Water Research 216 (2022) 118281

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

# Metagenomic evidence for the microbial transformation of carboxyl-rich alicyclic molecules: A long-term macrocosm experiment

Changfei He [a,b], Jihua Liu [a,b,c,*], Rui Wang [a,b], Yuanning Li [a,b], Qiang Zheng [b,d], Fanglue Jiao [b,f], Chen He [g], Quan Shi [g], Yongle Xu [a,b], Rui Zhang [b,d], Helmuth Thomas [b,f,h], John Batt [b,f], Paul Hill [f], Marlon Lewis [f], Hugh Maclntyre [f], Longfei Lu [d,i], Qinghua Zhang [d,j], Qichao Tu [a,b], Tuo Shi [a,b], Feng Chen [a,e], Nianzhi Jiao [b,c]

[a] Institute of Marine Science and Technology, Shandong University, Qingdao 266237, China
[b] Joint Laboratory for Ocean Research and Education at Dalhousie University, Shandong University and Xiamen University, Halifax, NS, B3H 4R2, Canada, Qingdao 266237, China, and Xiamen 361005, China
[c] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangzhou 510000, China
[d] State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Fujian Key Laboratory of Marine Carbon Sequestration, Xiamen University, Xiamen 361005, China
[e] Environmental Research Center, University of Maryland at Baltimore, United States
[f] Department of Oceanography, Dalhousie University, Halifax, NS B3H 4R2, Canada
[g] State Key Laboratory of Heavy Oil Processing, China University of Petroleum-Beijing, Beijing 102249, China
[h] Helmholtz-Center Geesthacht, Institute for Coastal Research, Max-Planck-Strasse 1, Geesthacht D-21502, Germany
[i] Weihai Changqing Ocean Science Technology Co., Ltd., Weihai, Shandong, China
[j] Marine Equipment Inspection & Testing Co. Ltd, China

## ARTICLE INFO

*Keywords:*
Aquatron tower tank
Refractory dissolved organic matter (RDOM)
Microbial metagenome
KEGG modules
CRAM production
FTICR MS

## ABSTRACT

Carboxyl-rich alicyclic molecules (CRAMs) widely exist in the ocean and constitute the central part of the refractory dissolved organic matter (RDOM) pool. Although a consensus has been reached that microbial activity forms CRAMs, the detailed molecular mechanisms remain largely unexplored. To better understand the underlying genetic mechanisms driving the microbial transformation of CRAM, a long-term macrocosm experiment spanning 220 days was conducted in the Aquatron Tower Tank at Dalhousie University, Halifax, Canada, with the supply of diatom-derived DOM as a carbon source. The DOM composition, community structure, and metabolic pathways were characterised using multi-omics approaches. The addition of diatom lysate introduced a mass of labile DOM into the incubation seawater, which led to a low degradation index ($I_{DEG}$) and refractory molecular lability boundary (RMLB) on days 1 and 18. The molecular compositions of the DOM molecules in the later incubation period (from day 120 to day 220) were more similar in composition to those on day 0, suggesting a rapid turnover of phytoplankton debris by microbial communities. Taxonomically, while Alpha proteobacteria dominated during the entire incubation period, Gamma proteobacteria became more sensitive and abundant than the other bacterial groups on days 1 and 18. Recalcitrant measurements such as $I_{DEG}$ and RMLB were closely related to the DOM molecules, bacterial community, and Kyoto encyclopaedia of Genes and Genomes (KEGG) modules, suggesting close associations between RDOM accumulation and microbial metabolism. KEGG modules that showed strong positive correlation with CRAMs were identified using a microbial ecological network approach. The identified KEGG modules produced the substrates, such as the acetyl-CoA or 3-hydroxy-3-methylglutaryl-CoA, which could participate in the mevalonate pathway to generate the precursor of CRAM analogues, isopentenyl-PP, suggesting a potential generation pathway of CRAM analogues in bacteria and archaea. This study revealed the potential genetic and molecular processes involved in the microbial origin of CRAM analogues, and thus indicated a vital ecological role of bacteria and archaea in RDOM production. This study also offered new perspectives on the carbon sequestration in the ocean.

* Corresponding author.
*E-mail address:* liujihua1982@foxmail.com (J. Liu).

https://doi.org/10.1016/j.watres.2022.118281
Received 1 December 2021; Received in revised form 7 March 2022; Accepted 10 March 2022
Available online 15 March 2022

## 1. Introduction

Dissolved organic matter (DOM) is the largest reservoir of organic carbon in the ocean and is approximately equivalent to the reservoir of atmospheric $CO_2$ (Ogawa and H., 2001; Quéré et al., 2014). Approximately 94% of DOM is refractory DOM (RDOM), which has been sequestered in the ocean for thousands of years (Dittmar et al., 2008; Hansell et al., 2009). Owing to its large inventory and long-term stability, the origin and fate of RDOM in the ocean impacts global carbon cycling. Microbial activity is largely responsible for RDOM production in the ocean (Jiao et al., 2018; Kaiser and Benner, 2008; Refractory, 2011). For example, previous studies have shown that bacteria respond actively to the different additions, such as cyanobacterial DOM (Zhao et al., 2019), *Synechococcus* lysates (Zheng et al., 2021), glucose (Lechtenfeld et al., 2015), and sediment extraction (Lian et al., 2021; Wu et al., 2018), suggesting a correlation between bacterial community composition and RDOM generation. In addition, particulate organic matter (POM) is the second largest organic carbon pool in the ocean, and can descend from surface waters to the seabed (Eppley et al., 1979). POM can be respired by microbial activities during the sinking process, and only a small fraction of POM escapes mineralization and reaches the sedimentary layer (Ducklow et al., 2001). Microbial activity is closely related to the carbon cycling in ocean.

DOM is a large molecular mixture in aquatic environments (Shen and Benner, 2020; Zhang, 2017). Owing to its molecular complexity, a full structural elucidation of DOM has been unattainable to date (Zark et al., 2018). Carboxyl-rich alicyclic molecules (CRAMs), defined as a diverse array of organic compounds enriched with carboxylated and fused alicyclic rings, are the most abundant components of deep-ocean DOM that have been characterised and constitute a major part of the RDOM pool in the ocean (Hertkorn et al., 2006). CRAM account for 64–68% of the bulk DOM below 200 m in the Mariana Trench (Li et al., 2019). CRAMs are derived from biomolecules with structural similarities to sterols and terpenoids (Arakawa et al., 2017; Woods et al., 2012). Isopentenyl diphosphate (isopentenyl-PP) is the primary precusor substance for the biosynthesis of sterols and terpenoids (Hertkorn et al., 2006; Repeta, 2015), which can be produced via the mevalonate or non-mevalonate pathways (Kuzuyama et al., 2002; Liu et al., 2013). Carotenoid photodegradation products constitute a significant fraction of marine CRAM (Arakawa et al., 2017). Microbes can produce carotenoids using the key intermediate, isopentenyl-PP, which indicates that isopentenyl-PP is an important precursor for CRAM generation in the ocean (Hertkorn et al., 2006; Arakawa et al., 2017; Jiao et al., 2018). ultra-high-resolution Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) has advanced the analytical capacity to characterise complex DOM molecules, not only in laboratory bacterial cultures (Koch et al., 2014; Zheng et al., 2021), but also in seawater (Gonsior et al., 2011; Hertkorn et al., 2013). Thousands of DOM molecules can be characterized with their detailed chemical compositions (C, H, O, N, and S) using FTICR MS (Kramer et al., 2004). Some proxies have been developed to predict the characterisation of DOM, such as the double bond equivalent (DBE), H/C ratio, and degradation index ($I_{DEG}$) (Flerus et al., 2012; Hertkorn et al., 2006). FTICR MS has been used to characterise the changes in DOM molecules in seawater, and to understand the correlation between DOM composition and microbial community succession (Chen et al., 2020, 2021; Zhou et al., 2021). Correspondences between complex microbial populations and DOM molecules have been explored, showing that bacterial communities can utilise the labile DOM molecules from the aquatic environments and convert them into relative refractory DOM molecules (Osterholz et al., 2016; Zhao et al., 2019). However, current studies exploring the associations between microbial communities and DOM compositions mainly rely on 16S rRNA gene amplicon sequencing, while the ability of the functional genes of microorganisms for this remains untapped.

As a culture-independent technology, shotgun metagenomic sequencing has been widely utilised to explore natural and artificial ecosystem diversity and metabolic potential (Cao et al., 2020; Grossart et al., 2020; Sun et al., 2021; Takami et al., 2012). Microbial metabolic pathways can be characterised using metagenomic data and compared to the variation amongst different samples (Tringe et al., 2005). Changes in metabolic pathways can be detected when environmental conditions change (Gianoulis et al., 2009). Such "functional" changes are caused by the variation of the microbial community, not by alteration at the transcriptomic level. Microbial KEGG pathways are more sensitive than community structures in identifying the relationship between microbes and environmental factors (Dinsdale et al., 2008). For example, the Spearman correlation analysis between the molecular characteristics of DOM and metabolic pathways (based on metagenomes) revealed that anaerobic carbon fixation, fermentation, and methanogenesis significantly affected DOM composition in the paddy soil (Li et al., 2018). These studies suggest that metabolic processes predicted from metagenomes can provide insight into the functional transformations resulting from community variations, which facilitates further exploration of the correspondence between metagenomic functions and environmental variables.

In this study, to better understand the transformation of diatom-derived DOM at the molecular and genetic level, a 220-day incubation experiment was conducted in the Aquatron Tower Tank, a macrocosm facility of 10.46 m in height, 3.66 m in diameter, and 117.1 $m^3$ in water volume. The Aquatron Tower Tank is a large-scale ecological facility that can minimise the bottle effect and offer sufficient samples for metagenomic studies (Lian et al., 2021; Wang et al., 2020). The microbial metagenomes and DOM compositions were characterized at different incubation times. Associations between the microbial community, microbial metabolism, and DOM molecules were investigated to address the following questions: (1) How do DOM composition and characteristics change during the incubation process? (2) What are the associations between microbial activity and DOM transformation during incubation? (3) What are the potential pathways used by microbes to convert the labile DOM into the CRAMs?

## 2. Method

### 2.1. Preparation of DOM from the diatom culture

The diatom *Chaetoceros gracilis* was obtained from the Provasoli-Guillard National center for Marine Algae and Microbiota and propagated at approximately 20 °C under a light intensity of 50 μE $m^{-2} s^{-1}$ in F/2 medium. When the diatoms were cultured till the plateau phase in a 200 L spherical glass bottle (number = 6), the suspended cells were sonicated for 15 min at a resonance of 50 Hz. The filtered diatom lysate was then pumped into the Aquatron Tower Tank.

### 2.2. Incubation experiments in the aquatron tower tank

The DOM incubation experiment was performed in an Aquatron Tower Tank (Dalhousie University, Halifax, Canada) (Fig. 1). Coastal seawater (Halifax, Canada) was filtered through a filter membrane with a pore size of 5 μm and pumped into the tank. The seawater (> 108 $m^3$) was stabilised for 94 days to weaken the influence of labile DOMs in coastal seawater and obtain a stable incubation environment (additional materials, Part 1). Approximately 1000 L of diatom lysate (TOC concentration, 925 μMC) was added to the seawater (TOC concentration, 95 μMC) in the tank for the incubation experiment. The mixed seawater was allowed to develop for 220 days at a temperature of 16–22 °C, salinity 31.54–31.56 PSU, and pH 7.95–8.00 in the Aquatron Tower Tank.

### 2.3. Sample collection

Samples were collected on days 0, 1, 18, 120, 140, 180, and 220 for the analysis of diversity of the microbial community, functional genes, and DOM molecules. A total of 50 L of seawater was collected from the

tank's surface at each sampling time, and filtered using 0.22 μm polycarbonate filter membranes (Whatman), and stored at −80 °C for DNA extraction. To analyse the chemical composition of DOM, 2 L of seawater was filtered through glass fibre (Whatman) filters, and DOM was extracted using a previously described solid-phase method (Dittmar et al., 2008). The adsorbed DOM was eluted with 5 mL of methanol (HPLC grade) and stored at −20 °C in glass bottles before FTICR MS analysis.

### 2.4. FTICR MS analysis

FTICR MS analysis was performed using a 9.4 T Apex Ultra FTICR MS instrument (Bruker) equipped with negative ion electrospray ionisation (ESI) housed at the State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Beijing, China. The DOM samples were diluted to approximately 50 mg/L and injected into the ESI source at a rate of 250 μL/h (Li et al., 2019). The mass spectrometer was initially calibrated using sodium formate and then recalibrated with a known mass series in the SRFA (Suwannee River Fulvic Acid), which contains a relatively high abundance of CHO molecules, providing a mass accuracy of 0.2 ppm or higher throughout the mass range of interest. Mass peaks with an S/N greater than six, ranging from 200 to 800 *m/z* unit; and their peak intensities were exported to a spreadsheet. Data analysis was performed using in-house software (He et al., 2020). A molecular formula calculator generated matching formulas according to the elemental combinations of $^{12}C_{1-60}$, $^{1}H_{1-120}$, $^{14}N_{0-3}$, $^{16}O_{0-30}$, and $^{32}S_{0-1}$. The mass accuracy window was set at 1.0 ppm in the formula assignment section. All elemental formulas should meet the basic chemical criteria: (Kujawinski et al., 2006) (1) the number of H atoms should be at least 1/3 that of C atoms and cannot be greater than that of $2C + N + 2$; (2) the sum number of N and H atoms should be even; and (3) the H/C and O/C values should be restricted to less than 3 and 1.5, respectively.

Each sample was normalised to the sum of FTICR MS peak intensities. Bulk molecular parameters were calculated using the magnitude average (Sleighter et al., 2010; Koch and Dittmar, 2006). The modified aromaticity index ($AI_{mod}$) of each sample was calculated based on the corrected version of formula. The degradation index ($I_{DEG}$) was used to reflect the refractory of DOM (higher $I_{DEX}$ indicates older age and greater resistance) (Flerus et al., 2012). The detailed equations are shown in Additional Materials (Section 2.1). Based on the molecular lability boundary (MLB, H/C = 1.5), the DOM (H/C < 1.5) exhibited a less labile and more resistant character (MLBR), which was applied to reflect the relative variation of RDOM during the incubation period (D'Andrilli et al., 2015). Based on the elemental ratio and DBE, CRAM molecules were identified through $0.3 < DBE/C < 0.68$ and $0.2 < DBE/H < 0.95$ and $0.77 < DBE/O < 1.75$ (Hertkorn et al., 2006).

### 2.5. Genomic DNA extraction and sequencing

Genomic DNA was extracted from each sample using PowerSoil Kit (MoBio®), according to the manufacturer's protocol. The concentration and quality of extracted DNA (A260/A280) were measured using a

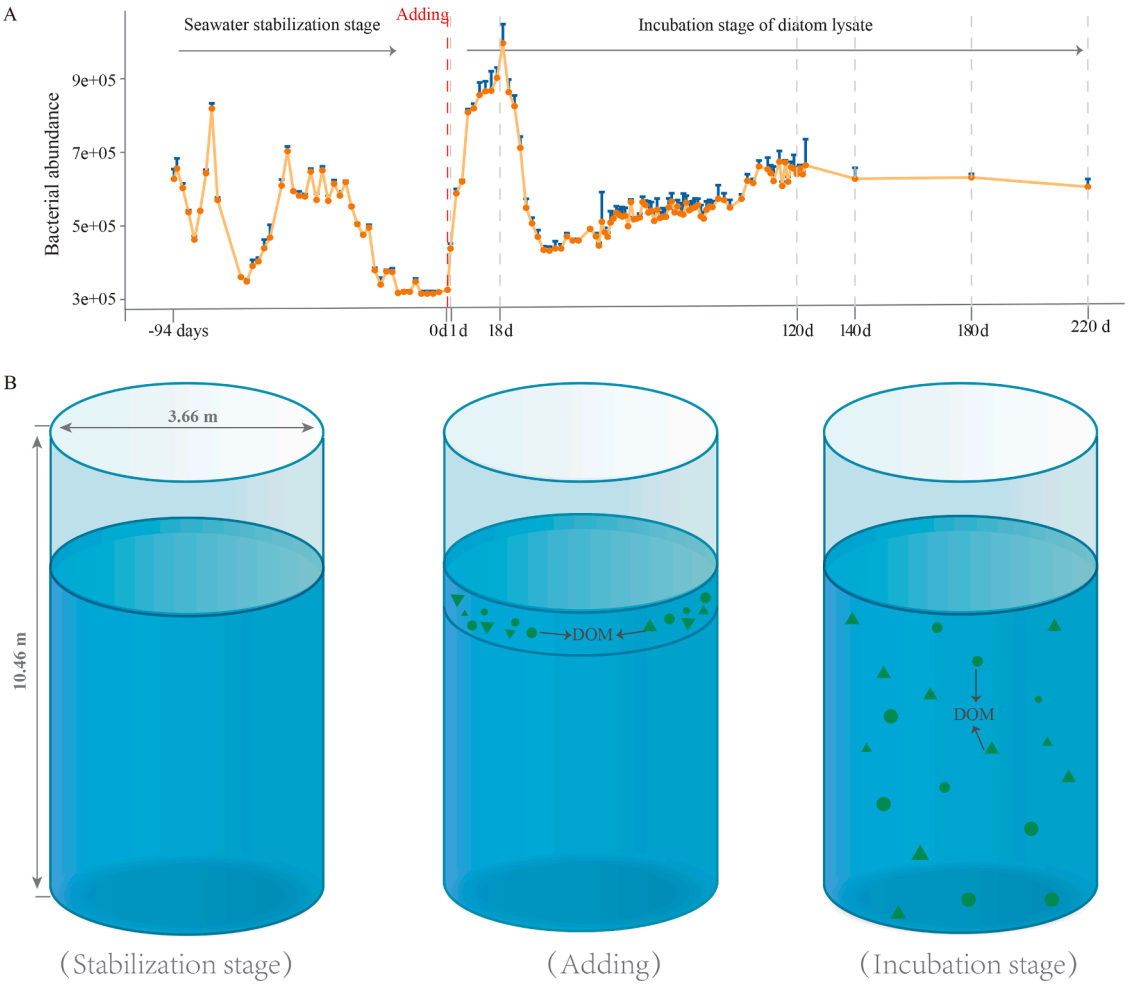


**Fig. 1.** The diagram of the incubation experiment at the Aquatron Tower Tank (Dalhousie University, Halifax, Canada). Bacterial abundance (A) and diagram of incubational experiment (B) during the incubation time. The seawater (> 108 $m^3$) was stabilized for 94 days, then added the diatom lysate on day 0, and the mixing seawater was incubated for 220 days. When diatom lysate is added, it quickly diffused in the incubational seawater. The error bar is shown with blue line.

NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). A total of 500 ng of DNA from each sample was used to construct paired-end libraries which were sequenced on an Illumina HiSeq 2500 instrument in BGI (Shenzhen, China). For each sample, an average of 34,129,578 reads, with an average length of 150 bp, were obtained. To obtain high-quality sequences (clean data), the Trimmomatic 0.36 (Bolger et al., 2014) with default settings was used to remove low-quality reads (LEADING:10, TRAILING:10, SLID-INGWINDOW:4:20, MINLEN:70).

### 2.6. Prokaryotic taxonomic, sequence assembly, and functional assignment

Kraken2 v2.0.9-beta (Wood et al., 2019), a taxonomic system, was used for microbial classification of metagenomes at the kingdom, phylum, class, order, family, genus, and species levels. The Megahit v1.2.9 (Li et al., 2016) was applied to achieve an iterative de novo assembly using the default Kmer size. The quality of genome assemblies was assessed using QUAST v5.0.2 (Gurevich et al., 2013). A total of 1545,046 contigs from all samples (N50 = 2482) were obtained to analyse the microbial function. Open reading frames (ORFs) were predicted by Prokka v1.14.6 (Seemann, 2014). Non-redundant genes were constructed (identity 95%, coverage 90%) using CD-HIT v4.6.6 (set as -c 0.95, -aS 0.9) (Fu et al., 2012), resulting in 1015,431 different genes. Salmon v0.13.1, was used to quantify genes in each sample (Patro et al., 2017), and read counts were normalised to TPM. Approximately 73% of the predicted ORFs were annotated to the EggNOG and Kyoto encyclopaedia for Genes and Genomes (KEGG) databases (Kanehisa and Goto, 2000) using diamond v0.9.14 (Buchfink et al., 2015).

### 2.7. Statistics and network analysis

All the calculations and plots were performed in an R environment (version 3.6.1). Redundancy analyses (RDA) were applied to explore the correlation amongst DOM molecules, microbial community, KEGG modules, and the calculated DOM parameters, which was conducted by the 'rda' function of the Vegan package (Oksanen et al., 2007). A Monte Carlo permutation test (permu = 999) was performed to detect the significance of the calculated DOM parameters (Lin, 2005).

A network was constructed to show the relationship between the DOM and microbial metabolism. The dominant 1617 DOM molecules (average intensity of each molecule > 10 000) and 112 KEGG modules (average relative abundance of each KEGG module > 1‰) were selected to construct the network. Spearman's rank correlation coefficient between the relative abundance of DOM molecules and the relative abundance of KEGG modules was calculated using the WGCNA package (Pei et al., 2017). DOM molecules, and KEGG modules with correlations

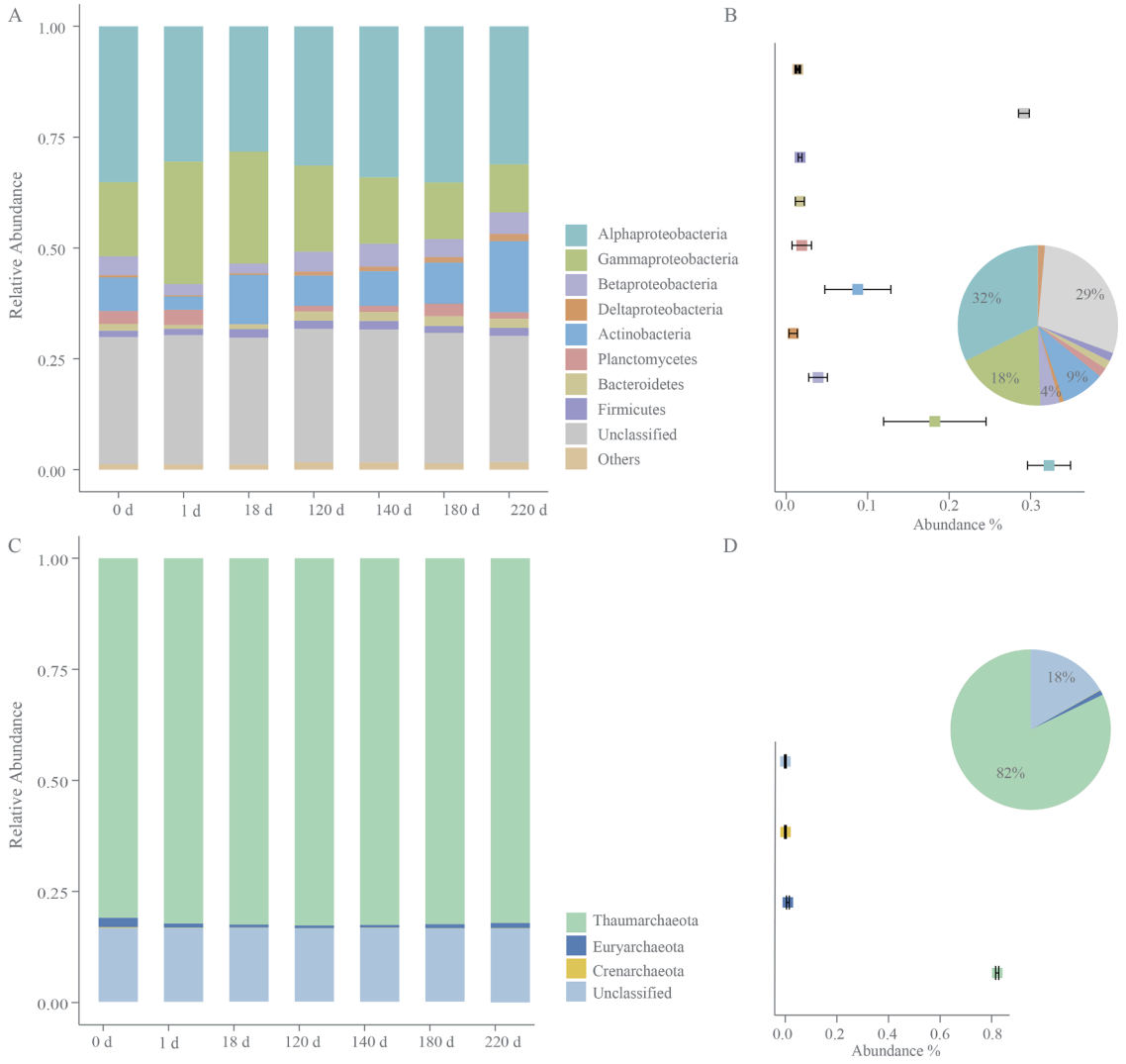


**Fig. 2.** Taxonomic structure of bacteria/archaea at different incubation times. A, the bar chart indicates the relative abundances of the bacterial phyla in each sample. B, the pie chart and error bar indicate the average relative abundances of the bacterial phyla in all samples. C, the bar chart indicates the relative abundances of the archaeal phyla in each sample. D, the pie chart and error bar indicate the average relative abundance of the archaeal phyla in all samples.

($|r| > 0.9$, $p < 0.05$) were set as nodes in the network. The network was analysed using the igraph package (Ju et al., 2016).

## 3. Results

### 3.1. Transformation of microbial community

Bacteria and archaea accounted for 78.5%–94.2% and 5.8%–21.5% of the total prokaryotic community, respectively. The relative abundance of archaea gradually increased during the incubation period (Fig. S6). Changes in the community composition during the incubation period were evident (Fig. 2). Proteobacteria, which constituted 56% of the total bacterial community, was the dominant phylum. Other abundant bacterial taxa included Actinobacteria (9%), Planctomycetes (2%), Bacteroidetes (2%), and Firmicutes (2%) (Fig. 2A). On day 0, the bacterial community was dominated by Alpha proteobacteria, Gamma proteobacteria, Beta proteobacteria, Actinobacteria, and Planctomycetes (Fig. 2A). The relative abundance of Gamma proteobacteria gradually increased on days 1 and 18, and then decreased gradually towards the end of incubation (Fig. 2A). In contrast to Gamma proteobacteria, Actinobacteria appeared to follow the opposite trend. In general, Alpha proteobacteria remained the most dominant bacterial group throughout the entire incubation period despite some variations. amongst the archaeal communities, the predominant phylum was Thaumarchaeota, which accounted for 80–82.5% of the archaeal community at different incubation times (Fig. 2B). At the genus level, two archaeal genera, *Candidatus Nitrosomarinus* and *Nitrosopumilus* (Thaumarchaeota) accounted for 31.6–32.6% of the archaeal community. Interestingly, the relative abundance of *Candidatus Nitrosomarinus* gradually increased as the incubation progressed, but *Nitrosopumilus* followed the opposite trend (Fig. S7).

### 3.2. Conversions of DOM molecules

The composition of DOM molecules was characterised based on the relative abundance of the 6837 formulas identified by FTICR MS. After the diatom lysate was added, the magnitude-weighted $O/C_W$, $DBE_W$, and $DBE/C_W$ showed slight decrease on days 1 and 18, and $H/C_W$, $DBE/O_W$, and $C\#_W$ increased slightly (Table 1). The degree of unsaturation of DOM molecules decreased on days 1 and 18, compared to days 120, 140, 180, and 220. The DOM molecular formulas ($C_{16}H_{32}O_2$ and $C_{18}H_{36}O_2$) were most abundant on days 1 and 18 (Fig. S8), which represent potential fatty acids. CRAMs were the most abundant, and constituted 67–76% of the total DOM during incubation. The MLBR and $I_{DEG}$ were used to predict the DOM resistance during the incubation times, and it was discovered that the DOM molecules were more refractory during days 120 to 220 (Table 1). Interestingly, these calculated DOM parameters on day 0 were mostly consistent with those in the later incubation period, suggesting a quick transformation upon the addition of labile DOMs.

The RDA and PCA results showed that the DOM molecules were divided into three groups during the incubation process: group 1 (day 1), group 2 (day 18), and group 3 (days of 0, 120, 140, 180, and 220) (Fig. 3A1, Fig. S9A). Significant correlations ($p < 0.05$) were observed between DOM molecules and $H/C_W$ ($r^2 = 0.99$), $O/C_W$ ($r^2 = 0.99$), $DBE_W$ ($r^2 = 0.83$), MLBR ($r^2 = 0.99$), and $I_{DEG}$ ($r^2 = 0.82$), indicating that DOM molecules were more recalcitrant in group3 than in other groups (Fig. 3A1).

The DOM molecules in different groups largely reflected the differences in molecular composition during the incubation periods. According to the eigenvalue of each DOM molecule, DOM molecules in different groups were obtained and used to understand the transformation of the chemical and molecular compositions during incubation times. Analysis of the chemical element composition showed that the DOM molecules in group 1 had a high H and N element contents, which was consistent with the high $H/C_W$ and CHON content on day 1. In group 2, DOMs had a high C content, which explains the high CHO content and $H/C_w$ value on day 18. In group 3, DOM molecules showed a low H element character and high S content, which was consistent with the high $O/C_w$, $DBE_w$, and CHOS content (Fig. 3A2 and Table 1). The van Krevelen diagram showed that CHON-containing molecules were the most crucial for the DOM composition in group 1 (Fig. 3A3), and the DOM formulas ($C_{17}H_{22}O_7N_2$ and $C_{18}H_{22}O_8N_2$) were highly abundant on day 1 (Fig. S8). The dominant molecules in group 2 were CHO-containing DOM molecules, and many molecules were potential lipid substances (Fig. 3A3). For example, formulas ($C_{16}H_{32}O_2$ and $C_{18}H_{36}O_2$) were abundant on day 18 (Fig. S8). The CHO-containing molecules in group 3 had a high O/C ratio and low H/C ratio compared with the other groups, which were understood to be potential condensed aromatics. It is noteworthy that a mass of S-containing DOM with high H/C and O/C ratios accumulated in group 3 (Fig. 3A3), implying that special S-containing organic matter accumulated with the turnover of diatom lysate.

### 3.3. Association between calculated DOM parameters and microbial community and KEGG modules

The microbial community and function shifted during the incubation period and were more similar in the later phase (days 120 to 220) (Fig. S9B, C). $O/C_W$, $DBE_W$, MLBR, and $I_{DEG}$ showed a positive relationship with the bacterial community on days 0, 120, 140, 180, and 220, and $H/C_W$ had a positive relationship with the bacterial community on days 1 and 18 (Fig. 3B1). These calculated DOM parameters could explain 66% of the total variation in the bacterial community, indicating that the microbial community was closely related to the change in DOM molecules.

KEGG modules are manually defined functional units of gene sets, and the shift of KEGG modules can be used to study changes in microbial functions. RDA analysis between the calculated DOM parameters and KEGG modules showed that $O/C_W$, $DBE_W$, MLBR, $I_{DEG}$, and $H/C_W$ were closely related to the KEGG modules (Fig. 3B2). The first two axes

**Table 1**
The chemical parameters of DOM molecules based on FTICR.

| Indexs | 0 d | 1 d | 18 d | 120 d | 140 d | 180 d | 220 d |
|---|---|---|---|---|---|---|---|
| $O/C_W$ | 0.50 | 0.44 | 0.35 | 0.50 | 0.48 | 0.48 | 0.48 |
| $H/C_W$ | 1.23 | 1.27 | 1.34 | 1.23 | 1.24 | 1.25 | 1.24 |
| $DBE_W$ | 8.60 | 8.50 | 8.15 | 8.60 | 8.52 | 8.37 | 8.50 |
| $DBE/C_W$ | 0.45 | 0.43 | 0.38 | 0.45 | 0.44 | 0.44 | 0.44 |
| $DBE/O_W$ | 0.91 | 1.02 | 1.13 | 0.91 | 0.95 | 0.94 | 0.95 |
| $C\#_W$ | 19.38 | 19.89 | 21.58 | 19.38 | 19.42 | 19.31 | 19.62 |
| MLBR | 0.93 | 0.91 | 0.82 | 0.93 | 0.93 | 0.92 | 0.93 |
| $I_{DEG}$ | 0.75 | 0.71 | 0.71 | 0.75 | 0.75 | 0.74 | 0.76 |
| MW | 415.40 | 397.31 | 411.25 | 409.25 | 406.82 | 411.19 | 411.46 |
| CRAM | 0.67 | 0.76 | 0.71 | 0.67 | 0.69 | 0.69 | 0.69 |
| CHO | 0.79 | 0.67 | 0.91 | 0.79 | 0.82 | 0.81 | 0.82 |
| CHON | 0.15 | 0.31 | 0.07 | 0.15 | 0.14 | 0.15 | 0.12 |
| CHOS | 0.05 | 0.02 | 0.02 | 0.05 | 0.05 | 0.04 | 0.06 |

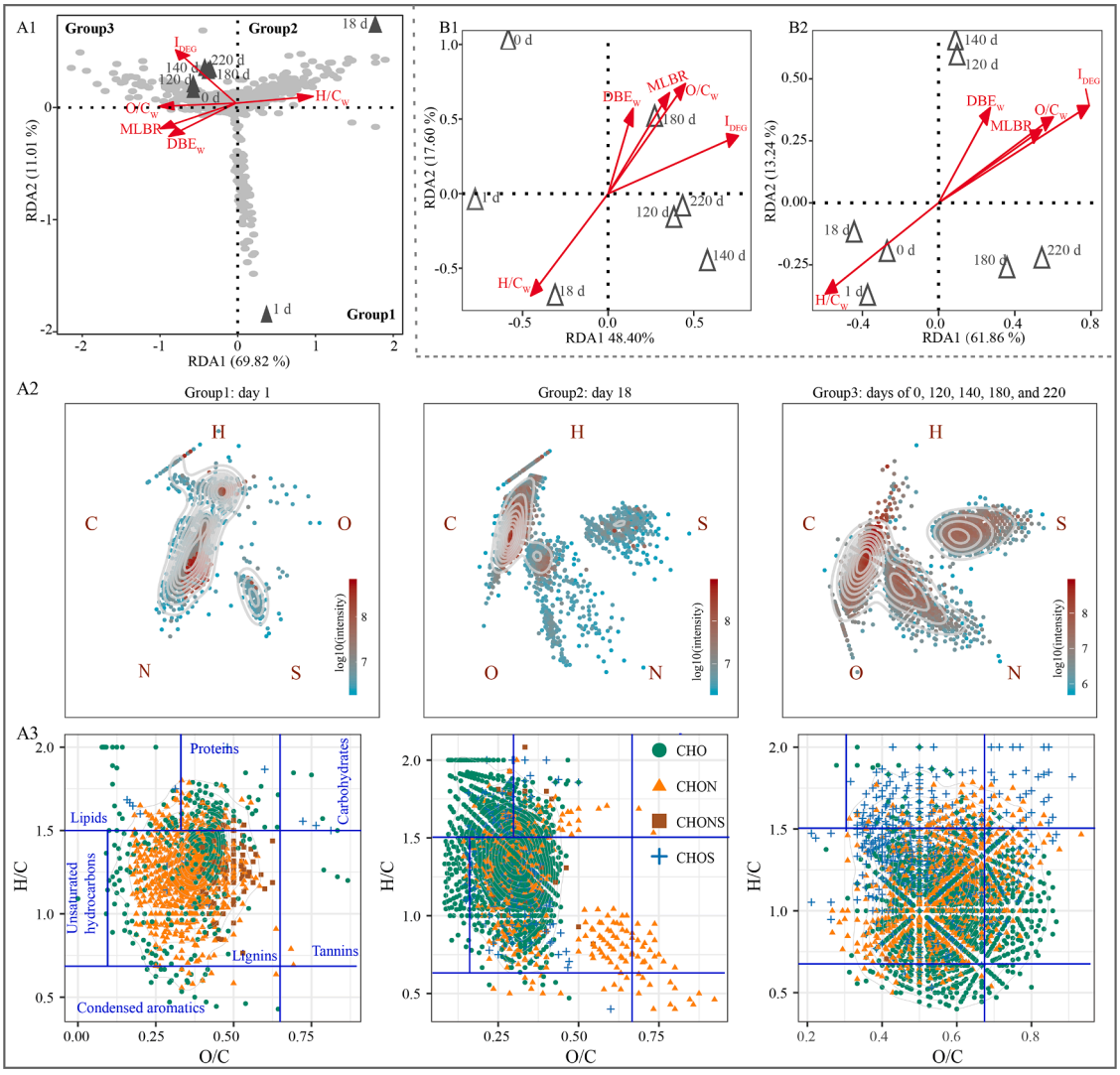


**Fig. 3.** DOM molecule composition during incubation time and the associations between microbes and DOM calculated parameters. A1, the relationship between DOM molecules and DOM calculated parameters during different times. The distance between black triangles represents the similarity of DOM molecules between different times. Grey points stand for DOM molecules. The DOM molecules can be divided into three groups (Group1, Group2, and Group3) with incubation time. A2, kernel density plots of DOMs based on chemical element composition in different groups. A3, the van Krevelen diagram of DOM molecules in different groups. The blue lines divide the DOMs into different organic categories. B1, the connection between the bacterial/archaeal community and DOM calculated parameters in the different incubation time points. The distance between black-framed triangles represents the similarity of community structure between times. B2, The connection between KEGG modules and DOM calculated parameters. The distance between black-framed triangles represents the similarity of metabolic pathways in the different incubation time points. The red arrows are DOM calculated parameters.

explained 75.1% of the variation in microbial function. Recalcitrant measurements such as MLBR and $I_{DEG}$ positively correlated with microbial community function on days 120, 140, 180, and 220 (Fig. 3B2), indicating a potential connection between microbial function transformation and RDOM generation.

### 3.4. Associations between KEGG modules and DOM formulas

Network analysis was used to explore and visualize the complex association between KEGG modules and DOM molecules (CRAMs and other DOMs). Direct connections were created as edges between the selected DOM molecular formulas and network KEGG modules (coloured nodes) (Fig. 4A). In the network, 472 DOM molecules and 76 KEGG modules were selected based on their high Spearman's correlation coefficients ($|r|>0.9$, $p < 0.05$). These DOM molecules are divided into two groups: CRAMs (340 molecules), and other DOM molecules (132 molecules). The label size of the KEGG module was proportional to the rank of the KEGG module nodes (the number of connections). The complex network had 548 nodes and 4852 edges, and the average degree and diameter were 17.7 and 15, respectively.

In the network, each KEGG module connects many CRAMs and other DOM molecules. To explore the contribution of KEGG module to the CRAM generation, we calculated the proportion of CRAM by dividing the number of CRAM molecules that were positively/negatively correlated with KEGGs by the number of KEGG module connections. KEGG modules (the proportion that positively correlated with CRAM is more than 60%) was chosen to explore the potential CRAM generation pathways, indicating that these KEGG modules are likely to facilitate the process of CRAM generation. The results showed that 11 KEGG modules strongly positively correlated with the CRAM accumulation (Fig. 4B). These KEGG modules included: 1) beta-oxidation (M00087 and M00086) – Beta-oxidation is the catabolic process by which fatty acid molecules are broken down in the cytosol of prokaryotes to generate acetyl-CoA, which is used in biosynthesis and energy metabolism; 2) Phenylacetate degradation (M00878) and Trans-cinnamate degradation (M00545) – Phenylacetate and trans-cinnamate are the major

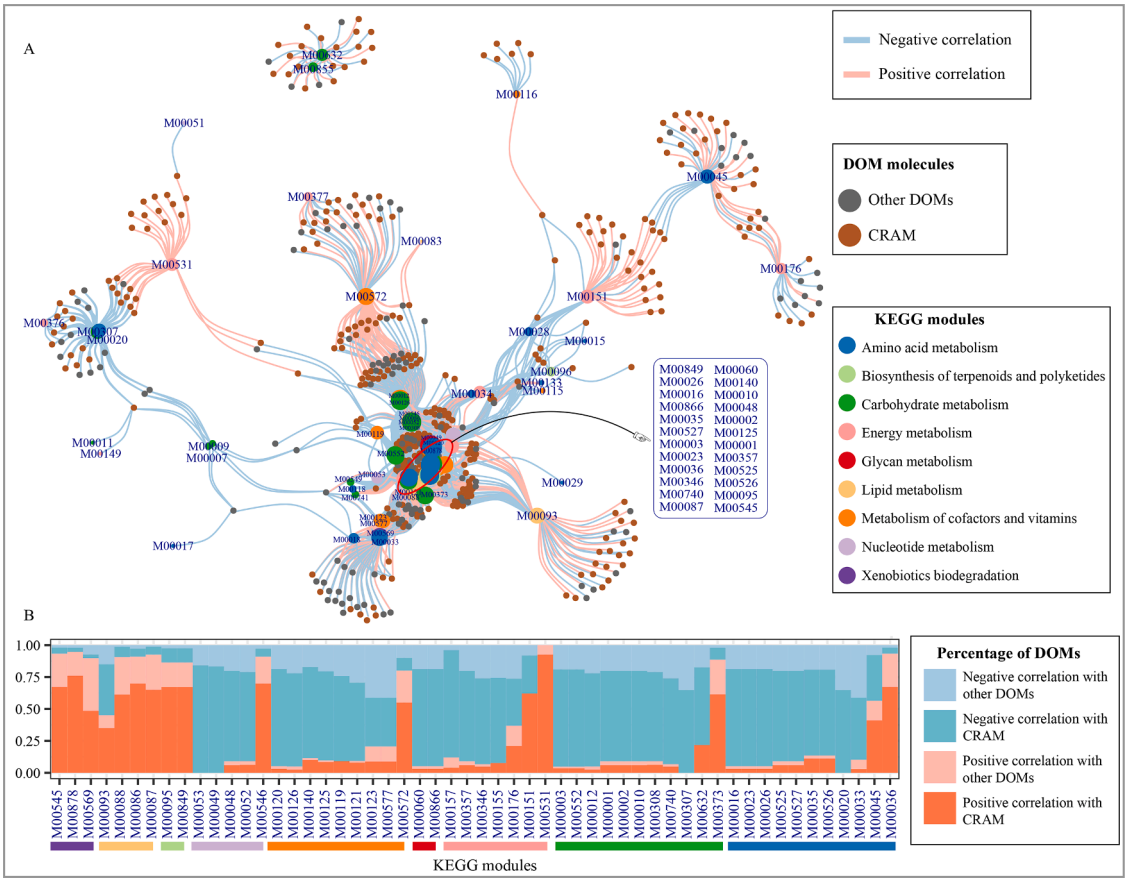


**Fig. 4.** Network layout showing the association between KEGG modules and DOM molecules. A The connection between KEGG modules and DOM molecules (CRAMs and other DOMs). Direct connections were made between DOM molecules and KEGG modules when the Spearmen's rank correlation coefficient is strong $r \geq 0.9$ (Positive network) and $r \leq -0.9$ (Negative network), $p < 0.05$. The brown and black points stand for CRAM and other DOMs, respectively. B The percentage of KEGG modules that contribute to the accumulation of CRAM and other DOM molecules. The connections of each KEGG module are > 20. Deep red is the percentage of each KEGG module that is positively correlated to the CRAM. Light red is the percentage of each KEGG module that is negatively correlated to the CRAM. Deep blue represents the percentage of each KEGG module that is positively correlated with the other DOM. Light blue is the percentage of each KEGG module that is negatively correlated to the other DOM.

intermediates in the bacterial degradation of many aromatic compounds, such as phenylalanine. Bacteria can oxidise phenylacetate and trans-cinnamate to produce acetyl-CoA via M00878 and M00545 modules finally; 3) Leucine degradation (M00036) – Leucine is one of the three main branched-chain amino acids that can be completely converted into three acetyl-CoA through the leucine degradation pathway; 4) Purine degradation (M00546) – This pathway describes the degradation of xanthine to ammonia and contributes to amino acid metabolism; 5) Assimilatory nitrate reduction (M00531) – Bacteria utilise nitrate and nitrite from the environment by reducing these nitrogen oxides to ammonium ions, which are readily metabolised into organic molecules, such as amino acids; 6) Mevalonate pathway of bacterial (M00095) and archaea (M00849) – The mevalonate pathway is an essential metabolic pathway presented in both archaea and bacteria. The mevalonate pathway begins with acetyl-CoA (the sole carbon feedstock) and ends with the production of isopentenyl-PP, which is used to biosynthesize various biomolecules, such as steroids, terpenoids, and carotenoids; 7) Ethyl malonyl pathway (M00373) – Bacteria utilise organic substances (such as fatty acids and so on) that are metabolised via acetyl-CoA. The ethyl malonyl pathway converts acetyl-CoA, a 2-carbon compound, to a 4-carbon compound, which is used in the citrate cycle to provide energy to the bacteria; 8) Cytochrome bc1 complex respiratory unit (M00151) – This module is a crucial component of bacterial energy metabolism. The above KEGG modules are involved in the production of acetyl-CoA, and the mevalonate pathway is involved in the utilisation of acetyl-CoA for production of the terpenoid backbone. For example, the formula $C_{19}H_{24}O_9$ ($m/z$ = 396.14) was the most abundant CRAM (Fig. S10), and positively correlated with leucine, phenylacetate, and trans-cinnamate degradation, beta-oxidation, and the mevalonate pathway (Fig. 4A). Leucine, phenylacetate, trans-cinnamate degradation can produce acetyl-CoA, which can be utilised to generate the terpenoid backbone via the mevalonate pathway.

In addition, some important KEGG modules related to gluconeogenesis (M00003), glycolysis (M00001 and M00002), pyruvate oxidation (M00307), and the citrate cycle (M00010) negatively correlated with the CRAMs, and these pathways were mostly involved in microbial energy metabolism.

### 3.5. *Potential pathways for microbial production of CRAM*

Microbial transport and secretion systems serve as conduits for material exchange between intracellular and extracellular environments (Dinsdale et al., 2008). In our study, the dominant membrane transport was the ABC transport family, which accounted for 2.1% of all genes, followed by the bacterial secretion system (1.3%), and PTS (0.13%). amongst ABC transporters, N/P transporters, especially those for branched-chain amino acids, were the most abundant, accounting for 33% of the ABC transport system. The fructose transporter was the most dominant in the PTS, and Sec-SRP was the most abundant in the secretion system (Fig. S12).

Eleven KEGG modules, that positively correlated with CRAM (Fig. 4) were concatenated to construct a putative schema for the biogenic origin

of CRAM (Fig. 5). Three potential pathways were identified, including the leucine, fatty acid, and phenylacetate pathways. Leucine from intracellular synthesis or extracellular transport was completely degraded to two 3-Hydroxy-3-methylglutaryl-CoA via the leucine degradation pathway, and this process fixed two inorganic $HCO_3^-$. The 3-Hydroxy-3-methylglutaryl-CoA enters the mevalonate pathway directly to produce isopentenyl-PP, which is the exclusive precursor for the production of steroid, terpenoid, or carotenoid (CRAM analogs) via specific metabolic pathways, such as monoterpenoid biosynthesis, diterpenoid biosynthesis, carotenoid biosynthesis, indole diterpene alkaloid biosynthesis, ubiquinone and other terpenoid-quinone biosynthesis, and steroid biosynthesis (Kuzuyama, 2017; Kuzuyama et al., 2002; Okamura et al., 2010; Sever et al., 2003). In addition, leucine can also be degraded completely to three acetyl-CoA molecules, which enter the mevalonate pathway to produce isopentenyl-PP. Fatty acids are progressively oxidized to produce acetyl-CoA by beta-oxidation which is the most abundant pathway. Acetyl-CoA is used by the mevalonate pathway to produce isopentenyl-PP, which enters those special metabolic pathways to generate CRAM analogues (Pang et al., 2021). Phenylacetate is degraded to acetyl-CoA and succinyl-CoA, respectively. Acetyl-CoA may contribute to the production of CRAM analogues based on the mevalonate pathway and special metabolic pathways. Succinyl-CoA enters the tricarboxylic acid (TCA) cycle to provide energy for microbial activity. In addition, assimilatory nitrate reduction and purine degradation could produce ammonia for amino acid synthesis, and its degradation may contribute to acetyl-CoA production (Fig. 5 and Fig S13). In general, microbes produce the CRAM though two key steps: accumulation of original substrates, such as acetyl-CoA and 3-Hydroxy-3-methylglutaryl-CoA, and biosynthesis of CRAM analogues by the mevalonate pathway and special metabolic pathways (Fig. 5).

To further confirm CRAM generation at the molecular weight level, network analysis showed that 217 CRAM were positively correlated with the KEGG modules (Fig. 4A), and the H/C and O/C ratios ranged from 0.75 to 1.5 and 0.3 to 0.6, respectively (Fig. S11A). The molecular weights of these 217 CRAMs (*m/z*) ranged from 225 to 625 *m/z*, and most CRAMs were concentrated between 325 and 500 *m/z* (Fig. S11B). We found that the molecular weight interval was overlapped and interweaved between CRAM and compounds in the following pathways, including steroid biosynthesis, ubiquinone and other terpenoid-quinone biosynthesis, indole diterpene alkaloid biosynthesis, monoterpenoid biosynthesis, diterpenoid biosynthesis, and carotenoid biosynthesis (Fig. S11 C, D, E, F, G, and H), suggesting a transformation of CRAM at the molecular level.

### 3.6. Microbes involved in the mevalonate pathway

The mevalonate pathway is the key KEGG module that converts acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA into precursors of CRAM in this study. To identify the taxa in the mevalonate pathway, the seven most abundant microbial lineages were extracted and analysed (Fig. 6). The dominant microbes were Alpha proteobacteria, Gamma proteobacteria, Actinobacteria, and Thaumarchaeota (Fig. 6A). The dominant genera shifted with incubation time (Fig. 6B). Upon the addition of the diatom lysate, *Nitrosopumilus* and *Nitrosarchaeum* were abundant on days 1 and 18, but *Candidatus Nitrosomarinus* was dominant on days 120, 140, 180, and 220. Other bacterial lineages became dominant at different times during the incubation period.

The cooperation between multiple KEGG modules and the mevalonate pathway potentially contributes to the generation of CRAM. To study the species' connection between KEGG modules and the mevalonate pathway, KEGG modules that produced acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA had the most similar species to the mevalonate pathway at the genus level (Fig. 6C). As a control, the gluconeogenesis (M00003) module had a strong negative correlation with CRAM in our study. A total of 737 different genera were extracted from gluconeogenesis, 480 from mevalonate pathway, 586 from leucine degradation pathway, 427 from branch-chain amino acid transporter, 570 from beta-oxidation, and 553 from phenylacetate degradation pathway. Approximately 45% of the genera involved in the gluconeogenesis was shared with the mevalonate pathway (Fig. 6Ca). However, the genera of the leucine degradation pathway (Fig. 6Cb), beta-oxidation (Fig. 6Cc), and phenylacetate degradation pathway (Fig. 6Cd) shared 62%, 65%, and 64% of the mevalonate pathway, respectively. In addition, 60% of the branch-chain amino acid transporter genera were like those in the mevalonate pathway. These results illustrate that the CRAM-related KEGG modules had a high probability occurrence in the genomes of the same species.

## 4. Discussion

CRAMs are recalcitrant DOM molecules (Li et al., 2019., Koch et al., 2006). In this study, CRAM accounted for a large proportion (> 67%) of all DOM molecules during the incubation period, suggesting a significant contribution to the accumulation of refractory DOM. Therefore, the biogenic origin of CRAM can further improve our understanding of the detailed process of RDOM accumulation caused by microbial activity. Previous studies have reported that bacteria can transform distinct exogenous substrates, such as trehalose (Lian et al., 2021),

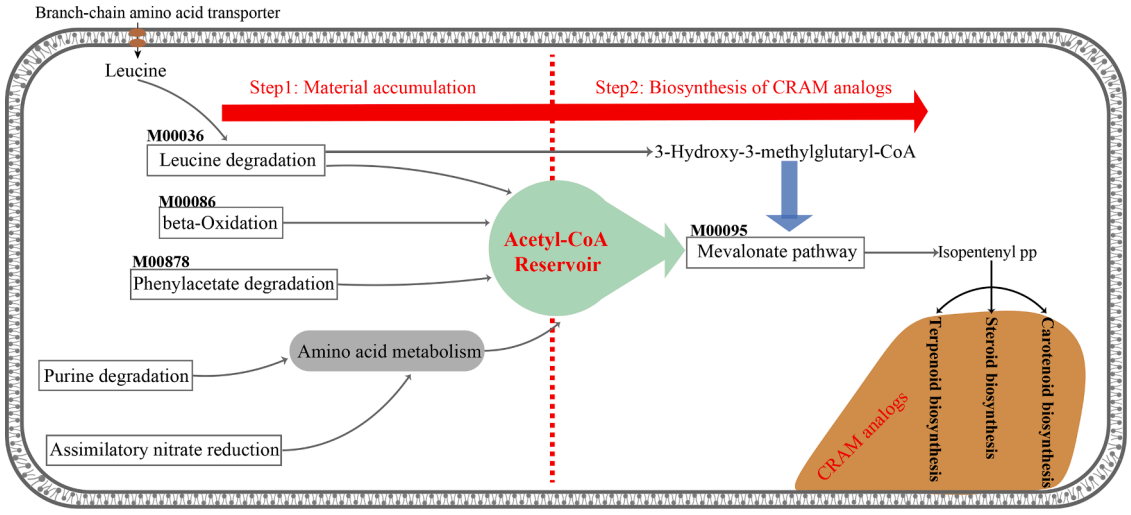


**Fig. 5.** A putative schema that illustrates the generation pathway of CRAMs by the microbe. All KEGG modules were positively correlated to the accumulation of CRAM, which marked with grey boxes. The outermost grey box represents the cell membrane. The detailed information was showed in Fig S12.

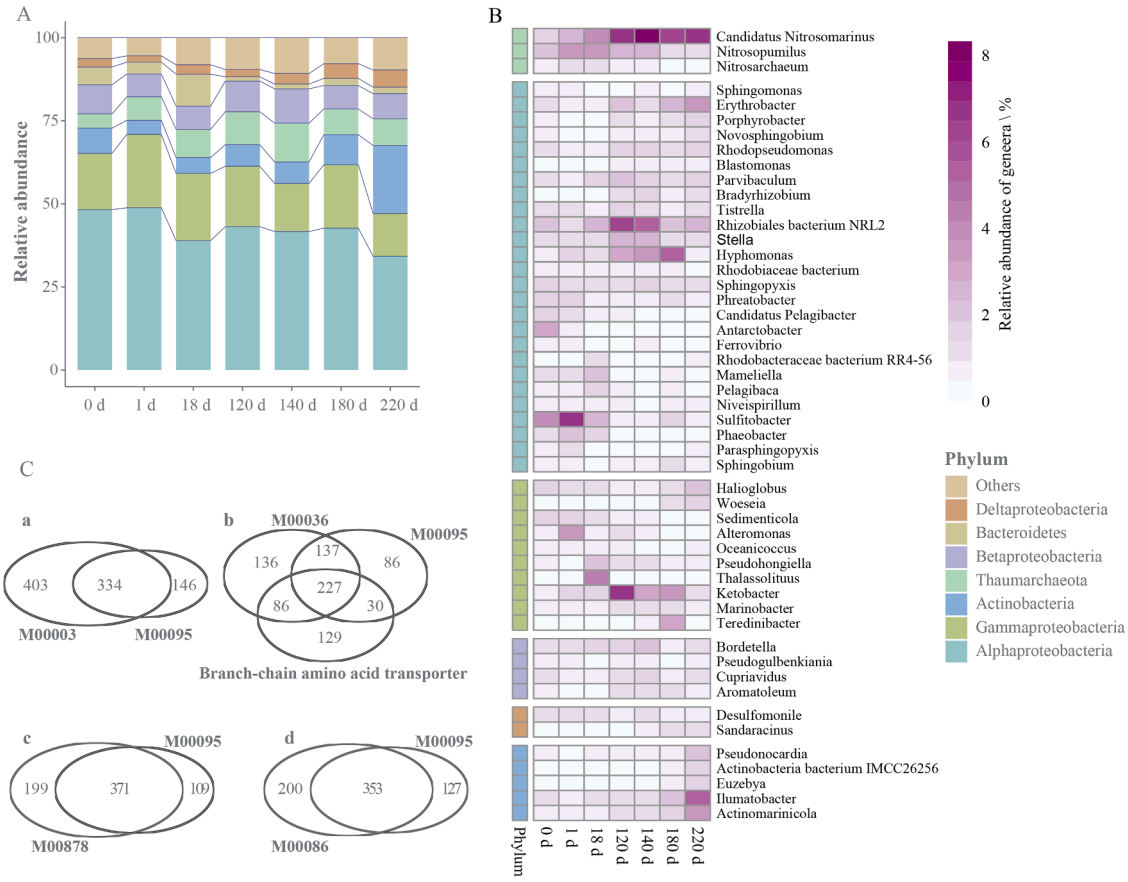


**Fig. 6.** The microbial diversity in the mevalonate pathway (M00095). The microbial community at the phylum (A) and genus (B) level. We selected the 50 most abundant genera. The different colors stand for the taxon at the phylum level. The heatmap showed the relative abundance of each dominated genera in the different incubation time points. C, Venn diagram showing the number of genera between different KEGG modules. Number of genera detected between gluconeogenesis (M00003) and mevalonate pathway (M00095) (Ca); amongst leucine degradation (M00036), mevalonate pathway (M00095), and branch-chain amino acid transporter (Cb); between phenylacetate degradation (M00878) and mevalonate pathway (M00095) (Cc); and between beta-Oxidation (M00086) and mevalonate pathway (M00095) (Cd).

cyanobacteria lysate (Zhao et al., 2019), and *Synechococcus* lysates (Zheng et al., 2021) into the analogous composition of CRAM at the community level. Our results indicated a close connection between refractory DOM generation and the microbial community and KEGG modules, not only at the community level but also at the genetic level. Recalcitrant measurements, such as MLBR and $I_{DEG}$, were found to be positively correlated with DOM molecules, bacterial and archaeal communities, and KEGG modules in the later period of the incubation experiment (from 120 to 220 days), indicating that microbial metabolism processes seem to be deeply involved in refractory DOM generation. Therefore, network analysis was applied to identify the KEGG modules that are potentially engaged in CRAM generation. Eleven KEGG modules were positively correlated with the CRAM (Fig. 4), which seems to have contributed to CRAM generation. Based on these KEGG modules, we discovered that CRAM generation requires two steps: first, production of acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA, and then transformation of acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA to the CRAM precursor, isopentenyl-PP (Fig. 5). This study provides a new horizon for carbon cycling and the origin of the ocean RDOM pool.

Acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA is an important primary material for CRAM generation, and participates in the mevalonate pathway to synthesise the steroids and terpenoids in prokaryotes (Bode et al., 2003; Siedenburg and Jendrossek, 2011). The acetyl-CoA family is prominent in facilitating carbon fixation, transportation, and CRAM generation in SAR202 genomes in the abyssal zone (Landry et al., 2017). Acetyl-CoA is produced by a variety of metabolic processes (Bertsch et al., 2015; Khomyakova et al., 2011). However, a few key acetyl-CoA-producing pathways, such as beta-oxidation, leucine, and phenylacetate and trans-cinnamate degradation, seem to contribute to the generation of CRAM in our study, which is strongly related to the conversion of the organic matter in the cell. Interestingly, glycolysis, pyruvate oxidation, and TCA, which are candidates for acetyl-CoA production, negatively correlated with the CRAM generation (Fig. 4). Pyruvate oxidation is the conversion of pyruvate to acetyl-CoA by the enzyme complex pyruvate dehydrogenase complex, which is the step that connects glycolysis and the TCA cycle (Amador-Noguez et al., 2010; Fattal-Valevski, 2011; Su et al., 2018), indicating that negative KEGG modules are closely related to microbial energy metabolism activities. The difference in acetyl-CoA producing pathways indicated that KEGG modules related to organic transformation pathways favour the generation of CRAM compared to those related to energy metabolism pathways. In addition, abundant KEGG modules were chosen for coupling analysis with CRAM in data processing, and some low abundant KEGG modules were discarded, even though they were involved in acetyl-CoA production in this study. The relative abundance of genes or KEGG modules may differ depending on the habitat (Liu et al., 2020; Tanvir et al., 2021). It is not surprising that other KEGG modules that produce acetyl-CoA may be potential candidate pathways for CRAM generation in the future.

Isopentenyl-PP production appeared to be a key step in the mevalonate pathway for CRAM formation in our study. Acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA participates in the mevalonate pathway to produce isopentenyl-PP, which is an important precursor for the biosynthesise of the CRAM analogue (steroid, terpenoid, or carotenoid)

through metabolic pathways, such as steroid, ubiquinone and other terpenoid-quinone, indole diterpene alkaloid, monoterpenoid, diterpenoid, and carotenoid biosynthesis pathways. The weight ($m/z$) of these metabolic pathways overlapped with that of CRAM which positively correlated with KEGG modules (Fig. S11), indicating that CRAM may be generated in these metabolic pathways. Alpha proteobacteria, Gamma proteobacteria, and Thaumarchaeota were the dominant microbes involved in the mevalonate pathway. Previous studies have also found that these three microbial groups are abundant and essential for the transformation of labile DOM into CRAM based on DOM incubation experiments (Chen et al., 2020; Lian et al., 2021). Our study provides evidence that the mevalonate pathway is active in response to CRAM generation at the genetic level. Veen's analysis found that CRAM-related KEGG modules had a high probability of occurrence in the genomes of the same species, indicating that the cooperation amongst different KEGG modules was essential for CRAM generation. The non-mevalonate pathway is another KEGG module that produces the precursor of CRAM, isopentenyl-PP (Eisenreich et al., 2004). However, there was no evidence in this study regarding non-mevalonate pathway's participation in CRAM generation. The non-mevalonate pathway utilises pyruvate to generate isopentenyl-PP, unlike the mevalonate pathway. Our study showed that pyruvate oxidation negatively correlated with CRAM generation (Fig. 4B), which may play a vital role in providing energy to bacteria and archaea (Fattal-Valevski, 2011). The primary role of pyruvate was to provide energy for microbial activity rather than CRAM generation. Therefore, there is little connection between the non-mevalonate pathway and CRAM production.

The leucine path was a highly efficient converter for CRAM generation in our study. Bacteria and archaea can degrade leucine into acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA, which is used in the mevalonate pathway for the generation of CRAM analogues. A previous study reported that leucine could be used to trace the refractory bacterial material accumulating in the ocean, using the $^{14}C$ age analysis (Broek et al., 2019), which provides evidence that leucine is the main substrate in the biosynthesis of refractory DOM in the ocean via microbial activity. In this study, branched-chain amino acid transporters were the most abundant during the course of incubation. Leucine can be transferred to the cell interior by branched-chain amino acid transporters (Haney and Oxender, 1992), indicating that extracellular leucine can be utilised to produce CRAM via the intracellular mevalonate pathway. In the leucine degradation processes, 3-Methylcrotonyl-CoA carboxylase uses an ATP molecule and $HCO_3^-$to catalyse 3-Methylcrotonyl-CoA into 3-Methylglutaconyl-CoA (Höschle et al., 2005; Massey et al., 1976). The inorganic carbon $HCO_3^-$, the largest oceanic carbon reservoir (Bartley and Kah, 2004), is fixed and transformed to acetyl-CoA or 3-Hydroxy-3-methylglutaryl-CoA in the leucine degradation process, which participates in the mevalonate pathway for CRAM generation. Our results indicate that bacteria and archaea have the potential to convert both labile leucine and inorganic $HCO_3^-$ into CRAM, suggesting a potential relationship between the RDOM reservoir and the dissolve inorganic carbon (DIC) pool. Here, a novel perspective is presented in which microbes could utilise the carbon from the DIC pool to contribute to the RDOM reservoir.

Fatty acid oxidation greatly contributes to CRAM generation. Bacteria and archaea can oxidise fatty acids to produce a large amount of acetyl-CoA, which can be used in the mevalonate pathway to produce CRAM analogues. In our study, the relative abundance of the beta-oxidation pathway was the highest in the CRAM generation pathways. Lipids and unsaturated hydrocarbons which was predicted by H/C and O/C ratio, accumulated in the incubation seawater at day 18 and disappeared on days 120 to 220 (Fig. 3A3, Fig. S8). Microbial community transcriptomes in seawater incubation experiment have shown that fatty acid oxidation is abundant and significantly related to the DOM transformation (McCarren et al., 2010).

During the incubation period, DOM molecules were more labile on days 1 and 18 than on days 0, 120, 140, 180, and 220 (the low $I_{DEG}$ and MLBR on days 1 and 18), suggesting the introduction of labile DOM from the diatom lysate. The main dynamics of labile DOM transformation is microbial activity (Lechtenfeld et al., 2015; Zheng et al., 2021). A long-term incubation experiment in the Aquatron Tower Tank indicated that the diatom lysate debris was effectively transformed into relatively refractory DOM during the incubation period. During the later period of the incubation experiment (days 120–220), many CHO-containing DOM molecules were predicted to be analogues of condensed aromatics, which had a low H/C ratio and high O/C ratio. This unsaturated organic matter resisted microbial degradation (Rossel et al., 2013) and was sequestrated in the seawater for thousands of years as a carbon sink. It is worth noting that many S-containing DOM molecules accumulated on days 120 to 220, which contained a high H/C and O/C ratio (Fig. A3). Dissolved organic sulphur (DOS) shows a higher content of hydrogen (saturation) and oxygen (oxidation) and forms a large pool of refractory DOS reservoirs in the ocean (Ksionzek et al., 2016). Our study demonstrated that microbial transformation of the diatom lysate potentially contributed to the accumulation of refractory DOS molecules in the ocean.

The incubation time in the incubation experiments varied in different studies, such as the 15 days' *Synechococcus* incubation experiment (Zhao et al., 2017), 25 days' *Synechococcus* lysate incubation experiment (Zheng et al., 2020), and 90 days' incubation experiment in distinct exogenous substrates (Lian et al., 2021). This study performed a long-term incubation experiment for 220 days in the Aquatron Tower Tank, which is important for understanding the relationship between DOM transformation and microbial metabolism at an ecological level with a high temporal resolution. Before adding the diatom lysate, a 94-day stabilization stage was conducted to obtain a clean background for the incubation experiment in the Aquatron Tower Tank.

## 5. Conclusion

Our study provides evidence at the genetic level, that the cooperation of microbial metabolic pathways plays a vital role in CRAM generation. We found that bacteria/archaea can degrade labile DOM, such as leucine, fatty acids, and phenylacetate, into acetyl-CoA or 3-hydroxy-3-methylglutaryl-CoA, which participate in the mevalonate pathway to produce precursors of the CRAM analogues, isopentenyl-PP. The leucine path is an effective converter for CRAM generation, which can translate intra-or extracellular leucine into CRAM and fix inorganic carbon $HCO_3^-$, suggesting a close relationship between the RDOM reservoir and the DIC pool in the ocean. In addition, many refractory S-containing DOMs was accumulated in the latter period of the experiment, indicating that microbial turnover of diatom lysate contributes not only to the RDOM pool but also to the DOS pool in the ocean.

## Availability of data and materials

The raw sequence data was deposited in the SRA of the NCBI (https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA76750. The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

## Author contributions

NZ, J and C, F were involved in the conception and designed of the study. JH, L; R, W; FL, J; LF, L and QH, Z were involved to perform the experiment. C, H and S, Q provided the technical support in FTICR. CF, H analysed the data and wrote the manuscript. F, C and JH, L provided the guidance during the writing. JH, L; YN, L; Q, Z; QC, T; T, S; YL, X; R, Z; H, T; J, B; P, H; M, L and H, M modified the manuscript. All authors read and approved the final manuscript.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the National Key Research and Development Program of China (2018YFA0605800), the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311021006), the Key Research and Development Program of Shandong Province (2020ZLYS04), and National Key Research and Development Program of China (2019YFA0606704).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.118281.

## References

Amador-Noguez, D., Feng, X., Fan, J., Roquet, N., Rabitz, H., Rabinowitz, J.D., 2010. Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in Clostridium acetobutylicum. J. Bacteriol. 192 (17), 4452–4461.
Arakawa, N., Aluwihare, L.I., Simpson, A.J., Soong, R., Stephens, B.M., Lane-Coplen, D., 2017. Carotenoids are the likely precursor of a significant fraction of marine dissolved organic matter. Sci. Adv. 3 (9), e1602976.
Bartley, J.K., Kah, L.C., 2004. Marine carbon reservoir, Corg-Ccarb coupling, and the evolution of the Proterozoic carbon cycle. Geology 32 (2), 129–132.
Bertsch, J., Öppinger, C., Hess, V., Langer, J.D., Müller, V., 2015. Heterotrimeric NADH-oxidizing methylenetetrahydrofolate reductase from the acetogenic bacterium Acetobacterium woodii. J. Bacteriol. 197 (9), 1681–1689.
Bode, H.B., Zeggel, B., Silakowski, B., Wenzel, S.C., Reichenbach, H., Müller, R., 2003. Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2, 3 (S)-oxidosqualene cyclase from the myxobacterium Stigmatella aurantiaca. Mol. Microbiol. 47 (2), 471–481.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120.
Broek, T.A., Bour, A.L., Ianiri, H.L., Guilderson, T.P., McCarthy, M.D., 2019. Amino acid enantiomers in old and young dissolved organic matter: implications for a microbial nitrogen pump. Geochim. Cosmochim. Ac. 247, 207–219.
Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12 (1), 59–60.
Cao, S., Zhang, W., Ding, W., Wang, M., Fan, S., Yang, B., Mcminn, A., Wang, M., Xie, B. B., Qin, Q.L., Chen, X.L., He, J., Zhang, Y.Z., 2020. Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. Microbiome 8 (1), 47.
Chen, J., Li, H., Zhang, Z., He, C., Shi, Q., Jiao, N., Zhang, Y., 2020. DOC dynamics and bacterial community succession during long-term degradation of Ulva prolifera and their implications for the legacy effect of green tides on refractory DOC pool in seawater. Water Res. 185, 116268.
Chen, Q., Chen, F., Gonsior, M., Li, Y., Wang, Y., He, C., Cai, R., Xu, J., Wang, Y., Xu, D., Sun, J., Zhang, T., Shi, Q., Jiao, N., Zheng, Q., 2021. Correspondence between DOM molecules and microbial community in a subtropical coastal estuary on a spatiotemporal scale. Environ. Int. 154, 106558.
D'Andrilli, J., Cooper, W.T., Foreman, C.M., Marshall, A.G., 2015. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. Rapid Commun. Mass Sp. 29 (24), 2385–2401.
Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., 2008. Functional metagenomic profiling of nine biomes. Nature 452 (7187), 629–632.
Dittmar, T., Koch, B., Hertkorn, N., Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnol. Oceanogr.: Methods 6 (6), 230–235.
Ducklow, H.W., Steinberg, D.K., Buesseler, K.O., 2001. Upper ocean carbon export and the biological pump. Oceanography 14 (4), 50–58.
Eisenreich, W., Bacher, A., Arigoni, D., Rohdich, F., 2004. Biosynthesis of isoprenoids via the non-mevalonate pathway. Cellular Mol. Life Sci. CMLS 61 (12), 1401–1426.
Eppley, R.W., Peterson, B.J., 1979. Particulate organic matter flux and planktonic new production in the deep ocean. Nature 282 (5740), 677–680.
Fattal-Valevski, A., 2011. Thiamine (vitamin B1). J. Evid. Based Complementary Altern. Med. 16 (1), 12–20.
Flerus, R., Lechtenfeld, O.J., Koch, B.P., McCallister, S.L., Schmitt-Kopplin, P., Benner, R., Kaiser, K., Kattner, G., 2012. A molecular perspective on the ageing of marine dissolved organic matter. Biogeosciences 9 (6), 1935–1955.
Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28 (23), 3150–3152.
Gianoulis, T.A., Raes, J., Patel, P.V., Bjornson, R., Korbel, J.O., Letunic, I., Yamada, T., Paccanaro, A., Jensen, L.J., Snyder, M., 2009. Quantifying environmental adaptation of metabolic pathways in metagenomics. Proc. Natl. Acad. Sci. 106 (5), 1374–1379.
Gonsior, M., Peake, B.M., Cooper, W.T., Podgorski, D.C., D'Andrilli, J., Dittmar, T., Cooper, W.J., 2011. Characterization of dissolved organic matter across the Subtropical Convergence off the South Island, New Zealand. Mar. Chem. 123 (1–4), 99–110.
Grossart, H.P., Massana, R., McMahon, K.D., Walsh, D.A., 2020. Linking metagenomics to aquatic microbial ecology and biogeochemical cycles. Limnol. Oceanogr. 65, S2–S20.
Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29 (8), 1072–1075.
Haney, S.A., Oxender, D.L., 1992. Amino acid transport in bacteria. Int. Rev. Cytol. 137, 37–95.
Hansell, D.A., Carlson, C.A., Repeta, D.J., Schlitzer, R., 2009. Dissolved organic matter in the ocean: a controversy stimulates new insights. Oceanography 22 (4), 202–211.
He, C., Zhang, Y., Li, Y., Zhuo, X., Li, Y., Zhang, C., Shi, Q., 2020. In-house standard method for molecular characterization of dissolved organic matter by FT-ICR mass spectrometry. ACS omega 5 (20), 11730–11736.
Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., Hedges, J.I., 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochim. Cosmochim. Ac. 70 (12), 2990–3010.
Hertkorn, N., Harir, M., Koch, B.P., Michalke, B., Schmitt-Kopplin, P., 2013. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. Biogeosciences 10 (3), 1583–1624.
Höschle, B., Gnau, V., Jendrossek, D., 2005. Methylcrotonyl-CoA and geranyl-CoA carboxylases are involved in leucine/isovalerate utilization (Liu) and acyclic terpene utilization (Atu), and are encoded by liuB/liuD and atuC/atuF, in Pseudomonas aeruginosa. Microbiology+ 151 (11), 3649–3656.
Jiao, N., Cai, R., Zheng, Q., Tang, K., Liu, J., Jiao, F., Wallace, D., Chen, F., Li, C., Amann, R., 2018. Unveiling the enigma of refractory carbon in the ocean. Natl. Sci. Rev. 5 (4), 459–463.
Ju, W., Li, J., Yu, W., Zhang, R., 2016. iGraph: an incremental data processing system for dynamic graph. Front. Comput. Sci.-Chi. 10 (3), 462–476.
Kaiser, K., Benner, R., 2008. Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. Limnol. Oceanogr. 53 (1), 99–112.
Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic. Acids. Res. 28 (1), 27–30.
Khomyakova, M., Bükmez, Ö., Thomas, L.K., Erb, T.J., Berg, I.A., 2011. A methylaspartate cycle in haloarchaea. Science 331 (6015), 334–337.
Koch, B.P., Dittmar, T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Sp. 20 (5), 926–932.
Koch, B.P., Kattner, G., Witt, M., Passow, U., 2014. Molecular insights into the microbial formation of marine dissolved organic matter: recalcitrant or labile? Biogeosciences 11 (15), 4173–4190.
Kramer, R.W., Kujawinski, E.B., Hatcher, P.G., 2004. Identification of black carbon derived structures in a volcanic ash soil humic acid by Fourier transform ion cyclotron resonance mass spectrometry. Environ. Sci. Technol. 38 (12), 3387–3395.
Ksionzek, K.B., Lechtenfeld, O.J., McCallister, S.L., Schmitt-Kopplin, P., Geuer, J.K., Geibert, W., Koch, B.P., 2016. Dissolved organic sulfur in the ocean: biogeochemistry of a petagram inventory. Science 354 (6311), 456–459.
Kujawinski, E.B., Behn, M.D., 2006. Automated analysis of electrospray ionization fourier transform ion cyclotron resonance mass spectra of natural organic matter. Anal. Chem. 78 (13), 4363–4373.
Kuzuyama, T., 2017. Biosynthetic studies on terpenoids produced by Streptomyces. J. Antibiot. 70 (7), 811–818.
Kuzuyama, T., Takahashi, S., Dairi, T., Seto, H., 2002. Detection of the mevalonate pathway in Streptomyces species using the 3-hydroxy-3-methylglutaryl coenzyme A reductase gene. J. Antibiot. 55 (10), 919–923.
Landry, Z., Swan, B.K., Herndl, G.J., Stepanauskas, R., Giovannoni, S.J., 2017. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. MBio 8 (2), e413–e417.
Lechtenfeld, O.J., Hertkorn, N., Shen, Y., Witt, M., Benner, R., 2015. Marine sequestration of carbon in bacterial metabolites. Nat. Commun. 6 (1), 1–8.
Li, D., Luo, R., Liu, C., Leung, C., Ting, H., Sadakane, K., Yamashita, H., Lam, T., 2016. MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102, 3–11.
Li, H.Y., Wang, H., Wang, H.T., Xin, P.Y., Xu, X.H., Ma, Y., et al., 2018. The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. Microbiome 6 (1), 1–16.
Li, P., Tao, J., Lin, J., He, C., Shi, Q., Li, X., Zhang, C., 2019. Stratification of dissolved organic matter in the upper 2000m water column at the Mariana Trench. Sci. Total Environ. 668, 1222–1231.
Lian, J., Zheng, X., Zhuo, X., Chen, Y.L., He, C., Zheng, Q., Lin, T.H., Sun, J., Guo, W., Shi, Q., 2021. Microbial transformation of distinct exogenous substrates into analogous composition of recalcitrant dissolved organic matter. Environ. Microbiol. 23 (5), 2389–2403.
Lin, D.Y., 2005. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics 21 (6), 781–787.
Liu, H., Sun, Y., Ramos, K.R.M., Nisola, G.M., Valdehuesa, K.N.G., Lee, W.K., Park, S.J., Chung, W., 2013. Combination of Entner-Doudoroff pathway with MEP increases isoprene production in engineered Escherichia coli. PLoS One 8 (12), e83290.
Liu, S., Parsons, R., Opalk, K., Baetge, N., Giovannoni, S., Bolaños, L.M., Kujawinski, E.B., Longnecker, K., Lu, Y., Halewood, E., 2020. Different carboxyl-rich alicyclic

molecules proxy compounds select distinct bacterioplankton for oxidation of dissolved organic matter in the mesopelagic Sargasso Sea. Limnol. Oceanogr. 65 (7), 1532–1553.
Massey, L.K., Sokatch, J.R., Conrad, R.S., 1976. Branched-chain amino acid catabolism in bacteria. Bacteriol. Rev. 40 (1), 42–54.
McCarren, J., Becker, J.W., Repeta, D.J., Shi, Y., Young, C.R., Malmstrom, R.R., Chisholm, S.W., DeLong, E.F., 2010. Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. Proc. Natl. Acad. Sci. 107 (38), 16420–16427.
Ogawa, H., 2001. Production of Refractory Dissolved Organic Matter by Bacteria. Science 292 (5518), 917–920.
Okamura, E., Tomita, T., Sawa, R., Nishiyama, M., Kuzuyama, T., 2010. Unprecedented acetoacetyl-coenzyme A synthesizing enzyme of the thiolase superfamily involved in the mevalonate pathway. Proc. Natl. Acad. Sci. 107 (25), 11265–11270.
Oksanen, J., Kindt, R., Legendre, P., O Hara, B., Stevens, M.H.H., Oksanen, M.J., et al., 2007. The vegan package. Community ecology package 10 (631–637), 719. https://CRAN.R-project.org/package=vegan.
Osterholz, H., Singer, G., Wemheuer, B., Daniel, R., Simon, M., Niggemann, J., Dittmar, T., 2016. Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. ISME J. 10 (7), 1717–1730.
Pang, B., Li, J., Eiben, C.B., Oksen, E., Barcelos, C., Chen, R., Englund, E., Sundstrom, E., Keasling, J.D., 2021. Lepidopteran mevalonate pathway optimization in Escherichia coli efficiently produces isoprenol analogs for next-generation biofuels. Metab. Eng. 68, 210–219.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14 (4), 417–419.
Pei, G., Chen, L., Zhang, W., 2017. WGCNA application to proteomic and metabolomic data analysis. Meth. Enzymol. 585, 135–158.
Quéré, C.L., Moriarty, R., Andrew, R.M., Peters, G.P., Ciais, P., Friedlingstein, P., Jones, S.D., Sitch, S., Tans, P., Arneth, A., 2014. Global carbon budget 2014. Earth Syst. Sci. Data 6, 235–263.
Refractory, D., 2011. Bacterially Derived Dissolved Organic Matter in the Microbial Carbon Pump. AAAS is here–connecting government to the scientific community 46.
Repeta, D.J., 2015., pp. 21–63, Elsevier.
Rossel, P.E., Vähätalo, A.V., Witt, M., Dittmar, T., 2013. Molecular composition of dissolved organic matter from a wetland plant (Juncus effusus) after photochemical and microbial decomposition (1.25yr): common features with deep sea dissolved organic matter. Org. Geochem. 60, 62–71.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069.
Sever, N., Song, B., Yabe, D., Goldstein, J.L., Brown, M.S., DeBose-Boyd, R.A., 2003. Insig-dependent ubiquitination and degradation of mammalian 3-hydroxy-3-methylglutaryl-CoA reductase stimulated by sterols and geranylgeraniol. J. Biol. Chem. 278 (52), 52479–52490.
Shen, Y., Benner, R., 2020. Molecular properties are a primary control on the microbial utilization of dissolved organic matter in the ocean. Limnol. Oceanogr. 65 (5), 1061–1071.
Siedenburg, G., Jendrossek, D., 2011. Squalene-hopene cyclases. Appl. Environ. Microb. 77 (12), 3905–3915.
Sleighter, R.L., Liu, Z., Xue, J., Hatcher, P.G., 2010. Multivariate statistical approaches for the characterization of dissolved organic matter analyzed by ultrahigh resolution mass spectrometry. Environ. Sci. Technol. 44 (19), 7576–7582.
Su, Y., Peng, B., Li, H., Cheng, Z., Zhang, T., Zhu, J., Li, D., Li, M., Ye, J., Du, C., 2018. Pyruvate cycle increases aminoglycoside efficacy and provides respiratory energy in bacteria. Proc. Natl. Acad. Sci. 115 (7), E1578–E1587.
Sun, Y., Debeljak, P., Obernosterer, I., 2021. Microbial iron and carbon metabolism as revealed by taxonomy-specific functional diversity in the Southern Ocean. ISME J. 1–14.
Takami, H., Taniguchi, T., Moriya, Y., Kuwahara, T., Kanehisa, M., Goto, S., 2012. Evaluation method for the potential functionome harbored in the genome and metagenome. BMC Genomics 13 (1), 1–15.
Tanvir, R.U., Zhang, J., Canter, T., Chen, D., Lu, J., Hu, Z., 2021. Harnessing solar energy using phototrophic microorganisms: a sustainable pathway to bioenergy, biomaterials, and environmental solutions. Renewable Sustainable Energy Rev. 146, 111181.
Tringe, S.G., Von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., 2005. Comparative metagenomics of microbial communities. Science 308 (5721), 554–557.
Wang, R., Zhang, Z., Sun, J., Jiao, N., 2020. Differences in bioavailability of canonical and non-canonical D-amino acids for marine microbes. Sci. Total Environ. 733, 139216.
Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20 (1), 1–13.
Woods, G.C., Simpson, M.J., Simpson, A.J., 2012. Oxidized sterols as a significant component of dissolved organic matter: evidence from 2D HPLC in combination with 2D and 3D NMR spectroscopy. Water Res. 46 (10), 3398–3408.
Wu, X., Wu, L., Liu, Y., Zhang, P., Li, Q., Zhou, J., Hess, N.J., Hazen, T.C., Yang, W., Chakraborty, R., 2018. Microbial interactions with dissolved organic matter drive carbon dynamics and community succession. Front. Microbiol. 9, 1234.
Zark, M., Dittmar, T., 2018. Universal molecular structures in natural dissolved organic matter. Nat. Commun. 9 (1), 1–8.
Zhang, C., 2017. Untangling the role that microbes play in ocean carbon cycle- A new paradigm in marine biogeochemistry. Science China Earth Sciences 60 (2), 409–412.
Zhao, Z., Gonsior, M., Luek, J., Timko, S., Ianiri, H., Hertkorn, N., Schmitt-Kopplin, P., Fang, X., Zeng, Q., Jiao, N., 2017. Picocyanobacteria and deep-ocean fluorescent dissolved organic matter share similar optical properties. Nat. Commun. 8 (1), 1–10.
Zhao, Z., Gonsior, M., Schmitt-Kopplin, P., Zhan, Y., Zhang, R., Jiao, N., Chen, F., 2019. Microbial transformation of virus-induced dissolved organic matter from picocyanobacteria: coupling of bacterial diversity and DOM chemodiversity. ISME J. 13 (10), 2551–2565.
Zheng, Q., Lin, W., Wang, Y., Li, Y., He, C., Shen, Y., Guo, W., Shi, Q., Jiao, N., 2021. Highly enriched N-containing organic molecules of Synechococcus lysates and their rapid transformation by heterotrophic bacteria. Limnol. Oceanogr. 66 (2), 335–348.
Zheng, Q., Wang, Y., Lu, J., Lin, W., Chen, F., Jiao, N., 2020. Metagenomic and metaproteomic insights into photoautotrophic and heterotrophic interactions in a Synechococcus culture. MBio 11 (1), e3219–e3261.
Zhou, L., Zhou, Y., Tang, X., Zhang, Y., Jang, K., Székely, A.J., Jeppesen, E., 2021. Resource aromaticity affects bacterial community successions in response to different sources of dissolved organic matter. Water Res. 190, 116776.