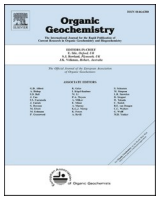
Organic Geochemistry 197 (2024) 104880

Contents lists available at ScienceDirect

Organic Geochemistry

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



Annotation of DOM metabolomes with an ultrahigh resolution mass spectrometry molecular formula library

Nicole R. Coffey [a], Christian Dewey [b], Kieran Manning [c], Yuri Corilo [d], William Kew [d], Lydia Babcock-Adams [e], Amy M. McKenna [e,f], Rhona K. Stuart [g], Rene M. Boiteau [b,c,*]

[a] Department of Earth and Environmental Sciences, University of Minnesota, Minneapolis, MN 55455, USA
[b] Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA
[c] College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, USA
[d] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354, USA
[e] National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32310, USA
[f] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA
[g] Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

ARTICLE INFO

Guest Editor: Jeffrey Hawkes; Executive Guest Editor: Hussain Abdulla

Keywords:
Algal exometabolome
Dissolved Organic Matter
FT-ICR MS
Metabolomics
Liquid Chromatography Mass Spectrometry
Iron stress

ABSTRACT

Increased accessibility of liquid chromatography mass spectrometry (LC-MS) metabolomics instrumentation and software have expanded their use in studies of dissolved organic matter (DOM) and exometabolites released by microbes. Current strategies to annotate metabolomes generally rely on matching tandem MS/MS spectra to databases of authentic standards. However, spectral matching approaches typically have low annotation rates for DOM. An alternative approach is to annotate molecular formula based on accurate mass and isotopic fine structure measurements that can be obtained from state-of-the-art ultrahigh resolution Fourier Transform Ion Cyclotron Resonance mass spectrometry (FT-ICR MS), but instrument accessibility for large metabolomic studies is generally limited. Here, we describe a strategy to annotate exometabolomes obtained from lower resolution LC-MS systems by matching metabolomic features to a molecular formula library generated for a representative sample analyzed by LC 21T- FT-ICR MS. The molecular formula library approach successfully annotated 53% of exometabolome features of the marine diatom *Phaeodactylum tricornutum* – a nearly ten-fold increase over the 6% annotation rate achieved using a conventional MS/MS approach. There was 94% agreement between assigned formula that were annotated with both approaches, and mass error analysis of the discrepancies suggested that the FT-ICR MS formula assignments were more reliable. Differences in the exometabolome of *P. tricornutum* grown under iron replete and iron limited conditions revealed 668 significant metabolites, including a suite of peptide-like molecules released by *P. tricornutum* in response to iron deficiency. These findings demonstrate the utility of FT-ICR MS formula libraries for extending the accuracy and comprehensiveness of metabolome annotations.

1. Introduction

Dissolved organic matter (DOM) represents a major conduit of energy, fixed carbon, and nutrients through the environment. Various factors, including species, growth phase, and nutrient stress, have been shown to alter the chemical characteristics and bioavailability of dissolved organic matter (DOM) released by microorganisms (Grossart and Simon, 2007; Becker et al., 2014; Wear et al., 2015; Tian et al., 2023). However, the specific metabolic processes that govern DOM composition remain a mystery. This limitation stems from the difficulty in separating and identifying individual molecules within the complex mixture of DOM (e.g., Hedges, 2002; Moran et al., 2016; Boiteau et al., 2024). Since the production and consumption of DOM directly influence microbial interactions, such as competition and resource partitioning (e.g., Wang et al., 2022a; Bao et al., 2023), our ability to understand microbial community dynamics is also limited.

Liquid chromatography mass spectrometry (LC-MS) based metabolomic analyses hold promise for filling this knowledge gap. These analyses aim to identify and quantify a wide range of small molecules simultaneously across a biosystem. In a standard LC-MS workflow, data

* Corresponding author at: 228 Smith Hall, 207 Pleasant St SE, Minneapolis, MN 55455, USA.
E-mail address: rboiteau@umn.edu (R.M. Boiteau).

https://doi.org/10.1016/j.orggeochem.2024.104880
Received 18 April 2024; Received in revised form 20 September 2024; Accepted 23 September 2024
Available online 25 September 2024

analysis software is used to determine the mass, retention time, and intensity of chromatographically resolved 'features' that correspond to distinct molecules. To identify molecules, their mass and retention time can be matched to those of authentic standards. When tandem MS/MS capable instruments are used, features are more commonly identified by comparing their fragmentation spectra to databases of MS/MS spectra from authentic standards (e.g., Milman, 2005; Schymanski et al., 2014; Luo et al., 2016, 2023; Folberth et al., 2020). This approach provides confident molecular formula identification of metabolites as well as structural information, with the expectation that molecular isomers with the greatest structural similarity will yield the highest scoring MS/MS spectral match (Schymanski et al., 2014).

Recent applications of liquid chromatography mass spectrometry-based metabolomics have resolved thousands of distinct chemical components that are released by microbes and microbial communities as DOM (the 'exometabolome'), enabling studies that compare their relative concentrations across environmentally relevant organisms (e.g., Becker et al., 2014; Shibl et al., 2020; Brisson et al., 2021; Ferrer-González et al., 2021). However, only a small fraction of the metabolome is typically identified, with the compounds available in spectral databases matching to only 10s to 100s of features out of thousands of those detected. (e.g., Jiménez-Sánchez et al., 2015; Folberth et al., 2020; Charpentier et al., 2022; Wang and Liu, 2023). Molecular networking tools such as SNAP-MS (Morehouse et al. 2023) and GNPS (Wang et al., 2016; Aron et al., 2020; Nothias et al., 2020) seek to expand beyond MS/MS matching to databases by linking related molecules, but still yield low annotation rates (e.g., Nazeh et al., 2024; Shama et al., 2024). Furthermore, structural predictions based on MS/MS fragmentation trees and in silico MS/MS spectral prediction have enabled further annotations of molecules that do not occur in databases (e.g., Wolf et al., 2010; Allen et al., 2014; Ruttkies et al., 2016; Dührkop et al., 2019; Hoffmann et al., 2022; Wang et al., 2022b). Though these tools provide a promising means to increase the annotation rate of metabolomics datasets, they have additional uncertainties and would also benefit from alternative annotation approaches (Chao et al., 2020).

As the leading tool for characterizing the molecular composition of dissolved organic matter, ultrahigh resolution Fourier Transform Ion Cyclotron Resonance mass spectrometry (FT-ICR MS) provides a means to identify a greater proportion of microbial exometabolites (Hendrickson et al., 2015; Shaw et al., 2016). Today's highest resolving mass analyzer, a hybrid linear ion trap/21 Tesla FT-ICR MS that enables automatic gain control (Page et al., 2005), routinely assigns tens of thousands of elemental compositions to marine DOM compounds at sub-ppm mass accuracy with resolving powers >3,000,000 at m/z 200. This custom-built instrument is capable of confidently assigning molecular formulae to complex organic mixtures (e.g. marine DOM, weathered oil; Smith et al., 2018) on chromatographic timescales (Boiteau et al., 2024). The higher resolution and mass accuracy of this instrument provide more confident molecular formula assignments than lower resolution instruments. Such analyses are enabled by advanced data processing tools such as CoreMS (Corilo et al., 2021), a Python-based platform that automates and curates signal processing, calibration, and annotation of mass spectrometry data. However, the cost and accessibility of state-of-the-art LC FT-ICR MS analyses is a significant barrier to conducting large scale metabolomic analyses on such platforms, compared to lower resolution commercially available Orbitrap or time-of-flight instruments that are most used in metabolomic studies.

Here, to address the low annotation rates of tandem MS approaches for metabolomic studies, we developed a novel CoreMS based pipeline that vastly increases the rate of metabolite annotation in environmental samples (Fig. 1). Our approach couples the unparalleled resolving power and mass accuracy of the 21T FT-ICR MS, which we use to generate a library of molecular formulas in small subset of samples, with the relative accessibility of Orbitrap instruments, which we use to detect and identify metabolites in the library across environmental gradients. Similar approaches have been previously employed to increase the confidence and throughput of proteome measurements (Smith et al., 2002).

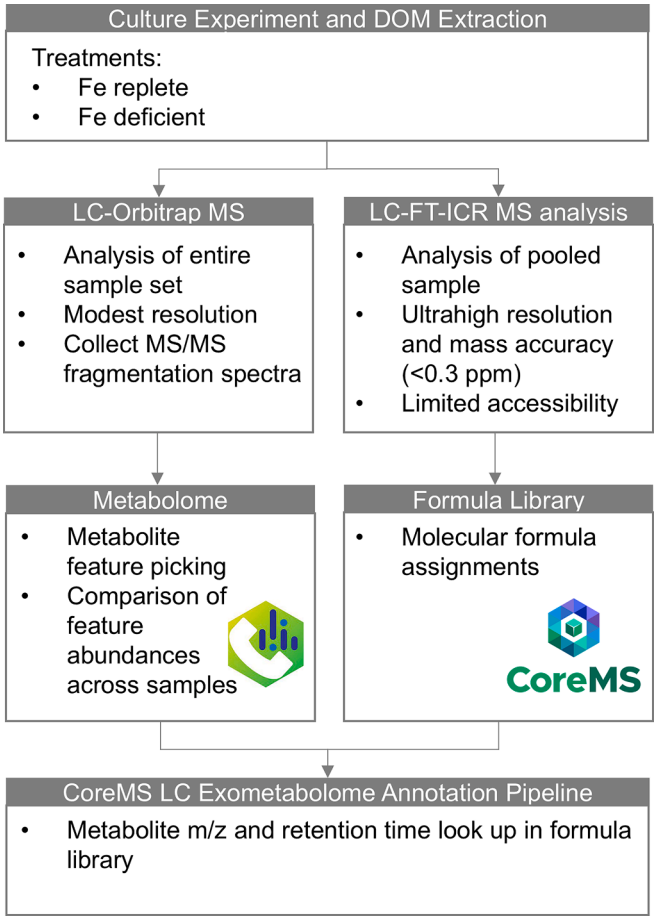


Fig. 1. Overview of CoreMS metabolomics formula assignment workflow. All samples were analyzed by LC-Orbitrap MS, and data was processed in MSDial to yield a feature list with 5403 features. Only 318 had annotations based on MS/MS similarity scoring (5312 of the features had collected MS/MS spectra), highlighting a limitation of this annotation approach. A pooled sample was analyzed with the same chromatography coupled to 21T FT-ICR MS. The mass spectra were binned in 2-minute time intervals and recalibrated based on a set of polysiloxanes observed across the chromatography. Molecular formula assignments were made based on elemental selection criteria described in the methods. This library was used to annotate molecular features in the metabolomics feature list based on matching the mass and retention time of the features in the metabolomic data set and the FT-ICR MS molecular formula library.

As a demonstration, we used our new approach to determine the effects of iron (Fe) deficiency on the exometabolome of *Phaeodactylum tricornutum* (*P. tricornutum)*, a well-studied genetic model for marine diatoms and biofuel algae (Martino et al., 2007; Bowler et al., 2008; Butler et al., 2020; Song et al., 2020). Previous studies found major transcriptional shifts in *P. tricornutum* in response to Fe deficiency, a commonly encountered nutrient stress in the ocean, but the influence on *P. tricornutum's* exometabolome remains largely unknown. Recent LC-MS metabolomic analyses of *P. tricornutum* identified 26 exometabolites and demonstrated how algal exometabolite composition may influence phycosphere bacteria community structure (Brisson et al. 2023). Measuring exometabolome changes under Fe limitation can reveal algal adaptation to stress and its impact on microbial interactions, but requires more complete metabolite annotations. Our results (1) demonstrate how FT-ICR MS formula library annotations complement MS/MS based annotations with added confidence and coverage, and (2) reveal specific changes in *P. tricornutum's* exometabolome in response to Fe

stress, such as decreased lipid production and shifts in protein and phytochemical composition. The formula library annotation strategy is designed to be openly accessible and flexible, enabling its adaptation to other metabolomic studies that would benefit from confident molecular formula annotations.

2. Material and methods

2.1. Culturing and sample generation

Cultures of the diatom (*P. tricornutum* Bohlin strain CCMP 2561) were maintained in modified f/2 medium (Guillard and Ryther, 1962; Smith and Chanley, 1975), with Fe added to a final concentration of 1 μM. Cultures were maintained at 19.0 ± 0.5 °C under cool white, fluorescent lights (2030 lx) on a 12-hour cycle. Parent culture density was monitored by manual cell counts using a hemocytometer, as well as by fluorescence measurements (Spectramax Gemini EM fluorescence microplate reader, λ_{Ex} = 440 nm, λ_{Em} = 680 nm). To pre-treat the diatom for this experiment, the axenic *P. tricornutum* culture was transferred to f/2 with 10 nM dissolved Fe at a cell density of 10^3 cells/mL. This concentration was selected based on initial experiments comparing *P. tricornutum* growth in full f/2 medium (11.7 μM Fe), with 1 μM or 10 nM added iron, and with no Fe added to the medium. The 10 nM Fe condition limited algal growth, so this was chosen as the "Fe deficient" treatment for these experiments. This is on par with findings in previous work, where Fe limitation of *P. tricornutum* was documented at total Fe concentrations of 5–10 nM (Allen et al., 2008; Castell et al., 2021). There was no observed difference between the full f/2 preparation and the 1 μM Fe treatment, so the 1 μM Fe treatment was selected as the "iron replete" condition. The pre-treated parent cultures were allowed to grow until they reached a cell density of ~10^5–10^6 cells/mL (Fe limited cultures could not reach the target cell density of 10^6 cells/mL), then transferred to fresh media to a final concentration of 10^3 cells/mL.

After three rounds of pre-conditioning in medium containing 10 nM Fe, the diatoms were transferred to the Fe replete (1 μM Fe) or Fe deficient (10 nM Fe) experimental conditions. 100 mL f/2 medium in 250 mL polycarbonate flasks (Triforest) were inoculated from the triple pre-treated Fe deficient parent for a starting cell density of 1000 cells/mL in triplicate for both the Fe replete and Fe deficient treatments. Cultures were maintained under the same incubation conditions as the parents and monitored by fluorescence measurements. By day 7, growth differences between the Fe replete and Fe deficient cultures emerged, with significantly greater abundance and relative fluorescence in the Fe replete treatment (p value < 0.05, *t*-test) (Fig. S1). 10 mL subsamples were collected and filtered (0.22 μm, Whatman) after 1, 7, and 16 days of growth (T1, T7, and T16, respectively) for LC-MS metabolomics (Fig. S1). Samples collected at T1 were used to constrain the abundance of exometabolites present within the initial growth media, as cells were still in lag phase and media composition was still reflective of background concentrations. At T7 both cultures were in early exponential growth phase. The relative intensity of exometabolites between Fe deficient versus Fe replete conditions were compared from samples collected when both cultures were in logarithmic growth phase at T16.

2.2. Sample Preparation

Samples were solid phase extracted at their natural pH using 0.1 g PPL cartridges (Agilent Technologies) following standard methods (e.g., Dittmar et al., 2008; Mitra et al., 2013; Boiteau et al., 2019). Cartridges were primed with 3 mL LC-MS grade methanol (VWR Chemical), 3 mL 0.1% trace metal-grade HCl (Fisher Chemical), and 3 mL ultrapure water. 10 mL samples from each culture were filtered (0.22 μm PES, Whatman) and loaded onto the primed columns. Columns were rinsed with 3 mL ultrapure water, and frozen at −20 ℃ until analysis. To prepare for analysis, samples were eluted in 1.5 mL LC-MS grade methanol, dried down to an approximate volume of 50 μL, and rehydrated with ultrapure water to a final volume of 1 mL. 500 μL aliquots of each sample were transferred to 2 mL polypropylene HPLC vials (VWR) and spiked to 1 μM cyanocobalamin (Sigma Aldrich), which served as an internal standard. Equal volumes of each sample from the same time point were mixed to create pooled samples for quality control.

2.3. LC-Orbitrap MS

Chromatographic separation of the sample was performed using a bioinert Ultimate RSLCnano 3000 LC system (Thermo Scientific). The sample loop was filled with 50 μL of sample, which was loaded onto a Zorbax XDB-C18 column (0.5 x 100 mm, 3.5 μm particle size) at 30 μL/min (95% ultrapure water + 5 mM ammonium formate, 5% LC-MS grade methanol + 5 mM ammonium formate). Samples were separated with a 30-minute gradient from 95% solvent A (ultrapure water + 5 mM ammonium formate) and 5% solvent B (methanol + 5 mM ammonium formate) to 95% solvent B, followed by a 5-minute isocratic elution period at 95% solvent B to flush remaining compounds off the column. Solvent composition was switched back to 95% solvent A, and held there for 9 min to re-equilibrate the column before introducing the next sample. All chemicals and solvents were LC-MS Grade (Optima; Fisher Scientific). The column was held at 40 °C throughout the run. Column eluent was directed into a Thermo Scientific Orbitrap IQ-X mass spectrometer equipped with a heated electrospray ionization (ESI) source. ESI source parameters were set to a capillary voltage of 3500 V, sheath, auxiliary and sweep gas flow rates of 5, 2, and 0 (arbitrary units), and ion transfer tube and vaporizer temperatures of 275 °C and 75 °C. MS^1 scans were collected over a *m/z* range of 100–1000 in high resolution (500 k at *m/z* 200, transient length 1024 ms) positive mode. Tandem MS/MS fragmentation spectra were collected in data-dependent acquisition mode in the ion trap, with an HCD energy of 30, a precursor quadrupole isolation width of 1.60, and at an acquisition rate of 35 spectra per second. Dynamic exclusion with a duration of 5 s and a mass tolerance of 10 ppm was used to limit repeated fragmentation of the same precursor.

2.4. LC 21T FT-ICR MS

Chromatographic separation of samples was identical to the method used for data collected using LC-Orbitrap MS. All chemicals and solvents were Honeywell. The column was held at 40 °C throughout the run. The eluent was coupled to a heated electrospray source operated in positive mode (3.75 kV) on the custom-built hybrid linear ion trap FT-ICR mass spectrometer equipped with a 21T superconducting solenoid magnet (Hendrickson et al., 2015). The inlet capillary and source heater temperatures were set to 350 °C and 75 °C, respectively. The sheath and auxiliary gas flow rates were set to 5 and 3 (arbitrary units). MS spectra were collected from *m/z* 200 to 2000 with a resolution of 1,200,000 at 200 *m/z* (transient length of 1536 ms), an AGC target of 1x10^6 charges, and a maximum ion injection time of 500 ms.

2.5. Feature list generation and MS/MS analysis

Individual samples were run by LC-Orbitrap as described above, and processed using MSDial (Tsugawa et al., 2015, 2020; ver.5.1.230517) to generate a feature list. Only peaks with signal intensity above 50,000 amplitude were considered in the feature list to exclude analytical noise. Mass accuracy for singly-charged species was constrained by 0.01 Da and 0.1 Da for MS and MS/MS spectra, respectively. An identification score cutoff of 70% was imposed to limit false matches to the Pos VS17 authentic standard spectral library (324,191 entries). To generate the feature list, features were aligned with a MS^1 tolerance of 0.01 Da and a retention time tolerance of 1 min.

2.6. Library generation and feature annotation

A python-based data processing pipeline was developed to generate a molecular formula library from pooled samples analyzed by LC FT-ICR MS using the open-source platform CoreMS (Corilo et al., 2021) and then used this library to annotate the metabolome feature list. To generate the molecular formula library, mass spectra of the pooled sample were averaged over two-minute time intervals and internally calibrated based on a series of polysiloxane masses that appeared throughout the chromatogram, which improved mass accuracy and reduced processing time compared to annotation of individual spectra. Molecular formulae were assigned using stoichiometric ranges that encompass most biologically relevant compound classes (e.g., Stubbins et al., 2010; Rivas-Ubach et al., 2018a): C 1–50, H 4–100, O 1–16, N 0–8, S 0–2, P 0–1, with a maximum DBE of 18, maximum O/C of 1.2, and maximum H/C of 3. 69.5% of the features found in the LC 21T FT-ICR MS-analyzed pooled sample were assigned a molecular formula. This library was then used to annotate features present in the metabolomics feature list based on matching the mass and retention time of annotated peaks in the formula library. A library match was defined as formula matching the metabolite mass within 3 ppm that occurred within the same 2-minute retention time window, after applying a time offset to account for minor differences in chromatography between the instruments by aligning the retention time of the cyanocobalamin internal standard.

3. Results and discussion

3.1. Exometabolome annotation using LC-MS/MS and standard reference database

Metabolites detected by LC-Orbitrap MS were annotated across the samples using MSDial. A total of 5,403 metabolite features (*m/z* detected at a specific retention time) were compiled. Annotation of these metabolites based on comparison to the MS/MS spectra of 324,191 authentic standards yielded 318 assignments, or an annotation rate of 5.9%. This low annotation rate reflects the need for alternative approaches that yield more comprehensive coverage and motivated the development of the ultrahigh resolution mass spectrometry molecular formula library-based annotation approach described in this manuscript.

The MS/MS annotation strategy was primarily limited by the number of metabolites in spectral libraries, which are focused on small molecules that are commercially available or present in well studied model organisms. Most of the MS/MS annotations were of low molecular weight ($m/z < 300$) features, though some features up to m/z 430 were annotated. 94% of *P. tricornutum* exometabolites' MS/MS spectra did not match any of the 324,191 database spectra, highlighting a major information gap associated with this approach. We note that these libraries continue to grow, and that there are promising emerging tools that can predict fragmentation spectra of molecules without a spectral library (e.g., Dührkop et al., 2019; Hoffmann et al., 2022; Morehouse et al., 2023). However, these tools continue to have significant uncertainty associated with them (Chao et al., 2020) and the size of the putative metabolome far exceeds – by orders of magnitude – the size of any current or future practical experimental database. A second limitation of MS/MS annotation was that fragmentation spectra were only measured for a subset of the most abundant detected features (5,312 out of 5,403 total features; Table 1). Furthermore, some spectra may contain noise or isobaric interferences that resulted in false negative spectral matches, although the frequency of negative matches appears to be low based on manual analysis of the MS/MS matches falling below the 70% threshold.

3.2. LC 21T FT-ICR MS library generation & exometabolome annotation

This study introduces a mass spectrometry data processing pipeline that annotates LC-MS exometabolomes (collected on lower resolution instruments such as Orbitrap or time-of-flight) with confident molecular formula assignment based on paired data from ultrahigh resolution 21T FT-ICR MS. Because this instrument is broadly accessible through user facilities but instrument time can be limited by high demand, the approach is designed to analyze a single or small number of representative 'reference' sample(s), such as a pooled sample containing a mixture of samples across all treatments and replicates.

Molecular formulas were assigned to mass peaks detected in the pooled sample analyzed by LC-FT-ICR MS, which were then compiled into a molecular formula library (Fig. 1). Of the 57,512 detected mass peaks in the FT-ICR MS data, 69.5% were assigned molecular formulas with a mass error of less than 0.3 ppm. The LC-Orbitrap metabolome was then annotated with molecular formula assigned in the FT-ICR MS library (Fig. 1, Fig. S2). A library match was defined as an FT-ICR MS mass matching the metabolite mass within 3 ppm (selected based on the mass accuracy and resolving power of the Orbitrap MS used in this study) that occurred within the same 2-minute retention time window to account for minor differences in chromatography between the instruments.

Of the 5,403 exometabolome features, 3,175 had library matches, and 39 of these had two library matches – these multiple library matches indicate features that were not fully resolved in the Orbitrap mass spectra. In the rare event of multiple library hits, the molecular formula of the library feature with the highest signal to noise ratio was reported, as the unresolved Orbitrap signal is dominated by the most abundant ion. Features with multiple library hits should be considered with caution as they may represent multiple distinct molecules. Of the library matches, 2,840 features were assigned molecular formulae, representing an annotation rate of 53.2% of the exometabolome. To demonstrate the advantage of FT-ICR MS annotations, we also generated a molecular formula library based on the pooled sample analyzed by LC-Orbitrap MS. Annotation of the 5,403 metabolome features using the LC-Orbitrap generated library yielded 2,730 assignments, with 2,320

Table 1
Number of features annotated and not annotated by the application of the FT-ICR MS Library approach and a MS/MS approach using MSDial. Fe replete and Fe deficient features were present in the corresponding treatments at T16 at an abundance > 2fold greater than T1. "Overlapping" features refer to those present in both the Fe deficient and replete treatments. "Significant" features are those that were expressed significantly differently (>2-fold intensity difference, $p < 0.05$) between the Fe deficient and Fe replete treatments.

	All exometabolome features	Background removed features	Fe replete features	Fe deficient features	Overlapping Fe features	Significant features
Total	5403	4031	3444	3003	2670	668
MS/MS assigned	318	212	181	162	148	61
Library assigned	2840	2255	1953	1772	1585	450
Outside library *m/z* range	1433	955	807	635	563	142
Inside library *m/z* range but not detected	834	600	515	446	401	51
Detected in library but not assigned	296	221	169	150	121	25

features overlapping with those assigned using the LC 21T FT-ICR MS library.

The quality of the LC 21T FT-ICR MS molecular formula assignments was assessed by the distributions of *m/z* errors. The *m/z* errors of the FT-ICR MS assignments show a narrow, symmetric, and unimodal distribution centered around zero (Fig. 2), as expected for well calibrated and accurately assigned data. The average *m/z* errors of these assigned features measured by Orbitrap MS also show a symmetric and unimodal distribution, but with tenfold greater standard deviation, highlighting the improved precision of 21T FT-ICR MS.

We also compared the annotations that were made when the metabolome was annotated with the LC 21T FT-ICR MS library versus the lower resolution LC-Orbitrap MS library. Overall, 87% of the features were annotated with the same molecular formulas; 297 formula assignments differed. Both the matched and unmatched features had similar *m/z* error distributions in the case of the LC 21T FT-ICR MS library-based assignments (Fig. S3a). In contrast, the *m/z* error distribution of unmatched features assigned with the Orbitrap MS library was broad and offset compared to the matched assignments (Fig. S3b). These results indicate that the 21T FT-ICR MS library assignments are more consistent with the mass accuracy of the measurements, and are thus more likely to be correct. The improved reliability of the 21T FT-ICR MS assignments results from both improved precision, which reduces the number of potential formulae that could be attributed to a specific *m/z* measurement, as well as resolution, which reduces the frequency of isobaric interferences that decrease *m/z* accuracy.

To gain additional insight into *P. tricornutum*'s annotated exometabolome, the molecular formulas were classified based on stoichiometric elemental ratios (Fig. 3; Fig. S4) following previously established criteria (Rivas-Ubach et al., 2018a). The molecular formulas identified in this analysis were classified as a wide range of biomolecules, encompassing peptides, lipids, carbohydrates, phytochemicals, sulfur-containing organics, and phospholipids. While this is a useful framework for assessing the composition of *P. tricornutum*'s exometabolome, it is worth noting that the stoichiometric classifications are accurate predictions for most, but not all biochemicals, due to some overlap in the stoichiometric ratios of different categories (Rivas-Ubach et al., 2018a). Most of the molecular formulas assigned using the pipeline fell into the lipid, peptide, or phytochemical categories (Fig. 3). As *P. tricornutum* is a model species for biofuel production due to its high lipid content, the dominance of lipids in its exometabolome in this experiment was unsurprising (Chauton et al., 2013; Vandamme et al., 2018). Furthermore, the abundance of peptides is also consistent with the high protein content of *P. tricornutum*, which can account for up to 25% of dry weight (Song et al., 2020). Diatoms are also known to produce a range of polyphenolic phytochemicals, and their production is in part regulated by metal stress (Rico et al., 2013). Our focus on these classes is also due in part to the methodological approach employed; solid phase extraction retains these classes well (Dittmar et al., 2008; Miranda et al., 2023), and reverse-phase liquid chromatography with positive mode electrospray ionization mass spectrometry separates and detects these classes effectively.

3.3. Comparison of MS/MS and LC-21T Library Annotations

To evaluate the strengths and limitations of each strategy, we next compared the annotations of the FT-ICR MS library approach to the MS/MS fragmentation library approach (Fig. 4). A total of 166 metabolites were annotated with both approaches, and 94% of the assigned molecular formula agreed between the two methods, providing added confidence in our library annotations. In the case of the 10 mis-matched formula, the mass error of the MS/MS assigned formula were consistently outside of the expected mass accuracy of the Orbitrap MS (>3 ppm error), except for one molecular formula that was annotated as an isotopologue of another peak by the FT-ICR MS library. These results indicate that the FT-ICR MS formula assignments were more robust. This analysis also demonstrates that MS/MS spectra matching does not

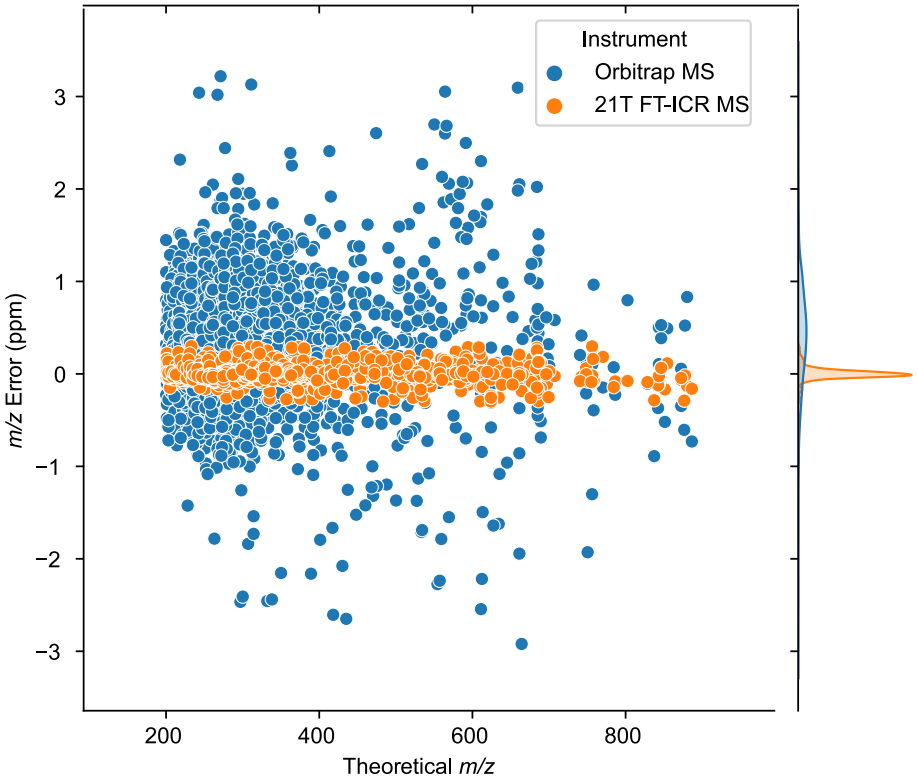


Fig. 2. Scatter plot showing the *m/z* error of FT-ICR MS formula library annotated metabolites. *M/z* error was calculated as the difference between the average *m/z* measured by LC-Orbitrap or LC-FT-ICR MS and the theoretical ion mass, with kernel density estimate distribution to the right. The 10-fold greater mass precision of the 21T FT-ICR MS measurements highlights the advantage of this instrumentation for assigning accurate formula.

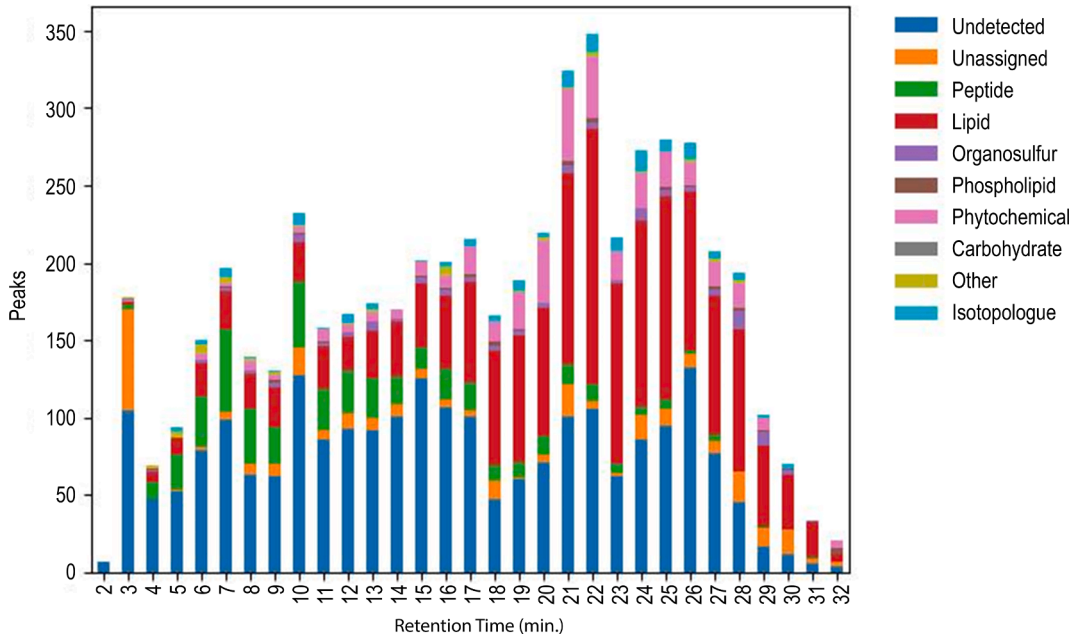


Fig. 3. Stoichiometric classifications of exometabolome features annotated with the formula library approach. Classification categories were based on elemental ratios of molecular formula adapted from Rivas-Ubach et al., 2018a. The 'Undetected' category corresponds to metabolite features that were not detected by LC-FT-ICR MS. The 'Unassigned' category corresponds to metabolite features that were detected by LC-FT-ICR MS but not assigned molecular formula.

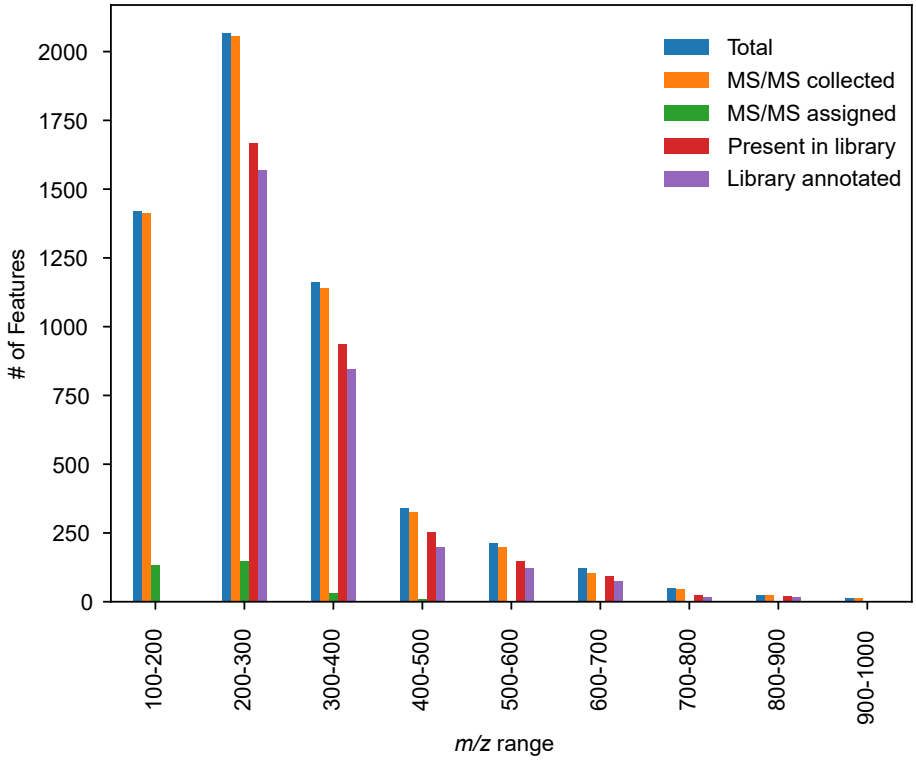


Fig. 4. *m/z* distribution of the total number of features (blue), features with collected MS/MS fragmentation spectra (ion trap of the Orbitrap IQ-X; orange), features assigned by a conventional MS/MS matching approach using MSDial (green), and features present in (red) and annotated with molecular formula (purple) using the 21T FT-ICR MS library annotation approach developed in this study. Despite the high availability of MS/MS spectra for metabolite features, less than 10% were annotated by matching to the MSDial authentic standard spectral library (Pos VS17). Features below *m/z* 200 were below the mass range of the 21T FT-ICR MS formula library and were thus not annotated with this approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

always yield correct molecular formula annotations, even when conservative criteria for mass tolerance and spectral similarity score are used. Narrowing the accepted precursor mass tolerance from 5 ppm to 3 ppm removed the mis-matched assignments, but also removed some annotations that were likely to be correct based on FT-ICR MS annotations. Despite these findings, a major benefit of the MS/MS

fragmentation approach is that the fragment analysis provides additional information on the putative molecular structure rather than just a formula. These two approaches could be used complementarily, with the structural information from the MS/MS match informing on the confident formula assignment from the library approach. However, as the rates of MS/MS annotation are still low, comprehensive coverage of features is still an area to be improved.

Though the annotation rate of the molecular formula library approach was considerably higher than that achieved by the MS/MS matching approach, it still left 46% of the exometabolome unannotated. 59% of these unannotated metabolites were smaller than *m/z* 200 and thus simply below the minimum detectable mass of the 21T FT-ICR MS. 33% of the unannotated features were not detected despite being inside of the FT-ICR MS mass range. This may have been due to differences in the sensitivity of the instruments or because many compounds were diluted in the pooled samples relative to individual samples collected from later time points where higher abundances were generally observed. Finally, only 296 features were detected by the 21T FT-ICR MS and not assigned a molecular formula (Table 1) – 5.5% of the entire dataset. The percent of detected but unassigned features was similar across the *m/z* range (Fig. 4). These metabolites likely include elements that were not considered in the formula assignment criteria used for this study. Of the 3,136 exometabolome features detected in the pooled sample by 21T FT-ICR MS, the molecular formula library pipeline achieved a 90.5% annotation rate. These results emphasize the potential of ultra-high resolution mass spectrometry to greatly improve the coverage of LC-MS exometabolome annotations.

3.4. Differentially expressed metabolites in Fe replete vs Fe deficient conditions

With the development of this new annotation pipeline, we were able to evaluate changes in *P. tricornutum*'s exometabolome between Fe replete and Fe deficient growth conditions. First, we filtered out 1372 background features that did not exhibit at least 2-fold greater maximum abundance in the T16 samples relative to the average of the initial timepoint samples. Of the remaining features, 3444 metabolites accumulated in the culture media under Fe replete conditions (2-fold greater average abundance at late-log versus initial timepoints), while 3003 accumulated under Fe deficient conditions. Many of these metabolites accumulated in both treatments (2670), indicating that they are generally released regardless of Fe condition (Fig. S5).

To investigate metabolites that reflected the Fe stress state of the diatoms, we identified features that were significantly different in abundance between treatments in the late-log samples (T16, >2-fold difference between treatments, Bonferroni adjusted two-tailed *t*-test p-value < 0.05), which yielded a list of 668 features. Of these features, MS/MS matching annotated 61 features, while the molecular formula library annotated 450 (Table 1, Fig. 5). Most of the statistically significant features were more abundant in the Fe replete condition (643 of 668 significant features; Fig. 5b). Since the abundances of most of these metabolites also increased in the Fe deficient cultures relative to the initial time point (Fig. S6), it is likely that their abundance reflects the cell density in the culture rather than the direct effect of Fe stress. In contrast, the 25 metabolites that were in greater abundance in the Fe deficient treatment were mostly absent from the Fe replete treatment, both at the late-log time point as well as the earlier time point when the Fe replete culture had similar cell density to the Fe deficient culture (T7; Fig. S7). Thus, these exometabolites appear to reflect a metabolic response to the Fe deficient culture conditions. Although MS/MS matching was not able to annotate any of these features, eleven were annotated using the FT-ICR MS library approach with *m/z* errors below 0.2 ppm (Fig. 5c; Table 2). This highlights the advantage of the library annotation pipeline for investigating metabolites that are not well-represented in MS/MS fragmentation databases.

The molecular annotations provide insight into the specific effect of Fe stress on exometabolite composition. Peptide and phytochemical-classified compounds exhibit greater shifts in response to Fe deficiency, where some molecules became more abundant while others less abundant. Here, phytochemicals are defined as oxy-aromatic compounds as per constraints proposed by Rivas-Ubach et al. (2018a) and encompass compounds such as flavonoids, lignin, and chlorophylls.

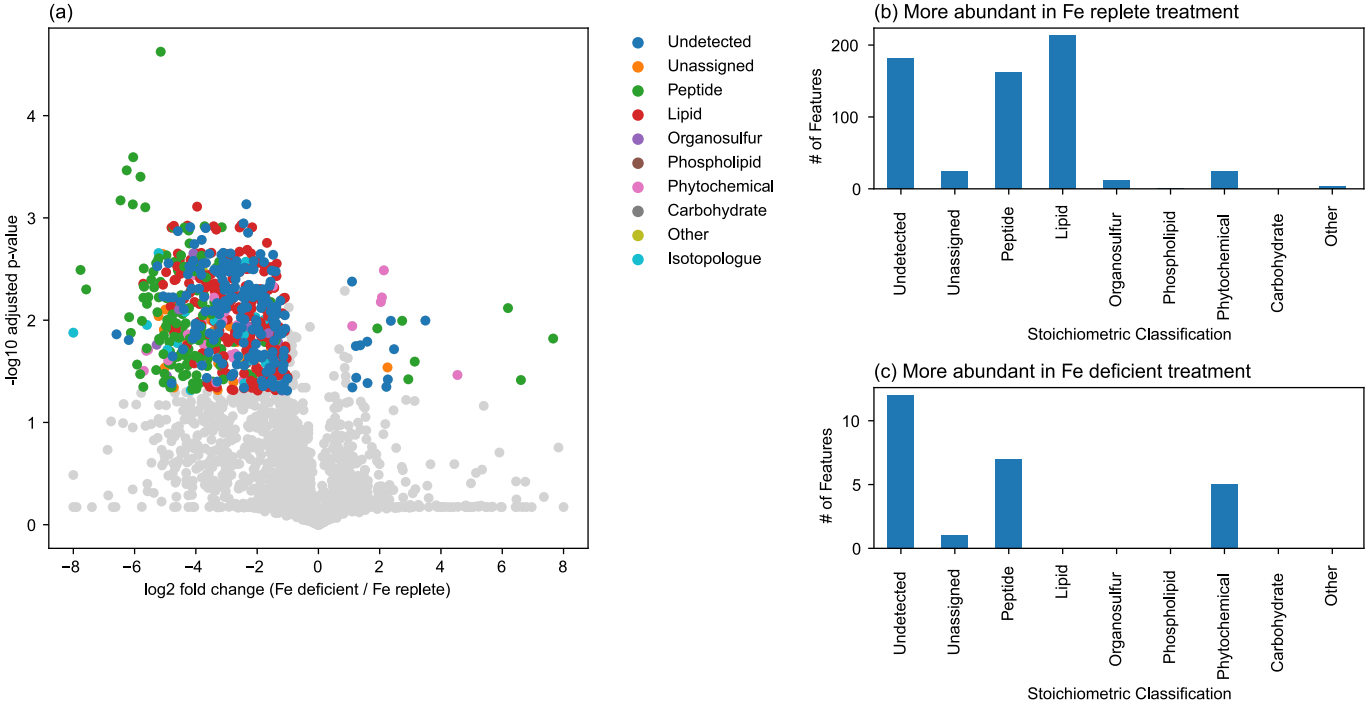


Fig. 5. (a) Volcano plot of metabolites, showing the stoichiometric classification of statistically significant features (adjusted p-value < 0.05) that differ (>2-fold) in relative abundance between the Fe deficient versus Fe replete treatments on Day 16. (b-c) Bar charts showing the number of significantly different features that were more abundant in the Fe replete (b) or Fe deficient treatments (c) belonging to each stoichiometric class. Annotated features in panel (c) are listed in Table 2.

Table 2
Annotated features present at higher abundances in the Fe-deficient condition. These features all fall within stoichiometric classifications of peptides or phytochemicals, likely reflecting changes in chlorophyll biosynthesis under Fe limitation (e.g., Allen et al., 2008; Zhao et al., 2018).

Molecular Formula	Average *m/z* $[M+H]^+$	Retention Time (min.)	FT-ICR MS Library *m/z* Error (ppm)	Stoichiometric Classification	Log 2-fold difference (Fe deficient/ Fe replete)	Adjusted p-value
$C_{13}H_{14}O_3$	219.1015	15.3	+0.072	Phytochemical	2.04	0.0066
$C_{14}H_{14}O_3$	231.1017	12.8	−0.010	Phytochemical	2.14	0.0032
$C_{14}H_{16}O_4$	249.1124	12.9	−0.026	Phytochemical	2.07	0.0060
$C_{11}H_{16}O_3N_4$	253.1297	7.6	−0.034	Peptide	1.92	0.012
$C_{13}H_{14}O_4S_1$	267.0687	16.7	+0.049	Phytochemical	1.10	0.011
$C_{15}H_{28}O_2N_2$	269.2225	19.4	+0.096	Peptide	2.73	0.010
$C_{12}H_{12}O_2N_4S_1$	277.0760	11.0	+0.052	Peptide	6.61	0.038
$C_{13}H_{20}O_3N_4$	281.1610	9.1	+0.041	Peptide	2.93	0.038
$C_{13}H_{17}O_2N_5S_1$	308.1179	8.3	−0.013	Peptide	6.19	0.0076
$C_{12}H_{12}O_4N_4S_1$	309.0655	9.8	+0.036	Peptide	7.66	0.015
$C_{14}H_{34}O_{10}N_5Na_1$	456.2277	12.9	−0.089	Peptide	3.14	0.025

Previous work with *P. tricornutum* and other marine diatoms has documented the down-regulation of genes related to processes with high Fe requirements such as nitrate reduction, as well as shifts in expression of different photosynthetic components (Allen et al., 2008; Lommer et al., 2012; Marchetti et al., 2012; Zhao et al., 2018). Interestingly, although overall chlorophyll content of Fe-starved diatoms is decreased relative to an iron-replete condition (Greene et al., 1991; Rivas-Ubach et al., 2018b), some genes involved in chlorophyll biosynthesis are up-regulated in response to Fe limitation (Allen et al., 2008). These observations may relate to the specific phytochemicals and peptides that were found to increase in relative abundance in the iron-deficient cultures (Fig. 5c). The three molecules that exhibited the largest fold increase in the Fe deficient conditions had very similar molecular formula, $C_{12}H_{12}O_4N_4S_1$, $C_{12}H_{12}O_2N_4S_1$, and $C_{12}H_{17}O_2N_5S_1$, suggesting that they belong to the same class of metabolites. Further work is needed to elucidate the chemical structures of these metabolites and determine their physiological roles. Nevertheless, this analysis demonstrates the utility of FT-ICR MS annotations for providing molecular insight into a new group of *P. tricornutum* exometabolites that relate to Fe stress.

In contrast to the diatom's mixed response with regard to proteins and phytochemical compounds, far fewer lipid-classified molecules accumulated in the Fe limited treatment compared to the Fe replete treatment. This is consistent with previous observations of a significant decline in *P. tricornutum*'s total fatty acid content under Fe limitation (Wang et al., 2023), and appears to reflect a general downregulation of lipid biosynthesis, though this does not seem to be universal for all algae. Lipid production is stimulated by only slight Fe supplementation in *Dunaliella tertiolecta* (Rizwan et al., 2017), and under low Fe conditions in *Nannochloropsis oculata* (Sabzi et al., 2021). However, the decrease in lipid content in response to Fe stress has been observed for other algal species, such as *Chlorella vulgaris* and *Chlorococcum oleofaciens* (Liu et al., 2008; Rajabi Islami and Assareh, 2020). Our results indicate that stress induced by Fe limitation results in a decrease in lipid production by *P. tricornutum*, emphasizing that algal species have unique lipid production responses to Fe stress, which must be considered when optimizing growth conditions for specific purposes, such as biofuel production.

4. Conclusions

The data processing pipeline developed in this study provides an effective strategy to extend LC-MS metabolome annotations based on molecular formula libraries generated by ultrahigh resolution FT-ICR MS. In the current study, the FT-ICR MS library approach annotated ten times the number of features in marine algal exometabolomes relative to MS/MS spectra matching, greatly expanding our window into organisms' molecular responses to their environment. Our findings revealed differences in the release of lipids, peptides, and phytochemicals produced as the diatom experienced Fe limitation. These molecular formula assignments are also useful stepping-stones to structural elucidation, and they enable future targeted studies that assess mechanisms of adaptation or algal-bacterial interactions in response to Fe stress.

Just as databases of MS/MS spectra can be applied to many studies, the LC 21T FT-ICR MS molecular formula library generated in this study can be used as a resource for other LC-MS metabolome studies of *Phaeodactylum tricornutum*. Both the CoreMS data processing pipeline and the molecular formula library are publicly available to enable their application to other LC-MS metabolome data sets that use similar chromatographic conditions. Furthermore, analyses of other exometabolome and DOM 'reference samples' by LC 21T-FT-ICR MS can be used to generate similar molecular formula libraries for other organisms or environments. Ultimately, this approach to increase annotation rates of metabolomic data collected on widely available instrumentation can expedite the discovery of processes that govern the release, transformations, and biological/environmental effects of exometabolites and other molecules that comprise DOM.

CRediT authorship contribution statement

Nicole R. Coffey: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Christian Dewey:** Writing – review & editing, Software, Methodology, Investigation. **Kieran Manning:** Writing – review & editing, Investigation. **Yuri Corilo:** Writing – review & editing, Software, Methodology. **William Kew:** Writing – review & editing, Software, Methodology. **Lydia Babcock-Adams:** Writing – review & editing, Resources, Methodology, Investigation. **Amy M. McKenna:** Writing – review & editing, Resources, Methodology. **Rhona K. Stuart:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Rene M. Boiteau:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

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.

Data availability

MS data is available in the MassIVE repository (accession #MSV000094385). CoreMS LC Metabolome scripts are available to download from Github at https://github.com/boiteaulab/CoreMS-LC-Metabolome-PT.

Acknowledgements

This this research was funded by U.S. Department of Energy (DOE) Office of Biological and Environmental Research, SCW1039. The work at LLNL was performed under DOE auspices of the by Lawrence Livermore National Laboratory, United States, under Contract DE-AC52-07NA27344. A portion of this work was performed at the Ion Cyclotron Resonance user facility at the National High Magnetic Field Laboratory, which is supported by the National Science Foundation Division of Chemistry and Division Materials Research through Cooperative Agreement No. DMR-2128556, NHMFL User Collaborative Grants Program, and the State of Florida. A portion of this research was performed on project awards 10.46936/intm.proj.2021.60081/60000413 and 10.46936/intm.proj.2019.51140/60006695 from the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830. The authors would like to thank Kathleen Kouba for assistance and support with algal culture maintenance. We would also like to thank the reviewers for their constructive comments on this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.orggeochem.2024.104880.

References

Allen, A.E., LaRoche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P.J., Finazzi, G., Fernie, A.R., Bowler, C., 2008. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. Proceedings of the National Academy of Sciences 105, 10438–10443.

Allen, F., Pon, A., Wilson, M., Greiner, R., Wishart, D.S., 2014. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. Nucleic Acids Research 42, W94–W99.

Aron, A.T., Gentry, E.C., McPhail, K.L., Nothias, L.-F., Nothias-Esposito, M., Bouslimani, A., Petras, D., Gauglitz, J.M., Sikora, N., Vargas, F., Van Der Hooft, J.J. J., Ernst, M., Kang, K.B., Aceves, C.M., Caraballo-Rodríguez, A.M., Koester, I., Weldon, K.C., Bertrand, S., Roullier, C., Sun, K., Tehan, R.M., Boya, P., Christian, M. H., Gutiérrez, M., Ulloa, A.M., Tejeda Mora, J.A., Mojica-Flores, R., Lakey-Beitia, J., Vásquez-Chaves, V., Zhang, Y., Calderón, A.I., Tayler, N., Keyzers, R.A., Tugizimana, F., Ndlovu, N., Aksenov, A.A., Jarmusch, A.K., Schmid, R., Truman, A. W., Bandeira, N., Wang, M., Dorrestein, P.C., 2020. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. Nature Protocols 15, 1954–1991.

Bao, Y., Huang, T., Ning, C., Sun, T., Tao, P., Wang, J., Sun, Q., 2023. Changes of DOM and its correlation with internal nutrient release during cyanobacterial growth and decline in Lake Chaohu, China. Journal of Environmental Sciences 124, 769–781.

Becker, J.W., Berube, P.M., Follett, C.L., Waterbury, J.B., Chisholm, S.W., DeLong, E.F., Repeta, D.J., 2014. Closely related phytoplankton species produce similar suites of dissolved organic matter. Frontiers in Microbiology 5, 111. https://doi.org/10.3389/fmicb.2014.00111.

Boiteau, R.M., Till, C.P., Coale, T.H., Fitzsimmons, J.N., Bruland, K.W., Repeta, D.J., 2019. Patterns of iron and siderophore distributions across the California current system. Limnology and Oceanography 64, 376–389.

Boiteau, R.M., Corilo, Y.E., Kew, W.R., Dewey, C., Alvarez Rodriguez, M.C., Carlson, C. A., Conway, T.M., 2024. Relating molecular properties to the persistence of marine dissolved organic matter with liquid chromatography–ultrahigh-resolution mass spectrometry. Environmental Science and Technology 58, 3267–3277. https://doi.org/10.1021/acs.est.3c08245.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R.P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J.A., Brownlee, C., Cadoret, J.-P., Chiovitti, A., Choi, C.J., Coesel, S., De Martino, A., Detter, J.C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M.J.J., Jenkins, B.D., Jiroutova, K., Jorgensen, R.E., Joubert, Y., Kaplan, A., Kröger, N., Kroth, P.G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P.J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L.K., Montsant, A., Secq, M.-P.-O., Napoli, C., Obornik, M., Parker, M.S., Petit, J.-L., Porcel, B.M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T.A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M.R., Taylor, A.R., Vardi, A., Von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L.S., Rokhsar, D.S., Weissenbach, J., Armbrust, E.V., Green, B.R., Van De Peer, Y., Grigoriev, I.V., 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456, 239–244.

Brisson, V., Mayali, X., Bowen, B., Golini, A., Thelen, M., Stuart, R.K., Northen, T.R., 2021. Identification of effector metabolites using exometabolite profiling of diverse microalgae. mSystems 6, e00835–e00921.

Brisson, V., Swink, C., Kimbrel, J., Mayali, X., Samo, T., Kosina, S.M., Thelen, M., Northen, T.R., Stuart, R.K., 2023. Dynamic *Phaeodactylum tricornutum* exometabolites shape surrounding bacterial communities. New Phytologist 239, 1420–1433.

Butler, T., Kapoore, R.V., Vaidyanathan, S., 2020. *Phaeodactylum tricornutum*: a diatom cell factory. Trends in Biotechnology 38, 606–622.

Castell, C., Bernal-Bayard, P., Ortega, J.M., Roncel, M., Hervás, M., Navarro, J.A., 2021. The heterologous expression of a plastocyanin in the diatom *Phaeodactylum tricornutum* improves cell growth under iron-deficient conditions. Physiologia Plantarum 171, 277–290.

Chao, A., Al-Ghoul, H., McEachran, A.D., Balabin, I., Transue, T., Cathey, T., Grossman, J.N., Singh, R.R., Ulrich, E.M., Williams, A.J., Sobus, J.R., 2020. In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples. Analytical and Bioanalytical Chemistry 412, 1303–1315.

Charpentier, T., Boisard, S., Le Ray, A.-M., Bréard, D., Chabrier, A., Esselin, H., Guilet, D., Ripoll, C., Richomme, P., 2022. A descriptive chemical composition of concentrated bud macerates through an optimized SPE-HPLC-UV-MS2 method—application to *Alnus glutinosa*, *Ribes nigrum*, *Rosa canina* ,*Rosmarinus Officinalis* and *Tilia Tomentosa*. Plants 11, 144.

Chauton, M.S., Olsen, Y., Vadstein, O., 2013. Biomass production from the microalga *Phaeodactylum tricornutum*: nutrient stress and chemical composition in exponential fed-batch cultures. Biomass Bioenergy 58, 87–94.

Corilo, Y., Kew, W.R., McCue, L.A., 2021. CoreMS.

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. Limnology and Oceanography: Methods 6, 230–235.

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M., Dorrestein, P.C., Rousu, J., Böcker, S., 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nature Methods 16, 299–302.

Ferrer-González, F.X., Widner, B., Holderman, N.R., Glushka, J., Edison, A.S., Kujawinski, E.B., Moran, M.A., 2021. Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. The ISME Journal 15, 762–773.

Folberth, J., Begemann, K., Jöhren, O., Schwaninger, M., Othman, A., 2020. MS2 and LC libraries for untargeted metabolomics: enhancing method development and identification confidence. Journal of Chromatography B 1145, 122105.

Greene, R.M., Geider, R.J., Falkowski, P.G., 1991. Effect of iron limitation on photosynthesis in a marine diatom. Limnology and Oceanography 36, 1772–1782.

Grossart, H., Simon, M., 2007. Interactions of planktonic algae and bacteria: effects on algal growth and organic matter dynamics. Aquatic Microbial Ecology 47, 163–176.

Guillard, R.R.L., Ryther, J.H., 1962. Studies of marine planktonic diatoms: I. *Cyclotella nana hustedt*, and *Detonula confervacea (cleve)* gran. Canadian Journal of Microbiology 8, 229–239.

Hedges, J.I., 2002. Why dissolved organics matter? In: Hansell, D.A., Carlson, C.A. (Eds.), Biogeochemistry of Marine Dissolved Organic Matter. Academic Press, London, pp. 1–33.

Hendrickson, C.L., Quinn, J.P., Kaiser, N.K., Smith, D.F., Blakney, G.T., Chen, T., Marshall, A.G., Weisbrod, C.R., Beu, S.C., 2015. 21 Tesla Fourier transform ion cyclotron resonance mass spectrometer: a national resource for ultrahigh resolution mass analysis. Journal of the American Society for Mass Spectrometry 26, 1626–1632.

Hoffmann, M.A., Nothias, L.-F., Ludwig, M., Fleischauer, M., Gentry, E.C., Witting, M., Dorrestein, P.C., Dührkop, K., Böcker, S., 2022. High-confidence structural annotation of metabolites absent from spectral libraries. Nature Biotechnology 40, 411–421.

Jiménez-Sánchez, C., Lozano-Sánchez, J., Gabaldón-Hernández, J.A., Segura-Carretero, A., Fernández-Gutiérrez, A., 2015. RP-HPLC–ESI–QTOF/MS2 based strategy for the comprehensive metabolite profiling of *Sclerocarya birrea* (marula) bark. Industrial Crops and Products 71, 214–234.

Liu, Z.-Y., Wang, G.-C., Zhou, B.-C., 2008. Effect of iron on growth and lipid accumulation in Chlorella vulgaris. Bioresource Technology 99, 4717–4722.

Lommer, M., Specht, M., Roy, A.-S., Kraemer, L., Andreson, R., Gutowska, M.A., Wolf, J., Bergner, S.V., Schilhabel, M.B., Klostermeier, U.C., Beiko, R.G., Rosenstiel, P., Hippler, M., LaRoche, J., 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biology 13, R66.

Luo, R.Y., Comstock, K., Ding, C., Wu, A.H.B., Lynch, K.L., 2023. Comparison of liquid chromatography-high-resolution tandem mass spectrometry (MS2) and multi-stage mass spectrometry (MS3) for screening toxic natural products. Journal of Mass Spectrometry and Advances in the Clinical Lab 30, 38–44.

Luo, X., Zhao, S., Huan, T., Sun, D., Friis, R.M.N., Schultz, M.C., Li, L., 2016. High-performance chemical isotope labeling liquid chromatography–mass spectrometry for profiling the metabolomic reprogramming elicited by ammonium limitation in yeast. Journal of Proteome Research 15, 1602–1612.

Marchetti, A., Schruth, D.M., Durkin, C.A., Parker, M.S., Kodner, R.B., Berthiaume, C.T., Morales, R., Allen, A.E., Armbrust, E.V., 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. Proceedings of the National Academy of Sciences 109, E317–E325. https://doi.org/10.1073/pnas.1118408109.

Martino, A.D., Meichenin, A., Shi, J., Pan, K., Bowler, C., 2007. Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. Journal of Phycology 43, 992–1009.

Milman, B.L., 2005. Towards a full reference library of MS^n spectra. Testing of a library containing 3126 MS^2 spectra of 1743 compounds. Rapid Communications in Mass Spectrometry 19, 2833–2839.
Miranda, C., Boiteau, R.M., McKenna, A.M., Knapp, A.N., 2023. Quantitative and qualitative comparison of marine dissolved organic nitrogen recovery using solid phase extraction. Limnology and Oceanography: Methods 21, 467–477.
Mitra, S., Wozniak, A.S., Miller, R., Hatcher, P.G., Buonassissi, C., Brown, M., 2013. Multiproxy probing of rainwater dissolved organic matter (DOM) composition in coastal storms as a function of trajectory. Marine Chemistry 154, 67–76.
Moran, M.A., Kujawinski, E.B., Stubbins, A., Fatland, R., Aluwihare, L.I., Buchan, A., Crump, B.C., Dorrestein, P.C., Dyhrman, S.T., Hess, N.J., Howe, B., Longnecker, K., Medeiros, P.M., Niggemann, J., Obernosterer, I., Repeta, D.J., Waldbauer, J.R., 2016. Deciphering ocean carbon in a changing world. Proceedings of the National Academy of Sciences 113, 3143–3151.
Morehouse, N.J., Clark, T.N., McMann, E.J., Van Santen, J.A., Haeckl, F.P.J., Gray, C.A., Linington, R.G., 2023. Annotation of natural product compound families using molecular networking topology and structural similarity fingerprinting. Nature Communications 14, 308.
Nazeh, M., Asmaey, M., Eladly, A., Afifi, M., 2024. Metabolomic profiling of streptomyces griseorubens with the evaluation of their antioxidant and anticancer potentialities. Egyptian Journal of Chemistry 67, 251–286.
Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A.A., Alka, O., Allard, P.-M., Barsch, A., Cachet, X., Caraballo-Rodriguez, A.M., Da Silva, R.R., Dang, T., Garg, N., Gauglitz, J.M., Gurevich, A., Isaac, G., Jarmusch, A.K., Kameník, Z., Kang, K.B., Kessler, N., Koester, I., Korf, A., Le Gouellec, A., Ludwig, M., Martin, H.C., McCall, L.-I., McSayles, J., Meyer, S.W., Mohimani, H., Morsy, M., Moyne, O., Neumann, S., Neuweger, H., Nguyen, N.H., Nothias-Esposito, M., Paolini, J., Phelan, V.V., Pluskal, T., Quinn, R.A., Rogers, S., Shrestha, B., Tripathi, A., Van Der Hooft, J.J.J., Vargas, F., Weldon, K.C., Witting, M., Yang, H., Zhang, Z., Zubeil, F., Kohlbacher, O., Böcker, S., Alexandrov, T., Bandeira, N., Wang, M., Dorrestein, P.C., 2020. Feature-based molecular networking in the GNPS analysis environment. Nature Methods 17, 905–908.
Page, J.S., Bogdanov, B., Vilkov, A.N., Prior, D.C., Buschbach, M.A., Tang, K., Smith, R. D., 2005. Automatic gain control in mass spectrometry using a jet disrupter electrode in an electrodynamic ion funnel. Journal of the American Society for Mass Spectrometry 16, 244–253.
Rajabi Islami, H., Assareh, R., 2020. Enhancement effects of ferric ion concentrations on growth and lipid characteristics of freshwater microalga *Chlorococcum oleofaciens* KF584224.1 for biodiesel production. Renewable Energy 149, 264–272.
Rico, M., López, A., Santana-Casiano, J.M., Gonzàlez, A.G., Gonzàlez-Dàvila, M., 2013. Variability of the phenolic profile in the diatom *Phaeodactylum tricornutum* growing under copper and iron stress. Limnology and Oceanography 58, 144–152.
Rivas-Ubach, A., Liu, Y., Bianchi, T.S., Tolić, N., Jansson, C., Paša-Tolić, L., 2018a. Moving beyond the van Krevelen Diagram: a new stoichiometric approach for compound classification in organisms. Analytical Chemistry 90, 6152–6160.
Rivas-Ubach, A., Poret-Peterson, A.T., Peñuelas, J., Sardans, J., Pérez-Trujillo, M., Legido-Quigley, C., Oravec, M., Urban, O., Elser, J.J., 2018b. Coping with iron limitation: a metabolomic study of *Synechocystis sp*. PCC 6803. Acta Physiologiae Plantarum 40, 28.
Rizwan, M., Mujtaba, G., Lee, K., 2017. Effects of iron sources on the growth and lipid/carbohydrate production of marine microalga *Dunaliella tertiolecta*. Biotechnology and Bioprocess Engineering 22, 68–75.
Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. Journal of Cheminformatics 8, 3. https://doi.org/10.1186/s13321-016-0115-9.
Sabzi, S., Shamsaie Mehrgan, M., Rajabi Islami, H., Hosseini Shekarabi, S.P., 2021. Changes in biochemical composition and fatty acid accumulation of *Nannochloropsis oculata* in response to different iron concentrations. Biofuels 12, 1–7.
Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environmental Science and Technology 48, 2097–2098.
Shama, S.M., Elissawy, A.M., Salem, M.A., Youssef, F.S., Elnaggar, M.S., El-Seedi, H.R., Khalifa, S.A.M., Briki, K., Hamdan, D.I., Singab, A.N.B., 2024. Comparative metabolomics study on the secondary metabolites of the red alga, Corallina officinalis and its associated endosymbiotic fungi. RSC Advances 14, 18553–18566.
Shaw, J.B., Lin, T.-Y., Leach, F.E., Tolmachev, A.V., Tolić, N., Robinson, E.W., Koppenaal, D.W., Paša-Tolić, L., 2016. 21 Tesla Fourier transform ion cyclotron resonance mass spectrometer greatly expands mass spectrometry toolbox. Journal of the American Society for Mass Spectrometry 27, 1929–1936.
Shibl, A.A., Isaac, A., Ochsenkühn, M.A., Cárdenas, A., Fei, C., Behringer, G., Arnoux, M., Drou, N., Santos, M.P., Gunsalus, K.C., Voolstra, C.R., Amin, S.A., 2020. Diatom modulation of select bacteria through use of two unique secondary metabolites. Proceedings of the National Academy of Sciences 117, 27445–27455.
Smith, W.L., Chanley, M.H. (Eds.), 1975. Culture of marine invertebrate animals: proceedings — 1st conference on culture of marine invertebrate animals Greenport. Springer US, Boston, MA. doi:10.1007/978-1-4615-8714-9.
Smith, R.D., Anderson, G.A., Lipton, M.S., Pasa-Tolic, L., Shen, Y., Conrads, T.P., Veenstra, T.D., Udseth, H.R., 2002. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. Proteomics 2, 513–523.
Smith, D.F., Podgorski, D.C., Rodgers, R.P., Blakney, G.T., Hendrickson, C.L., 2018. 21 Tesla FT-ICR mass spectrometer for ultrahigh-resolution analysis of complex organic mixtures. Analytical Chemistry 90, 2041–2047.
Song, Z., Lye, G.J., Parker, B.M., 2020. Morphological and biochemical changes in *Phaeodactylum tricornutum* triggered by culture media: implications for industrial exploitation. Algal Research 47, 101822.
Stubbins, A., Spencer, R.G.M., Chen, H., Hatcher, P.G., Mopper, K., Hernes, P.J., Mwamba, V.L., Mangangu, A.M., Wabakanghanzi, J.N., Six, J., 2010. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. Limnology and Oceanography 55, 1467–1477.
Tian, L., Zhang, Z., Wang, Z., Zhang, P., Xiong, C., Kuang, Y., Peng, X., Yu, M., Qian, Y., 2023. Compositional variations in algal organic matter during distinct growth phases in karst water. Frontiers in Environmental Science 11, 1112522.
Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., Arita, M., 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nature Methods 12, 523–526.
Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., Okahashi, N., Yamada, Y., Tada, I., Bonini, P., Higashi, Y., Okazaki, Y., Zhou, Z., Zhu, Z.-J., Koelmel, J., Cajka, T., Fiehn, O., Saito, K., Arita, M., Arita, M., 2020. A lipidome atlas in MS-DIAL 4. Nature Biotechnology 38, 1159–1163.
Vandamme, D., Gheysen, L., Muylaert, K., Foubert, I., 2018. Impact of harvesting method on total lipid content and extraction efficiency for *Phaeodactylum tricornutum*. Separation and Purification Technology 194, 362–367.
Wang, F., Allen, D., Tian, S., Oler, E., Gautam, V., Greiner, R., Metz, T.O., Wishart, D.S., 2022b. CFM-ID 4.0–a web server for accurate MS-based metabolite identification. Acta Physiologiae Plantarum 50, W165–W174.
Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M.J., Liu, W.-T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval- Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.-C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.-C., Yang, Y.-L., Humpf, H.-U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya, P., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.-M., Phapale, P., Nothias, L.-F., Alexandrov, T., Litaudon, M., Wolfender, J.-L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.-T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B.Ø., Pogliano, K., Linington, R. G., Gutiérrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nature Biotechnology 34, 828–837.
Wang, Y., Liu, S., 2023. BoxCar data-dependent acquisition improves the MS/MS coverage in liquid chromatography-mass spectrometry-based metabolomics analysis. Arabian Journal of Chemistry 16, 105325.
Wang, H., Su, Q., Zhuang, Y., Wu, C., Tong, S., Guan, B., Zhao, Y., Qiao, H., 2023. Effects of iron valence on the growth, photosynthesis, and fatty acid composition of *Phaeodactylum tricornutum*. Journal of Marine Science and Engineering 11, 316.
Wang, Y., Xie, R., Shen, Y., Cai, R., He, C., Chen, Q., Guo, W., Shi, Q., Jiao, N., Zheng, Q., 2022a. Linking microbial population succession and dom molecular changes in *Synechococcus* -derived organic matter addition incubation. Microbiology Spectrum 10, e02308–e02321.
Wear, E.K., Carlson, C.A., Windecker, L.A., Brzezinski, M.A., 2015. Roles of diatom nutrient stress and species identity in determining the short- and long-term bioavailability of diatom exudates to bacterioplankton. Marine Chemistry 177, 335–348.
Wolf, S., Schmidt, S., Müller-Hannemann, M., Neumann, S., 2010. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinformatics 11, 148.
Zhao, P., Gu, W., Huang, A., Wu, S., Liu, C., Huan, L., Gao, S., Xie, X., Wang, G., 2018. Effect of iron on the growth of *Phaeodactylum tricornutum* via photosynthesis. Journal of Phycology 54, 34–43.