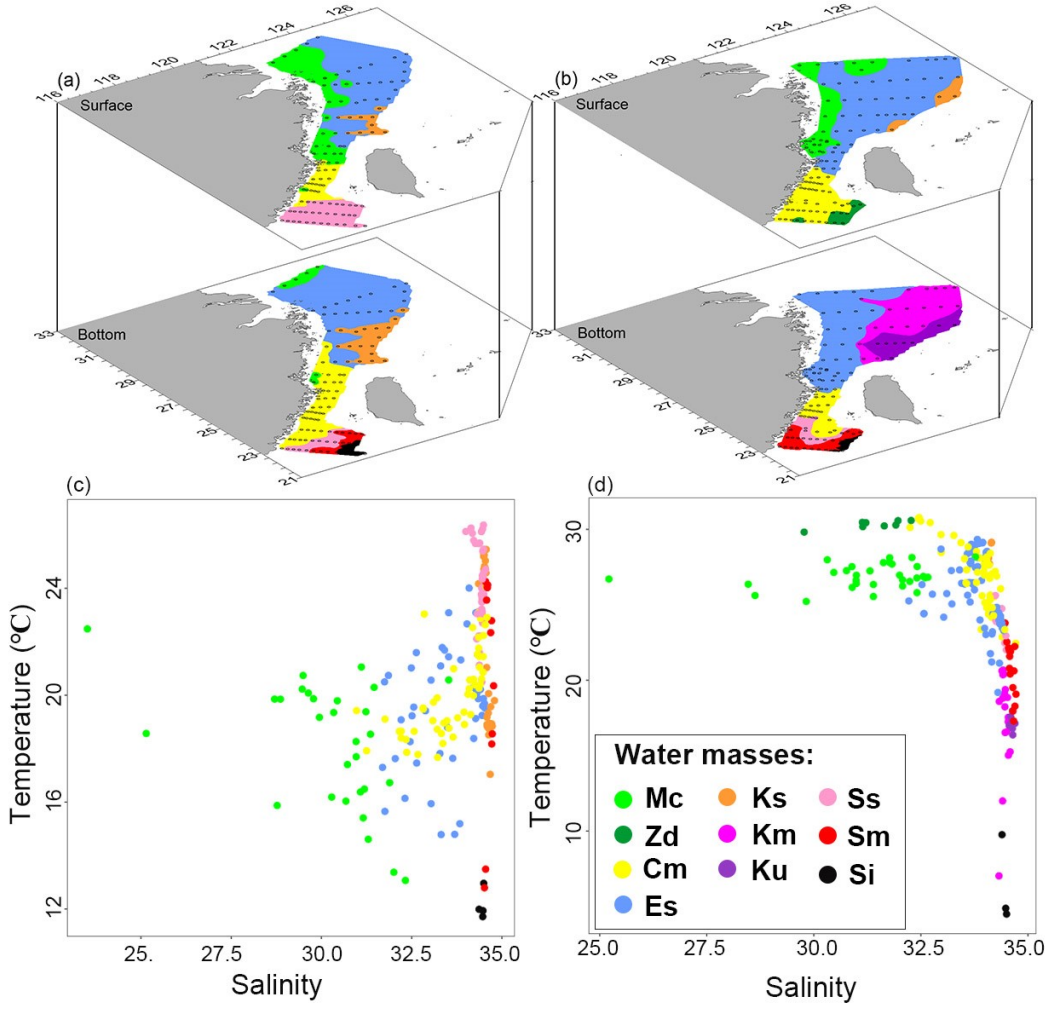

Fig. S5 Spatial distribution of temperature and salinity in three layers (surface, deep Chlorophyll maximum (DCM), and the bottom of the photic zone (Bottom)) (a) in the northern South China Sea (SCS) during summer in 2017. Black dots in the map represent all sampling spots during the cruises that had conductivity-temperature-depth (CTD) data used to classify the water mass. **According to temperature-salinity diagrams (b) and principal component analysis (c. PCA), we distinguished three water masses in the SCS in summer in 2017 from the other three SCS water masses (Ss, Sm, and Si, shown in Fig. S2) in summer in 2019.** The water masses are as follows: South China Sea surface water (**Ss**), South China Sea surface-subsurface mixed water (**Sm**), South China Sea subsurface-intermediate mixed water (**Si**), South China Sea surface water of 2017 (**Ss-2017**), South China Sea subsurface water of 2017 (**Su-2017**), and Pacific Ocean subsurface water of 2017 (**Pu-2017**).

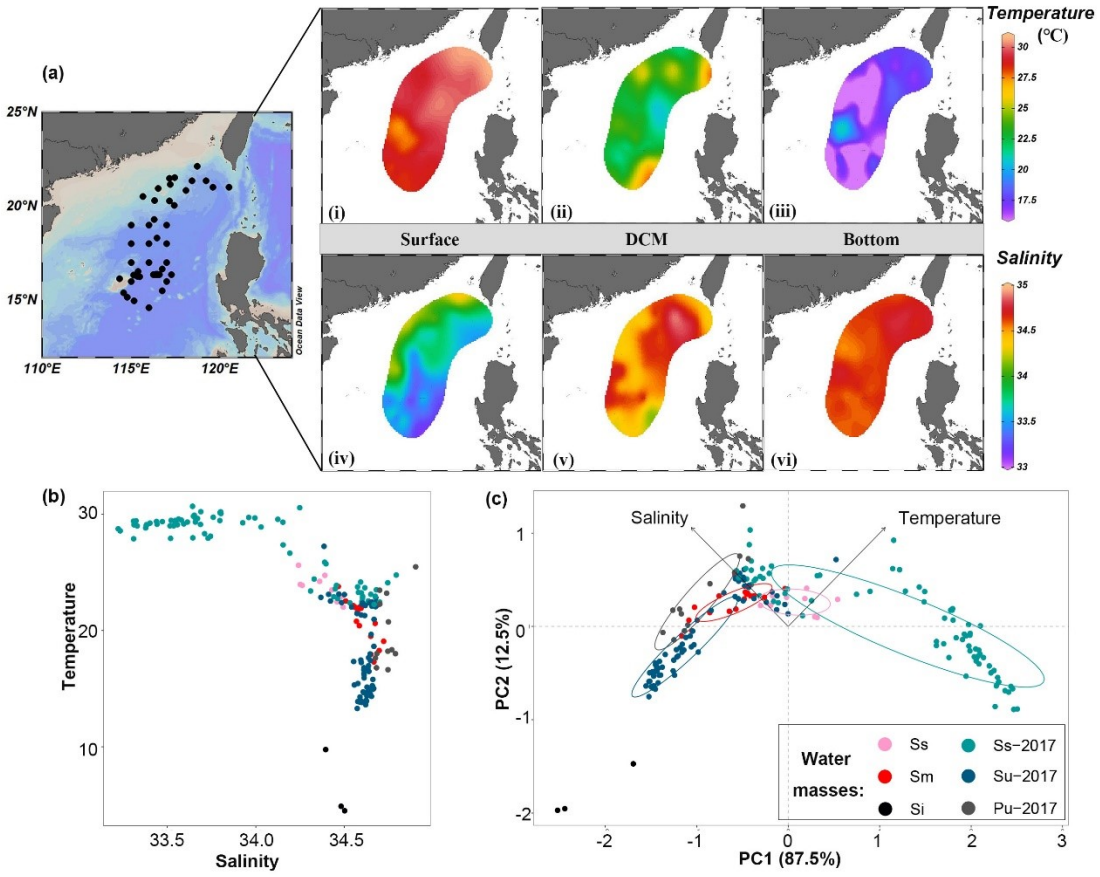

Fig. S6 Schematic diagrams of the distribution of water masses in spring (a) and summer-autumn (b).

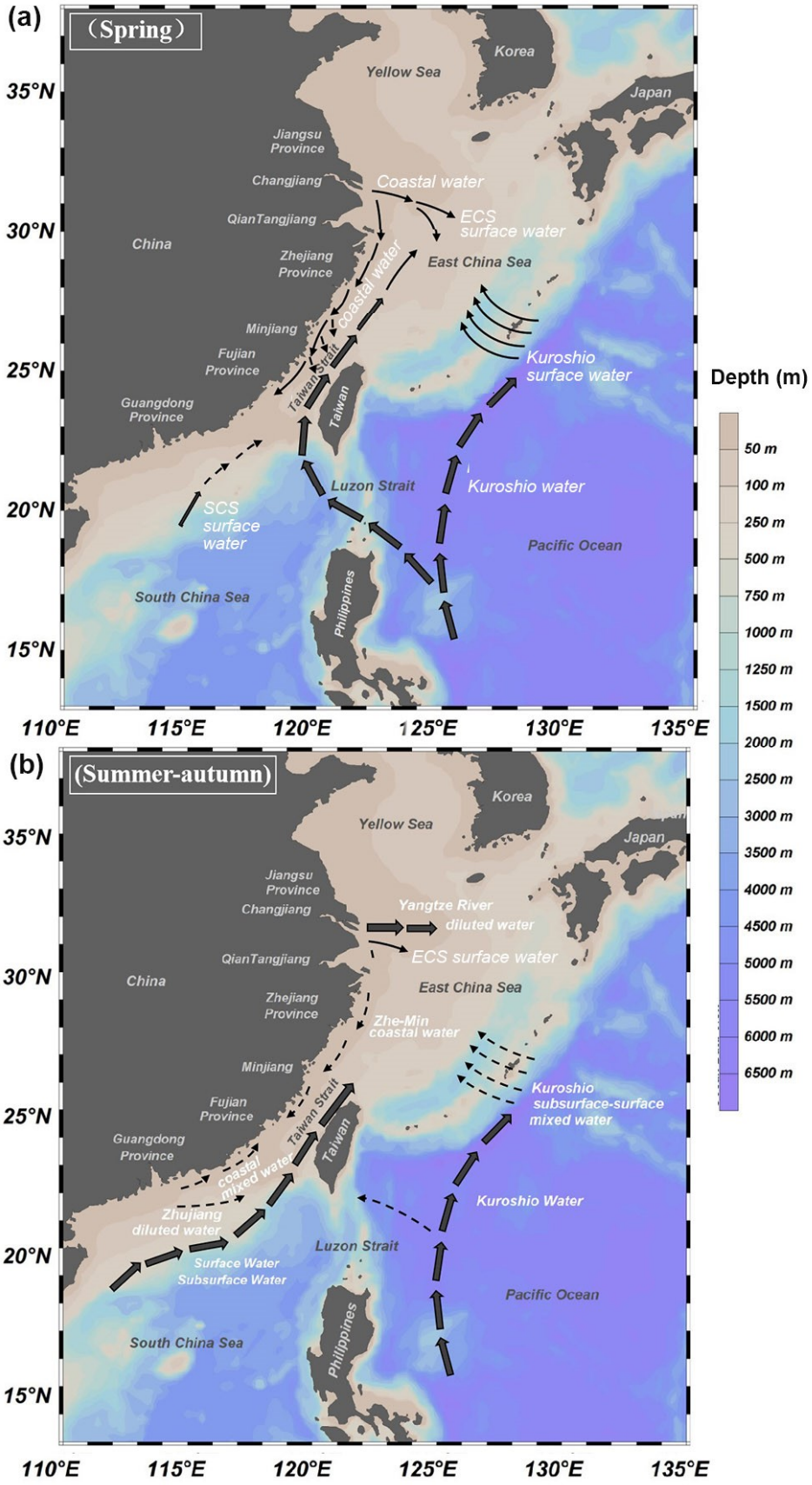

Fig. S7 Comparisons of richness (a. Chao 1) and diversity (b. Shannon) of the microbial flagellate communities. Different letters indicate significant differences at the level of $p < 0.05$. The hollow diamonds represented average values of individual index in each group.

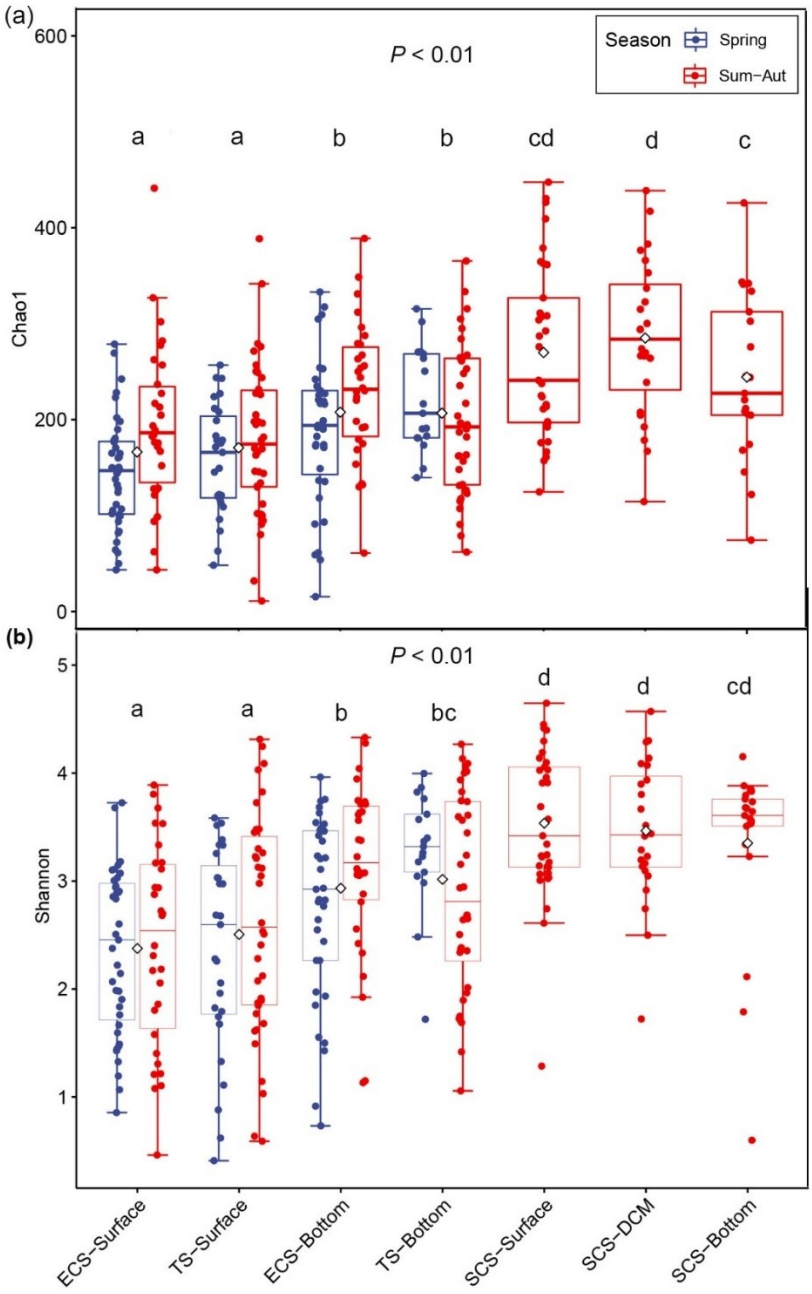

Fig. S8 Environmental drivers of Alpha diversity (represented by Chao 1 and Shannon) and composition of pico- and nano-sized microbial flagellate communities in all samples (a), all samples without SCS (b), spring (c), summer-autumn (d), and summer-autumn without SCS (e). Pairwise comparisons of environmental variables are shown with a color gradient denoting Spearman's correlation coefficient. The alpha diversity (Euclidean distance including Chao 1 and Shannon) and community composition (Bray-Curtis distance) were related to each environment variable (Euclidean distance) by Mantel tests. Only significant relationships ($p < 0.05$) based on 9999 permutations were shown.

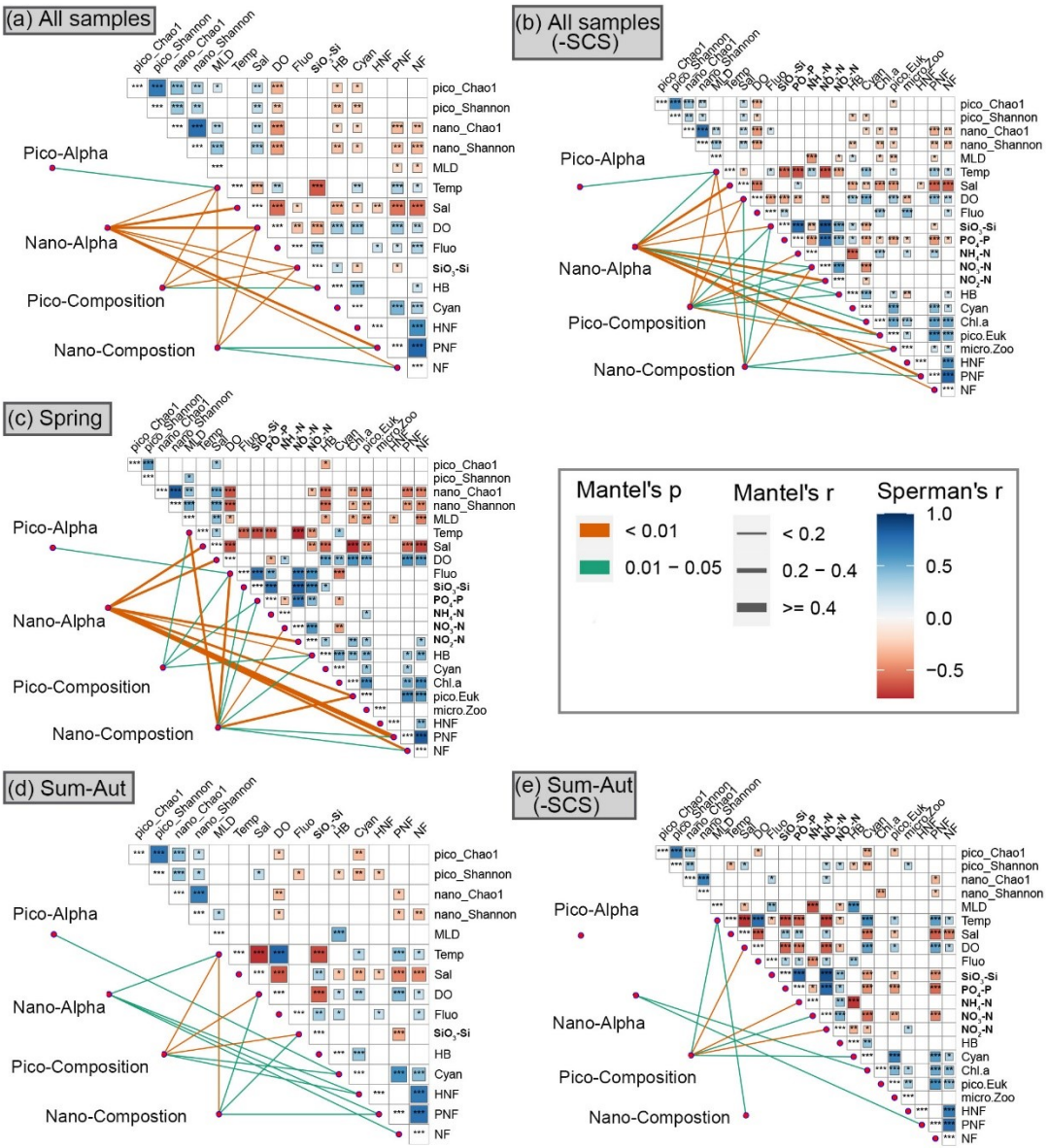

Fig. S9 Community compositions of the pico- and nano-sized microbial flagellates at supergroup level and lower taxonomic levels in the studied area. Different supergroups were showed in distinct colors while filling patterns represented lineages from lower taxonomic levels.

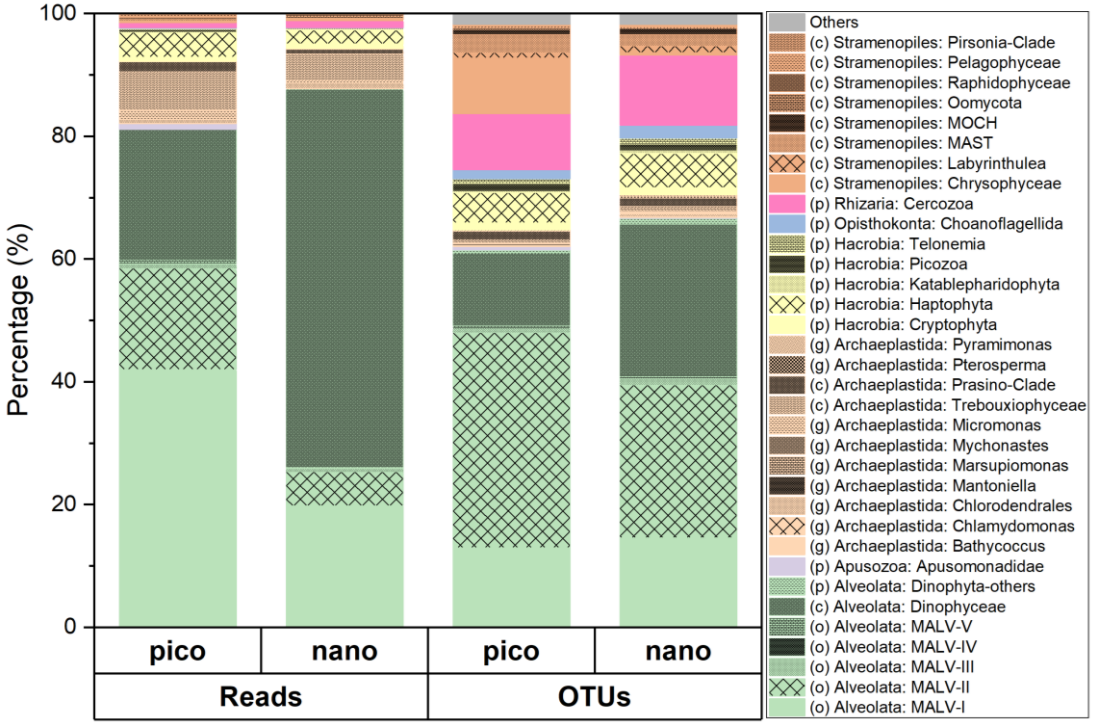

Fig. S10 Principal coordinate analysis (PCoA) showing the distributions of the pico-sized (a, b) and nano-sized (c, d) microbial flagellate communities in different water masses in spring (a, c) and summer-autumn (b, d).

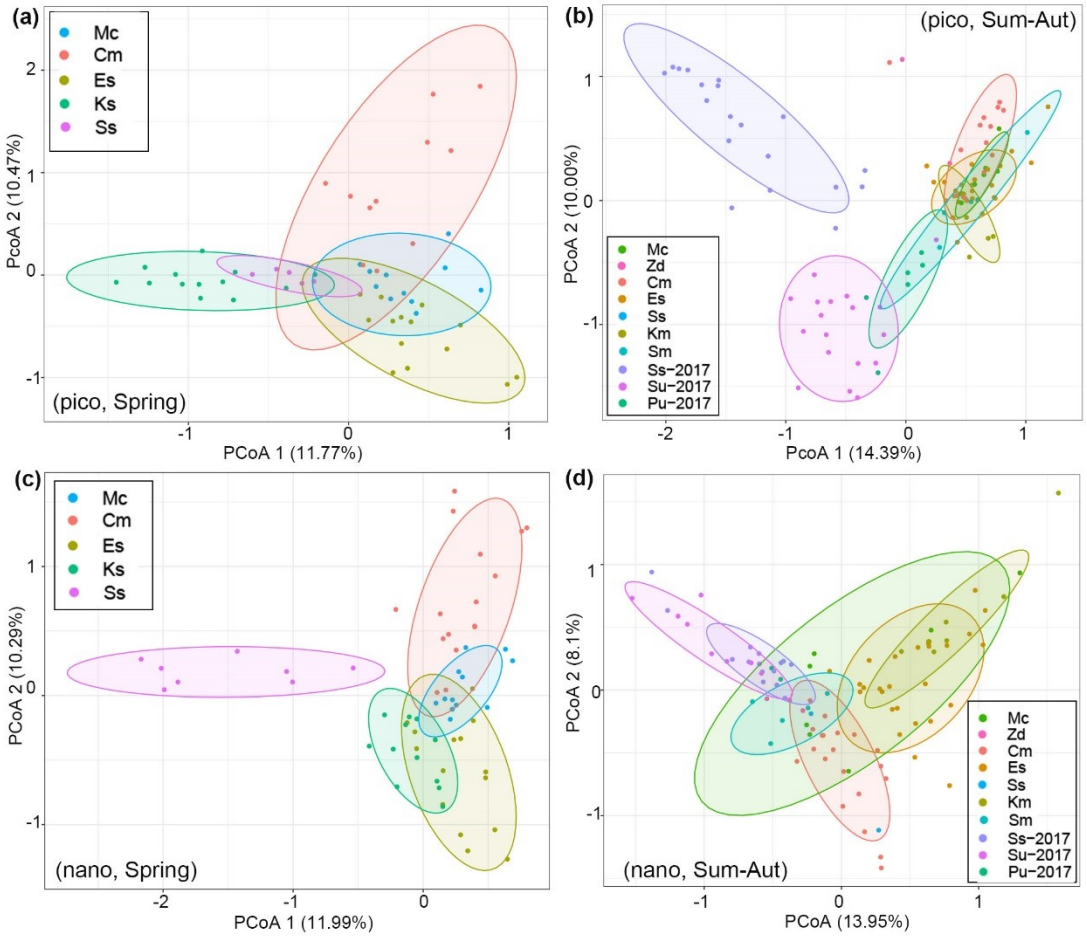

Fig. S11 Variation partitioning analyses (VPA) show the contributions of water masses (WM), environmental variables (E), and geographic distance (S) on the community structure of pico-sized (a) and nano-sized (b) microbial flagellate in the regional scale; Redundancy analyses (RDA) show the compositions of pico-sized (c) and nano-sized (d) microbial flagellate communities in relation to significant ($p < 0.01$) environmental and spatial (PCNM) variables in the regional area. Only the dominant taxa (percentage of reads > 0.5% within the entire dataset) in each community were marked.

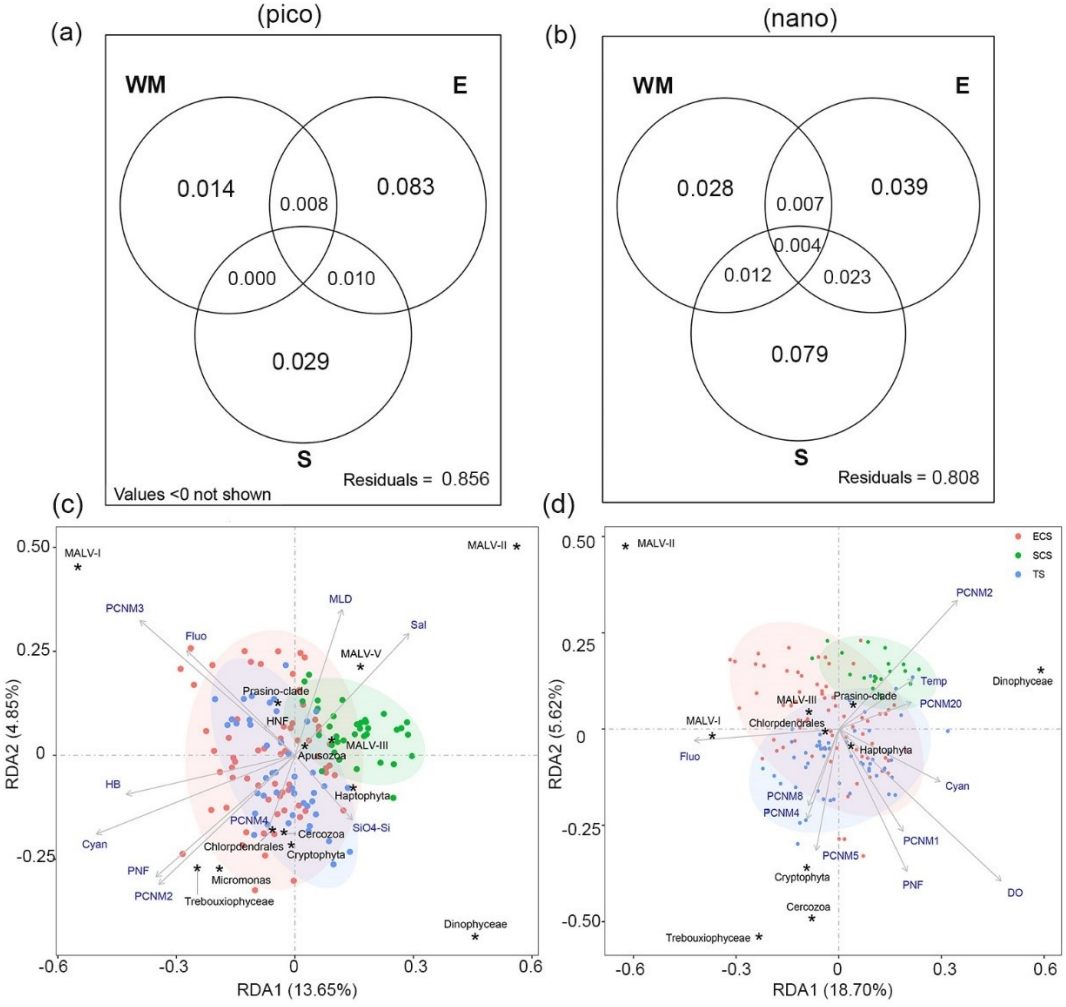

Fig. S12 Co-occurrence networks among OTUs of microbial flagellate communities in the regional scale. The nodes were colored according to different size fractions (a), taxonomic lineages (b), and modules (c). Connections between two individual nodes stand for strong (Spearman's $|r| > 0.6$) and significant ($p < 0.05$) correlation. The size of each node is proportional to the number of connections.