Supporting Information

The ecological responses of bacterioplankton during a *Phaeocystis globosa* bloom in Beibu Gulf, China highlighted by integrated metagenomics and metatranscriptomics

Sha Xu [a, c], Cheng He [a, c], Zhenjun Kang[e], Shuqun Song [a, b], Caiwen Li [a, b, d, *]

[a] CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

[b] Marine Ecology and Environmental Science laboratory, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266237, China

[c] University of Chinese Academy of Sciences, Beijing 100049, China

[d] Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China

[e] Guangxi Key Laboratory of Marine Disaster in the Beibu Gulf, Beibu Gulf University, Qinzhou 535011, China

***Corresponding author:** Dr. Caiwen Li

CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; Tel/Fax: 011-86-532-82898995; E-mail: cwli@qdio.ac.cn

Summary of the supporting information:

Additional details on material and methods

Fig. S1 Sampling area in Beibu Gulf (Guangxi Province, China)

Fig. S2 Bacterial variation of the bP and rP communities

Fig. S3. Taxonomic composition and active compositionof the bacterial community

Fig. S4 Taxonomic composition of the six communities

Fig. S5 Transcriptionally active taxa of the six communities

Fig. S6 Gene functions-associated taxa of the six communities

Fig. S7 Functional clustering analyses of the cDNA datasets of the bacterial samples

Fig. S8 The relative abundances distribution of key metabolite pathway-matched active bacterial taxa in bP and rP communities

Fig. S9 Relative abundances of transcripts associated with the EPS synthesis

Reference in the supporting information

Additional details on material and methods

Sampling and water chemistry

The water samples were collected from the surface waters (depth of 0.5-m) each station using large-capacity water sampler, immediately processed onboard after prefiltered through a 200-μm mesh to filter out mesoplankton and macroplankton and impurity particles. For bacterial RNA extraction, approximately 50-L of seawater was prefiltered through a 20-μm sieve and filtered through a 0.22-μm pore size polycarbonate filter (142-mm diameter, Millipore, Germany) successively using a peristaltic pump. The duration of the filtration step was no longer than 20 min to minimize RNA degradation. For DNA extraction of bacterial community analysis, additional 20-L seawater was simultaneously filtered (triplicate for each sample), and processed similarly to extract RNA. All filters were flash-frozen in liquid nitrogen and stored at -80℃ for further processing.

Seawater temperature (T) and salinity (S) were surveyed on site (depth of 2-m) with a shipborne conductivity-temperature-depth logger (CTD, RBR maestro, Canada). Chl *a* was extracted from the fiberglass filters (GF/F, 25mm diameter) with 90% acetone for 24h and was measured by spectrophotometry (reference). Other environmental parameters including dissolved oxygen (DO), total nitrogen (TN), ammonium salt (NH_4^+), nitrate plus nitrite (NO_2^-, NO_3^-), soluble phosphate (PO_4^{3-}), and silicate (SiO_3^-) were collected and measured according to the China Marine Survey Code

(GB/T12763.4−2007). The major nutrients were analyzed with a nutrient continuous flow analyzer (San Plus, SKALAR; Table 1).

Bacteria enumeration was performed by using a flow cytometer (Accuri C6, Becton, Dickinson and Company, the United States) after stained with SYBR Green I (nucleic acid dye, final concentration of 10^{-4}, v v^{-1}) for 20 min at room temperature (Zhao, 2010). Phytoplankton samples were trawled from the water bottom to the water surface. 1-L seawater was assessed on site before disintegration by fixation and sonification. The number of *P. globosa* colonies and the diameter of colonies were observed and measured, separately (Rousseau et al. 1990).

Extraction and Purification of Environmental DNA and RNA

Environmental total DNA was extracted from the filter sandwich by using CTAB (cetyltrimethylammonium bromide) method (Doyle and Doyle 1990) and processed with a PowerSoil DNA kit (Mo-Bio, Carlsbad, CA) according to the manufacturer's instruction. Briefly, 3-mL of lysis buffer (100-mM NaCl, 10-mM Tris, 1-mM EDTA) with lysozyme (5 mg mL^{-1}) was added to the 0.22-μm PES filter upon thawing, followed by vortexing for 10 min to lyse attached cells with beads. After incubating at 37℃ for 30 min, proteinase K (0.5 mg mL^{-1}) and SDS (1%) were added into the filters, and incubated at 55℃ for 20 min, followed by a further incubation at 70℃ for 5 min. Lysate was extracted twice with phenol: chloroform: IAA (25:24:1) and once with chloroform: isoamyl alcohol (24:1) and then precipitated with ethanol. DNA was collected by centrifugation, resuspended in TE buffer and subjected to a final clean-up

step (QIAamp mini spin columns, Qiagen). DNA concentration and purity were monitored on 1% agarose gels.

Environmental total RNA was extracted using TRIzol Reagent Kit (Thermo Scientific, Shanghai, China), subsequently purified employing the RNase-free DNase I (TaKaRa) to eliminate residual genomic DNA, and concentrated using the RNeasy MinElute Kit (Qiagen). Then, RNA degradation and contamination were monitored on 1% agarose gels. RNA purity and concentration were checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) and Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA), respectively. RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologiuhes, CA, USA).

Sequencing, Processing and Analysis of 16S rRNA Datasets

To assess the community composition of bacterioplankton, a total amount of 10-μg DNA per sample was processed and analyzed by Illumina HiSeq sequencing of 16S rRNA gene amplicons. The V3-V4 region of the 16S rRNA genes was amplified with the primer set 341 F (5′-CCTAYGGGRBGCASCAG-3′)/806 R (5′-GGACTACHVGGGTWTCTAAT-3′) with the barcode. PCR reactions were performed in a 30-μL reaction with 15-μL of Phusion® High-Fidelity PCR Master Mix (New England Biolabs). The thermal cycling conditions were as follows: initial duration at 98 ℃ for 1 min, followed by 30 cycles of 10 s at 98 ℃, 30 s at 50 ℃ and 30 s at 72℃, with a final extension at 72℃ for 5 min. The amplified PCR products were visualized

by electrophoresis on 2% agarose gel containing the nucleic acid stain SYBR green 1 under UV light. Then, PCR products were purified with GeneJET^{TM} Gel Extraction Kit (Thermo Scientific, Shanghai, China) according to the manufacturer's instructions.

Sequencing libraries were generated using Ion Plus Fragment Library Kit 48 rxns (Thermo Scientific, Shanghai, China) following manufacturer's recommendations. The library quality was assessed on the Qubit@ 2.0 Fluorometer (Thermo Scientific, Shanghai, China). Quality control and the processing of Single-end sequence reads were conducted on the Ion S5^{TM} XL platform. After cutting off the primers, quality filtering on the raw reads (Martin 2011), deleting chimera sequences (Edgar 2017) and high-quality sequences in the average length of 371 nt were kept for subsequent analyses.

Processing and analysis of environmental metatranscriptomic samples

A total amount of 3-μg RNA per sample was used for Illumina high-throughput sequencing (RNA-seq). Sequencing libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) according to the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. The cDNA library was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X) followed by the NEBNext mRNA Second Strand Synthesis Modul. After the library was constructed, preliminary quantification was performed using Qubit 2.0

117 Fluorometer, and the library was diluted to 1.5 ng μL^{-1}. The library fragments were
118 subsequently assessed on the Agilent Bioanalyzer 2100 system to select cDNA
119 fragments of preferentially 250~300 nt in length. At last, clonally amplified library
120 fragments were sequenced on Illumina HiSeq 2000 platform and 150 nt paired-end
121 reads were generated.

122 With RNA-seq, after quality filtering by removing reads containing adapter, reads
123 containing ploy-N and low-quality reads from raw data, we obtained an average of
124 35,568,957 paired reads per sample (Table S3). Ribosomal RNA reads were filtered out
125 employing SortMeRNA. The remaining non-ribosomal sequences were assembled de
126 novo with the splicing software Trinity (version: r20140413p1). The sequences were
127 integrated and eliminated redundancy using CD-HIT-EST with the sequence identity
128 threshold of 0.95, and obtained 1,585,720 unigenes for further analyses (Table S3).

129 For taxonomic classification of protein-coding gene sequences, the cDNA sequences
130 were identified by BLAST against the NCBI non-redundant protein database (Nr,
131 Version: 2016-11-05) using DIAMOND software (Conesa et al. 2016) with pair (blastp,
132 evalue $\leqslant 1e^{-5}$). Alignment with bacterial and archaeal protein-coding gene sequences
133 extracted from Nr (the LCA algorithm), using the system classification of the MEGAN
134 software (Huson et al. 2011) was parsed to analyze the taxonomic information, and an
135 average of 314,123 bacterial/archaea sequences were obtained (Table S3).

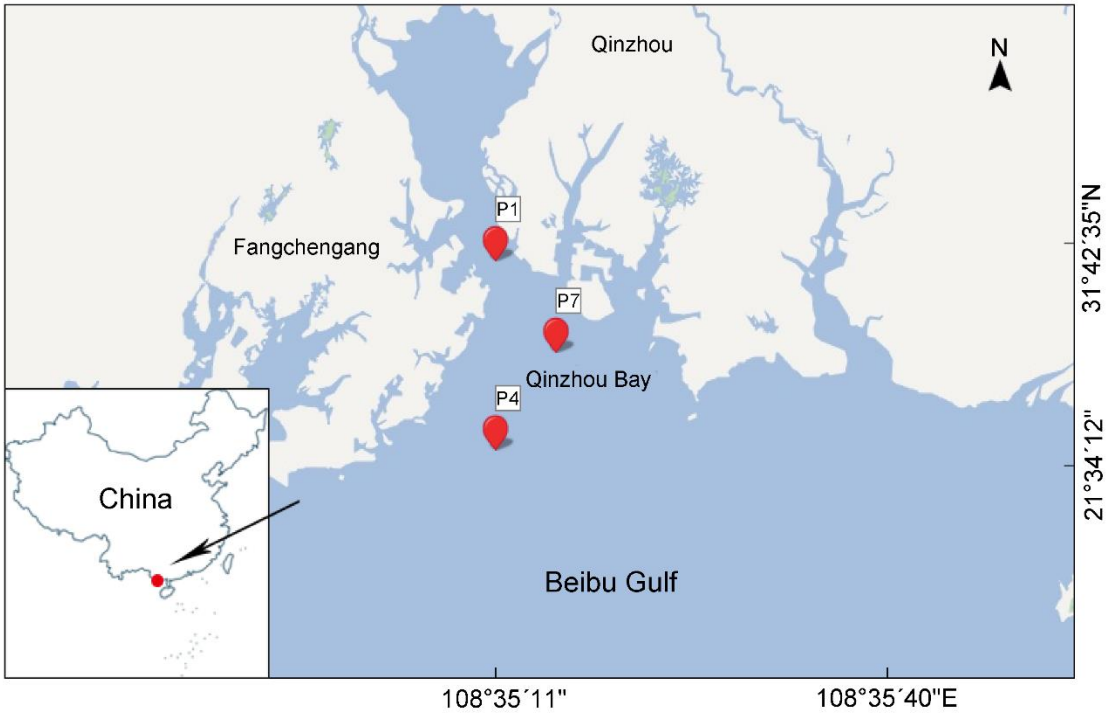
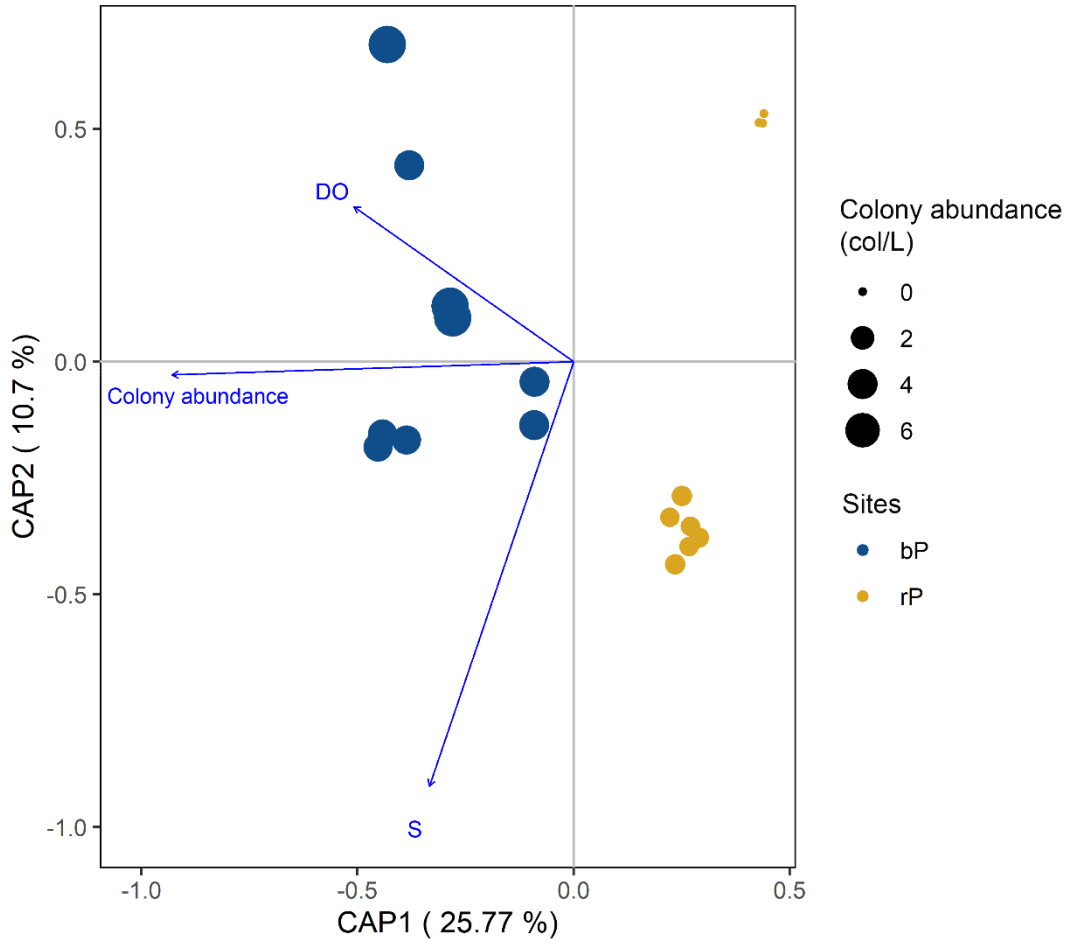
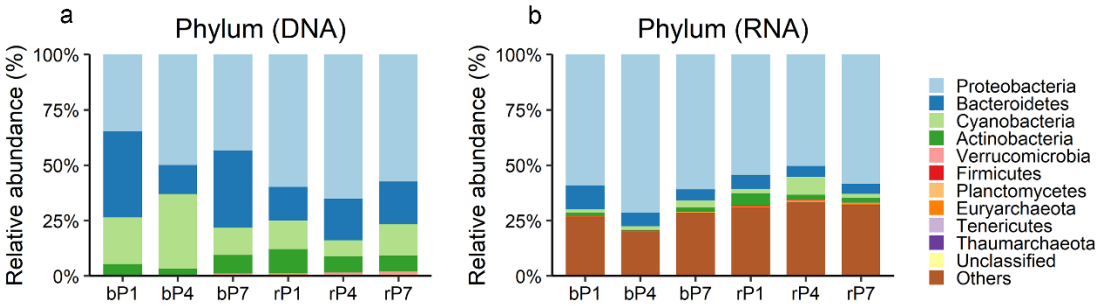


Fig. S1 Sampling area in Beibu Gulf (Guangxi Province, China). Red marks indicate the sampling stations.



139

140 **Fig. S2** Bacterial variation of the bP (blooming period) and rP (recession period)
141 communities. Constrained analysis of principal coordinates (CAP) based on Bray-
142 Curtis distance and environmental variables indicated physicochemical characteristics
143 that influenced the assembly of bP and rP communities. The size of each point is
144 proportional to the colony abundance of *P. globosa*. Only statistical significance of
145 explanatory variables was shown ($P < 0.05$; perMANOVA test).



146

147 **Fig. S3** Taxonomic composition (base on DNA) and active composition (base on RNA)

148 of the bacterial community at the phylum between bP (blooming period) and rP

149 (recession period).

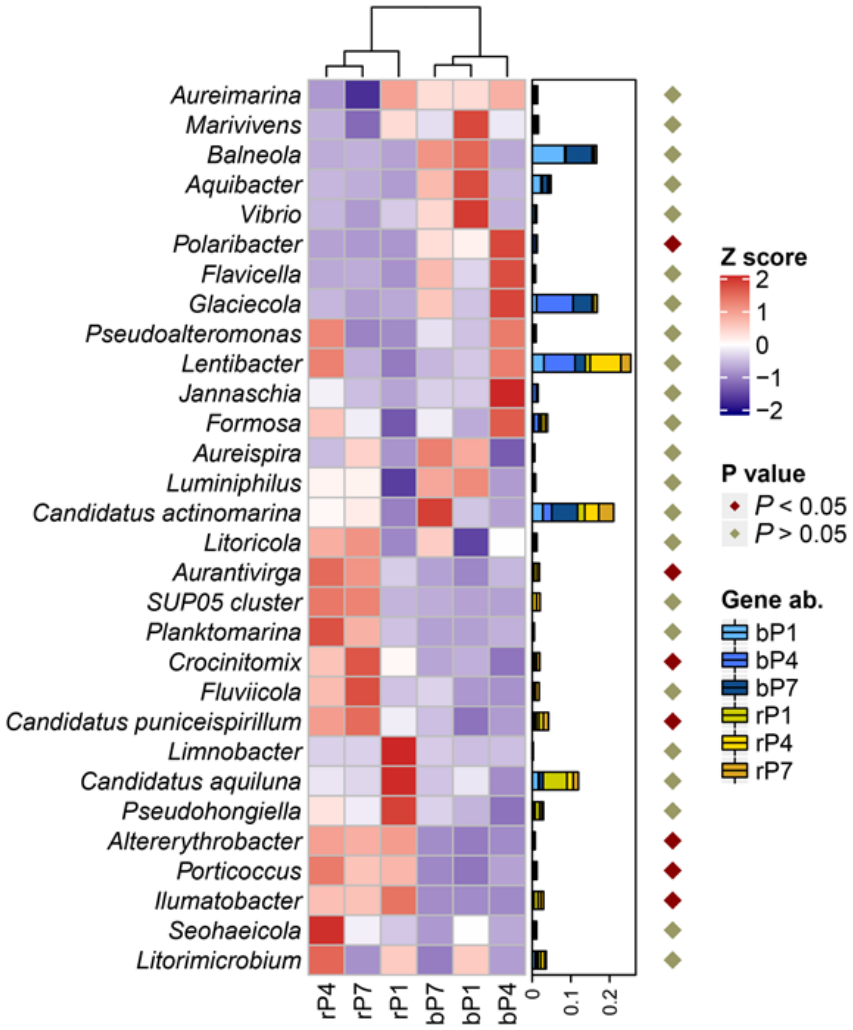


Fig S4 Taxonomic composition of the six communities. The stacked-Column (in the middle) showing the relative abundances of OTUs assigned to *Silva* taxonomy of microbial taxa (only top 30 are shown). Hierarchical clustering was performed based on a correlation matrix (with spearman correlation coefficient) of relative abundances. Heatmaps scaled and centred by row (Z score). The Z score is brighter red or blue coloring indicates higher or lower signal intensities. The red solid diamond indicates significantly change (by ANOVA analysis) of relative abundances between bP and rP.

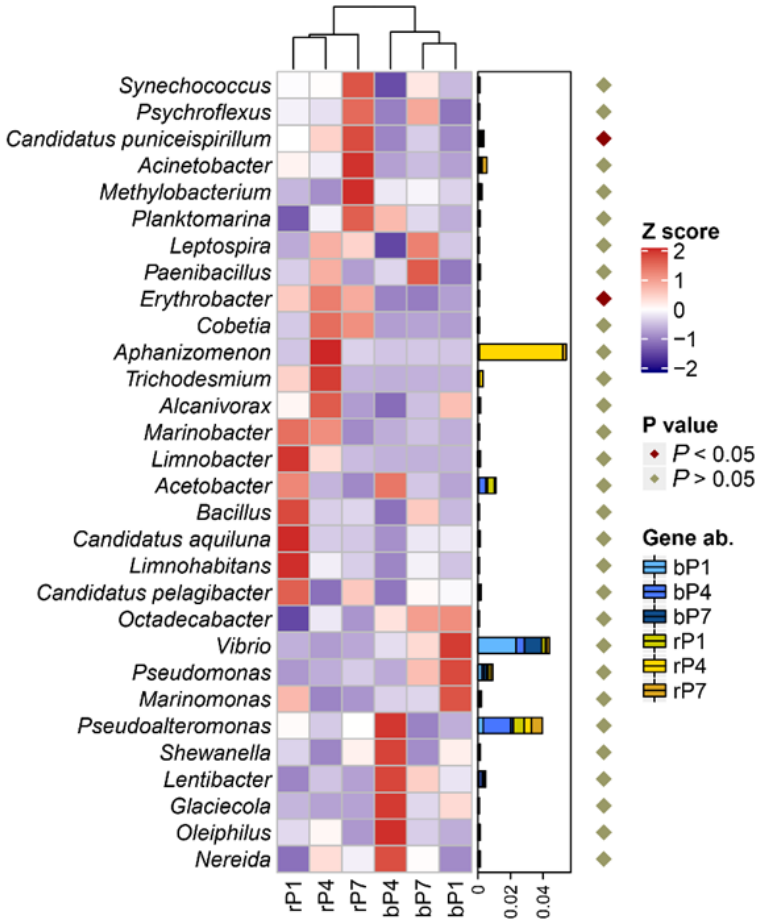
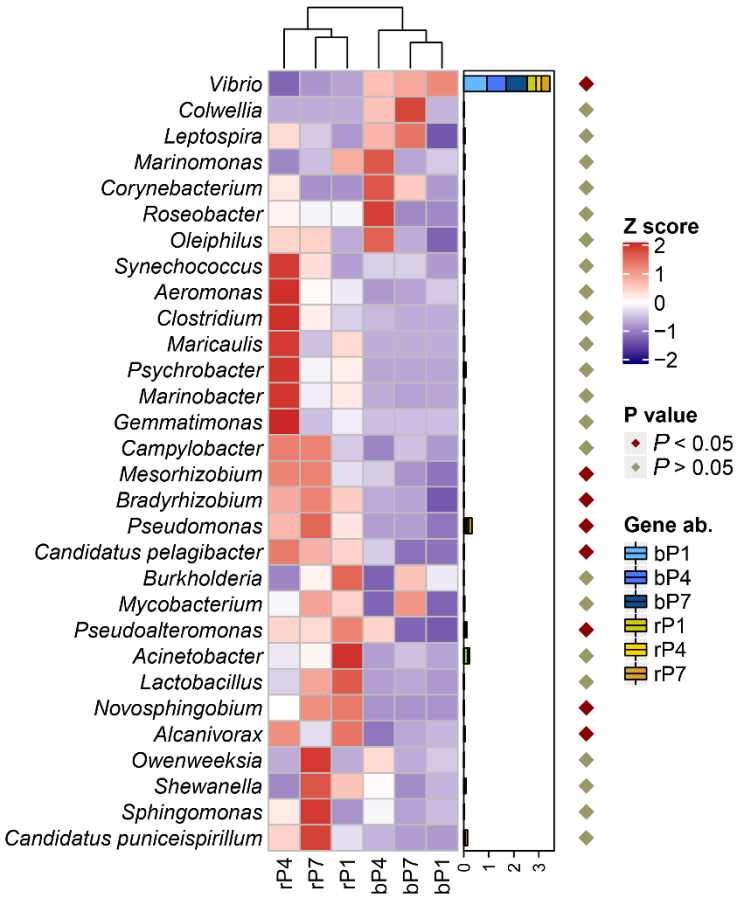


Fig S5 Transcriptionally active taxa of the six communities. The stacked-Column (in the middle) showing the relative abundances of protein-coding gene sequences assigned to NCBI taxonomy of microbial taxa (only top 30 are shown). Hierarchical clustering was performed based on a correlation matrix (with spearman correlation coefficient) of relative abundances. Heatmaps scaled and centred by row (Z score). The Z score is brighter red or blue coloring indicates higher or lower signal intensities. The red solid diamond indicates significantly change (by ANOVA analysis) of relative abundances between bP and rP.



167

168 **Fig S6** Gene functions-associated taxa of the six communities. The stacked-Column (in

169 the middle) showing the relative abundances of significantly different genes assigned

170 to NCBI taxonomy of microbial taxa (only top 30 are shown). Hierarchical clustering

171 was performed based on a correlation matrix (with spearman correlation coefficient) of

172 relative abundances. Heatmaps scaled and centred by row (Z score). The Z score is

173 brighter red or blue coloring indicates higher or lower signal intensities. The red solid

174 diamond indicates significantly change (by ANOVA analysis) of relative abundances

175 between bP and rP.

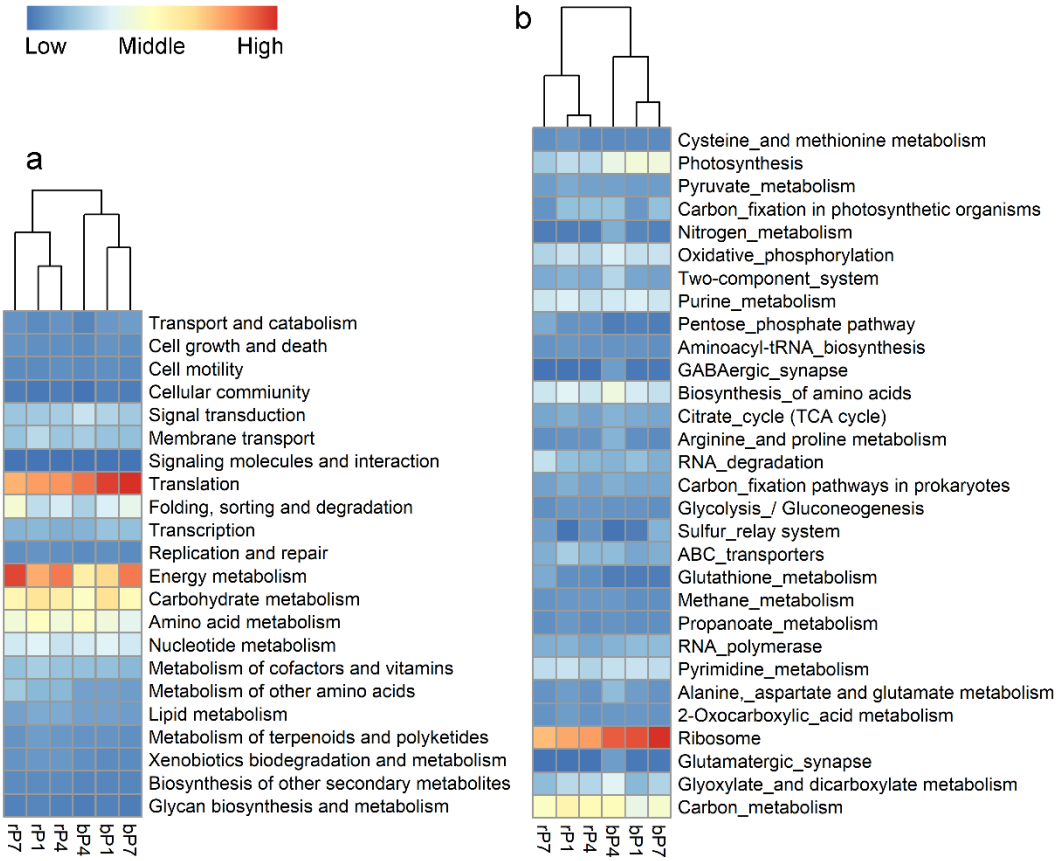
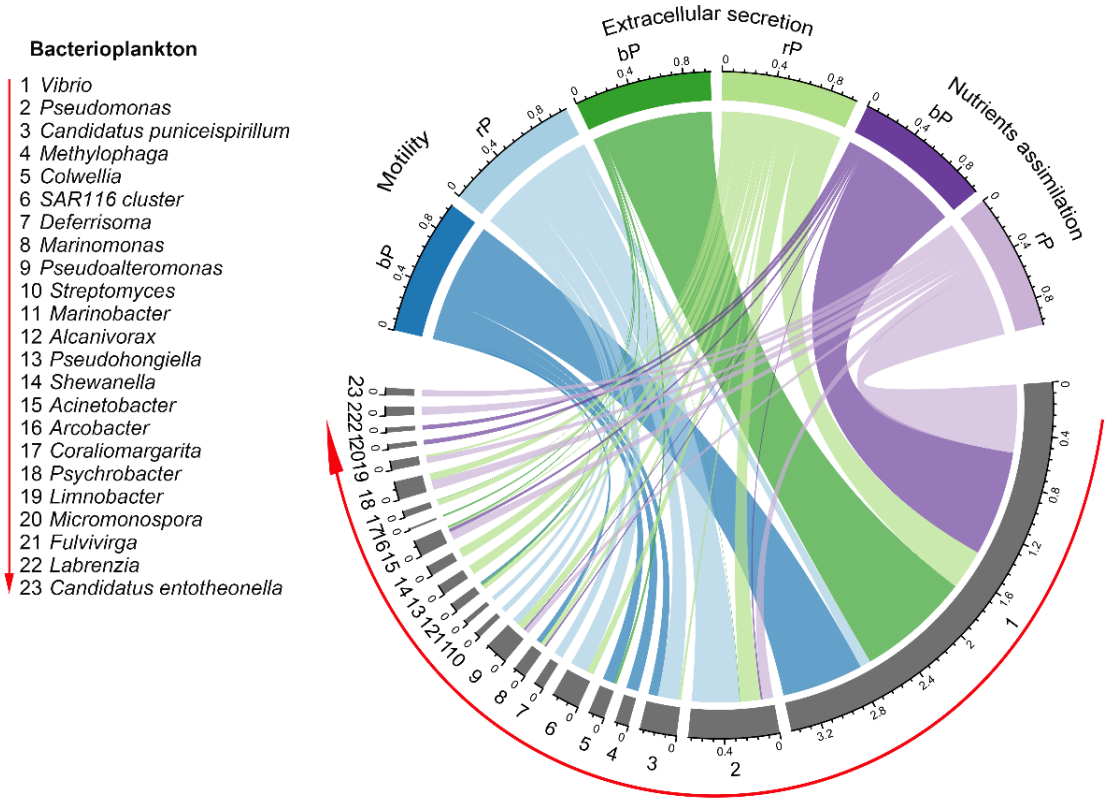
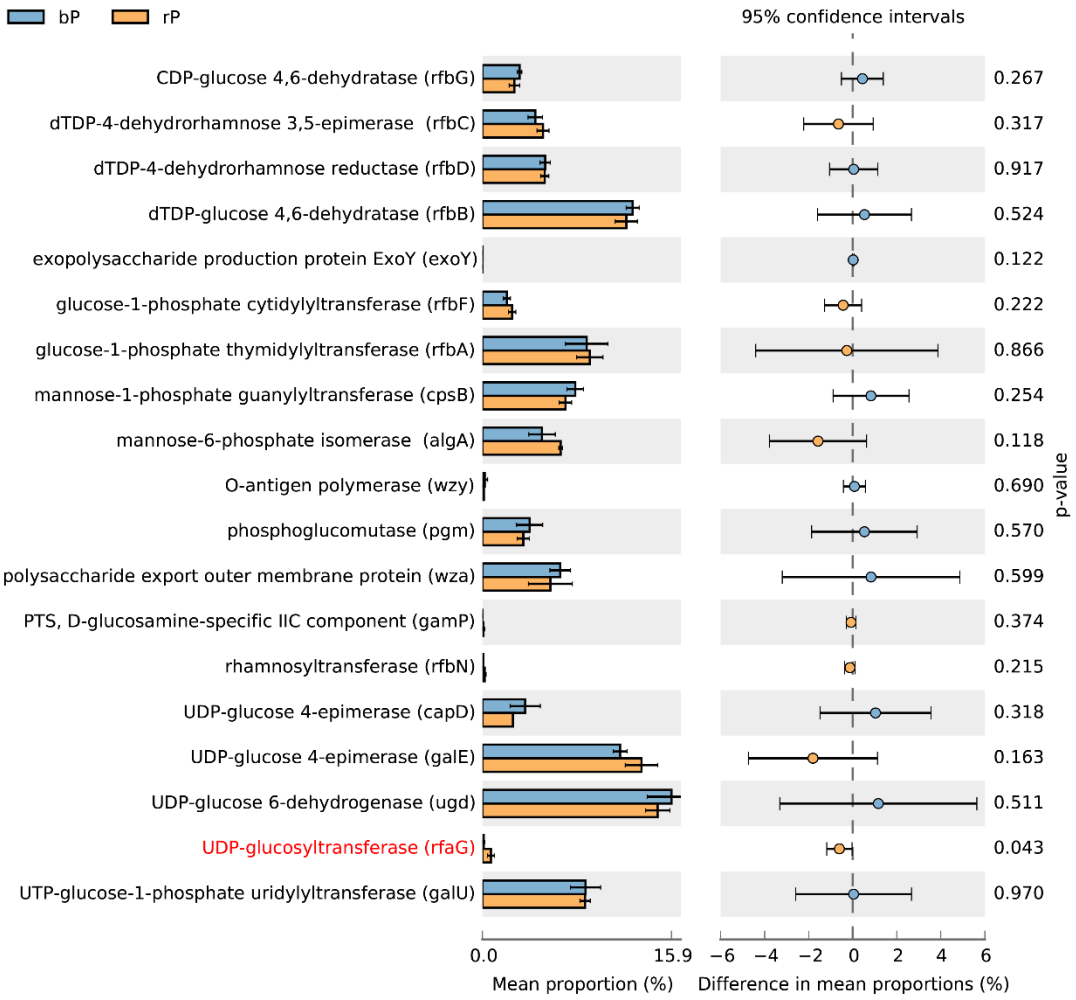


Fig S7 Functional clustering analyses of the cDNA datasets of the bacterial samples. Hierarchical clustering based on their relative abundance of protein-coding gene sequences assigned to (a) KEGG subcategories and (b) KEGG pathways (relative abundance top 30), was performed with complete linkage method based on a correlation matrix (1- Pearson correlation coefficient).



182

183 **Fig. S8** The relative abundances distribution of bacterial taxonomic transcripts
184 matching with cell motility, two-component system, and membrane transport in bP
185 (blooming period) and rP (recession period) communities. The chord diagram arranges
186 the nodes radially (the upper nodes and lower nodes represents key metabolic pathway
187 and bacterioplankton, respectively), drawing thick curves between nodes. Each
188 bacterial taxon has a link for each a metabolic pathway in bP or rP communities, and
189 the thicker the links, the more relative abundances distribution of bacterial transcripts
190 in this metabolic pathway. See Table S5-S8 for detailed information of the key
191 metabolite pathway-matched active bacterial taxa.



192

193 **Fig. S9** Relative abundances (%) of transcripts associated with the *in situ* functional

194 activities of exopolysaccharide (EPS) synthesis in bP (blooming period) and rP

195 (recession period) community. Red color showed the $P < 0.05$.

196 **Reference**

197 Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, Mcpherson A,
198 Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) Erratum to:
199 a survey of best practices for RNA-seq data analysis. Genome Biol 17:181.
200 https://doi.org/10.1186/s13059-016-1047-4

201 Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13−15.

202 Edgar RC (2017) SEARCH-16S: A new algorithm for identifying 16S ribosomal RNA
203 genes in contigs and chromosomes. BioRxiv doi: https://doi.org/10.1101/124131.

204 Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis
205 of environmental sequences using Megan4. Genome Res 21:1552−1560.
206 https://doi.org/10.1101/gr.120618.111

207 Martin M (2011) Cutadapt removes adapter sequences from high-throughput
208 sequencing reads. Embnet J 17. https://doi.org/10.14806/ej.17.1.200

209 Rousseau V, Mathot S, Lancelot C (1990) Calculating carbon biomass of *Phaeocystis*
210 sp. from microscopic observations. Mar Biol 107:305−314.
211 https://doi.org/10.1007/BF01319830

212 Zhao Y (2010) Spatial and temporal distribution of pieoplankton in Yellow Sea and
213 East China Sea. PhD thesis. Qingdao: Ocean University of China.