

LOBSTAHS: An Adduct-Based Lipidomics Strategy for Discovery and Identification of Oxidative Stress Biomarkers

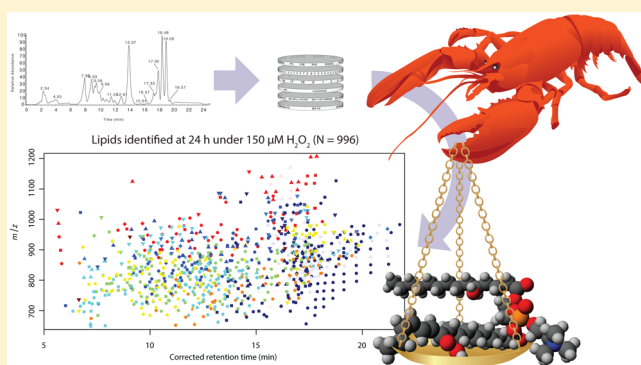
James R. Collins,^{*,†,‡} Bethanie R. Edwards,^{†,‡,§} Helen F. Fredricks,[‡] and Benjamin A. S. Van Mooy[‡]

[†]Massachusetts Institute of Technology/Woods Hole Oceanographic Institution Joint Program in Oceanography, Woods Hole, Massachusetts 02543, United States

[‡]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, United States

S Supporting Information

ABSTRACT: Discovery and identification of molecular biomarkers in large LC/MS data sets requires significant automation without loss of accuracy in the compound screening and annotation process. Here, we describe a lipidomics workflow and open-source software package for high-throughput annotation and putative identification of lipid, oxidized lipid, and oxylipin biomarkers in high-mass-accuracy HPLC-MS data. Lipid and oxylipin biomarker screening through adduct hierarchy sequences, or LOBSTAHS, uses orthogonal screening criteria based on adduct ion formation patterns and other properties to identify thousands of compounds while providing the user with a confidence score for each assignment. Assignments are made from one of two customizable databases; the default databases contain 14 068 unique entries. To demonstrate the software's functionality, we screened more than 340 000 mass spectral features from an experiment in which hydrogen peroxide was used to induce oxidative stress in the marine diatom *Phaeodactylum tricornutum*. LOBSTAHS putatively identified 1969 unique parent compounds in 21 869 features that survived the multistage screening process. While *P. tricornutum* maintained more than 92% of its core lipidome under oxidative stress, patterns in biomarker distribution and abundance indicated remodeling was both subtle and pervasive. Treatment with 150 μM H_2O_2 promoted statistically significant carbon-chain elongation across lipid classes, with the strongest elongation accompanying oxidation in moieties of monogalactosyldiacylglycerol, a lipid typically localized to the chloroplast. Oxidative stress also induced a pronounced reallocation of lipidome peak area to triacylglycerols. LOBSTAHS can be used with environmental or experimental data from a variety of systems and is freely available at <https://github.com/vanmooylipidomics/LOBSTAHS>.



Reactive oxygen species (ROS) represent a persistent source of stress in virtually all biological systems.^{1,2} The negative cellular effects of ROS include protein damage, mutation of DNA, and lipid peroxidation.³ While the profound effects of oxidative stress have been documented extensively in the lipids of mammals,^{4–6} ROS can induce equally significant and wide-ranging remodeling of cell lipidomes in terrestrial and marine plants.^{7–9} These ROS can act through a variety of enzymatic and abiotic mechanisms to produce a broad and heterogeneous suite of lipid products whose bioactivity and diversity make them ideal as molecular biomarkers. These products include both oxidized intact polar lipids (ox-IPL; e.g., oxidized phospholipids)⁸ and oxylipins, the smaller, direct derivatives of fatty acids. Lipid biomarkers (both oxidized and unoxidized) can be used to characterize the effects of ROS in humans from cancer¹⁰ and other diseases such as atherosclerosis;¹¹ in the marine environment, lipids can be used to diagnose various sources of biological and abiotic stress, including those imposed by nutrient limitation^{12,13} and viral

infection.^{14,15} The potency and specificity that make lipids useful as biomarkers of oxidative stress also support their function as bioactive “infochemicals”.^{16–18} In the ocean, for example, oxylipins have been shown to regulate different interspecific interactions among marine microbiota^{19–22} and the metabolism of sinking marine particles by heterotrophic bacteria.²³

For several reasons, there exist few comprehensive methods to screen, identify, and annotate large numbers of these oxylipins and oxidized lipids alongside the many unoxidized lipids from which they can originate.^{4,6} Oxylipins, like their parent lipids, have a wide diversity of structures and biochemical functions.^{4,24,25} They can be produced enzymatically^{9,26,27} or abiotically,^{28,29} often occurring in very low abundance relative to their intact polar lipid (IPL)

Received: March 31, 2016

Accepted: June 20, 2016

precursors.^{6,30} Finally, unique and tailored computational strategies are required to analyze the large volumes of data necessary for comprehensive lipidomics or metabolomics.^{25,31}

The limited number of analytical strategies developed specifically to assess the effects of oxidative stress on the lipidomes of humans⁴ and mammals^{32,33} have generally focused on traditional oxylipins, such as hydroperoxy, hydroxy, epoxy, oxo, and ketol fatty acids, while ignoring most molecular precursors and intermediates. For example, direct-infusion mass spectrometry has been used to identify select oxylipins simultaneously with their unoxidized IPL and ox-IPL precursors in the model plant *Arabidopsis thaliana*, but these studies used manual data analysis methods to examine oxidation of compounds containing only C₁₆ and C₁₈ fatty acids.^{7,8} Ni et al. employed a shotgun approach to identify oxidized lipids in rat cardiomyocytes, but their analysis was limited only to intact carbonylated phospholipids.³⁴ The commercial LipidSearch software (Thermo Scientific) can identify some oxidized lipids using MS/MS fragmentation spectra, but this capability does not extend to oxylipins derived from fatty acids.

We developed a new, rules-based screening approach that can be used to identify a broad range of IPL, nonpolar lipids, ox-IPL, and oxylipins in large, high-mass-accuracy HPLC-ESI-MS data sets. Lipid and oxylipin biomarker screening through adduct hierarchy sequences, or LOBSTAHS, is implemented as an open-source package for R³⁵ and integrates with the existing packages *xcms*^{36–38} and *CAMERA*.³⁹ The software is centered around a novel screening methodology that exploits the unique tendency of lipids to form adduct ions in consistent, diagnostic patterns of abundance that remain relatively consistent across sample types (e.g., for phosphatidylcholine in positive ion mode, $[M + H]^+ > [M + Na]^+ > [M + NH_4 + ACN]^+ > [M + 2Na - H]^+ > [M + K]^+$). Using lipid data from cultures of a mutant strain of a marine diatom designed for studies of oxidative stress,⁴⁰ we demonstrate how LOBSTAHS can be used to resolve conflicting compound assignments, examine differential expression of compounds across experimental treatments, discover and identify potential ox-IPL and oxylipin biomarkers, and identify potential isomers and isobars.

LOBSTAHS annotates each compound assignment with a confidence score, allowing the user to define subsets for further analysis. While we apply LOBSTAHS here to identify potential biomarkers in a marine microorganism, it can be applied to any HPLC-ESI-MS data set where the user expects the relative proportions of the various adduct ions of each analyte to remain constant across samples. LOBSTAHS requires data from a mass spectrometer having both high resolving power and high mass accuracy. While we developed the software using data acquired on a Thermo Exactive Plus Orbitrap instrument, LOBSTAHS could be used to analyze data acquired via FTICR-MS or, when sufficient allowances are made for mass resolution, a Q-TOF instrument.

THEORY AND DESIGN OF SOFTWARE AND DATABASES

Design and Scope of Lipid-Oxylipin Databases. LOBSTAHS draws compound assignments from customizable onboard databases that contain structural and adduct ion abundance data for various IPL, nonpolar lipids, ox-IPL, and oxylipins (Tables S1 and S3). Each database entry represents a different adduct ion of a potential analyte; because analytes present differently in positive and negative ionization modes,

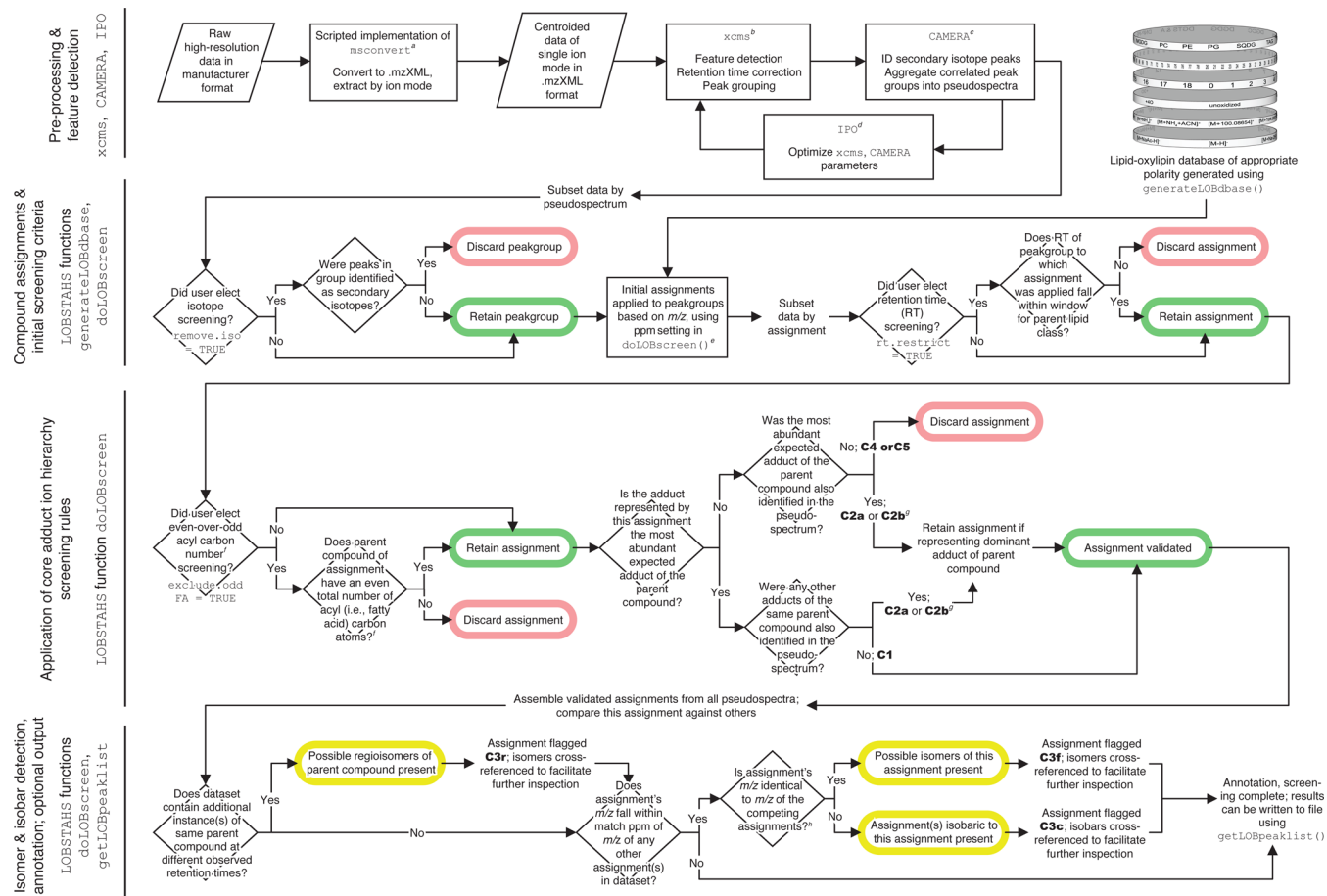
separate databases must be generated for compound identification in each mode. The standard package installation includes two default databases that contain entries for 14 068 unique compounds, some of them particular to marine algae (Tables S1 and S3). Alternatively, LOBSTAHS allows the user to generate his or her own databases; instructions are provided in the package documentation. The use of onboard databases distinguishes LOBSTAHS from other software packages that rely exclusively on external databases.

Database Generation. Databases are created in LOBSTAHS by pairing empirical data with an in silico simulation. To generate the default databases, we first calculated exact masses for various triacylglycerols (TAG), free fatty acids (FFA), polyunsaturated aldehydes (PUA), and molecules belonging to eight different classes of intact polar diacylglycerol (IP-DAG). Within each of these classes, we calculated the masses of a wide range of possible structures having fatty acid (FA) moieties of different acyl chain length, unsaturation, and oxidation (Table S1). While the default databases include entries for IPL and ox-IPL that contain primarily medium- and long-chain fatty acids, users may generate additional databases with entries for IPL composed of fatty acids of any length. The databases also include exact masses for several photosynthetic pigments common to the marine environment (Tables S1 and S3). TAG and IP-DAG are identified by the “sum composition”⁴¹ of double bonds and acyl carbon atoms in each compound (e.g., PC 34:1, rather than PC 16:0–18:1).

Determination of Relative Abundances of Adduct Ions for Inclusion in Databases. During database generation, LOBSTAHS uses empirical data for the LC/MS adduct ion(s) typically formed by each compound’s parent lipid class (Table S2) to create several entries for each compound. Each of the entries represents a different adduct ion of its parent; the relative ranks of the adducts form the basis for the hierarchy-based screening of compound assignments at the core of our method. Relative adduct ion abundance data were gathered from previous work⁴² and analysis of compounds commonly observed in cultures and environmental samples from marine microorganisms. Where possible, we confirmed the results using authentic standards for representative compounds. For ox-IPL, we applied adduct hierarchies observed for the corresponding, unoxidized IPL. We assumed ox-IPL would be unlikely to take charge on their oxidized functional group(s) during ionization, thus forming adducts similar to those of the corresponding unoxidized molecule. We confirmed this similarity in ionization behavior through manual inspection of several samples. Long-chain ox-IPL standards other than those containing aldehyde moieties are not available commercially.

A series of simple tables can be used to define additional analytes and/or adducts beyond those which are included in the default databases. For each new lipid or lipid class, LOBSTAHS requires (1) the elemental composition of the new lipid or parent moiety of the new lipid class, (2) a tabulation of expected adducts (defining, as necessary, any new adducts), (3) empirical adduct hierarchy data for any new adducts, and (4) if applicable, the ranges of acyl carbon atoms, double bonds, and oxidation states for which entries are to be generated. Specific instructions are contained in the online documentation for the software.

Lipidomics Workflow Based on *xcms*, *CAMERA*, and LOBSTAHS. Because high resolving power and high mass accuracy alone are often not sufficient to resolve isobaric ions

Scheme 1. Preparation, Screening, and Annotation of HPLC-MS Lipid Data Using LOBSTAHSⁱ

^aWe automate several functions of the ProteoWizard msConvert tool.⁴³ ^{b,c}xcms^{36–38} was chosen for its command-line features and because it permits follow-on use of the R package CAMERA³⁹ to identify isotopes. ^dIPO⁵⁴ can be used to optimize the values of parameters for some xcms and CAMERA functions. ^eMultiple assignments will likely exist for many peakgroups in a typical data set. ^fThis criterion may be useful when the subject data set contains lipids of exclusively eukaryotic origin. ^gIn the case of C2a, the adduct ion hierarchy for the parent compound is completely satisfied; i.e., the pseudospectrum contains peakgroups representing every adduct ion of the compound of greater theoretical abundance than the least abundant adduct ion present. In the case of C2b, the adduct ion of greatest theoretical abundance and some lesser adduct ion is present, but adduct ions of intermediate abundance are not observed. ^hBoth outcomes may apply simultaneously at this decision point if the data set contains isobars and isomers of the assignment. ⁱAnnotation codes (in bold) may be applied as indicated; these are designed to assist the user in evaluating assignment confidence during subsequent data analysis.

or to distinguish among species that have identical mono-isotopic masses but different elemental compositions,⁴⁴ we employ high-performance liquid chromatography (HPLC) in lieu of a simple direct-infusion approach. Data must first be converted to an open-source format (.mzXML) and, if necessary, from profile to centroid mode. If data are acquired using ion mode switching, LOBSTAHS requires that positive and negative ion mode scan data also be extracted into separate files. Procedural details and an R script that can be used to automate these steps for data obtained from an Orbitrap instrument are described in the [Supporting Information](#). For purposes of the present study, we consider two ions to be isobaric when the underlying features have different exact masses but the m/z difference is less than the instrument's demonstrated mass accuracy (in this case, 2.5 ppm). Once data files have been converted and extracted, the existing R packages xcms^{36–38} and CAMERA³⁹ are used for feature detection, peak grouping, chromatographic alignment, identification of pseudospectra, and discovery of features representing possible

secondary isotope peaks (Scheme 1, “Pre-processing & feature detection”).

Database Assignments and Progressive Screening in LOBSTAHS Using Orthogonal Criteria. After preprocessing, screening and annotation are performed according to the workflow in Scheme 1. First, initial compound assignments are applied to features from the database using a narrow m/z mass tolerance specified by the user in ppm. A series of optional orthogonal screening criteria can then be applied to the features and their assignments. First, users may exclude from the data set any features representing secondary isotope peaks; the presence of these features can be a significant obstacle in metabolomics.^{45–47} We programmed LOBSTAHS to exclude these secondary peaks, rather than merge the elements of each feature's isotopic envelope into a single parameter.⁴⁶

Next, LOBSTAHS can screen the feature's retention time against a retention time “window” defined for the accompanying assignment's parent lipid class. LOBSTAHS includes a set of default retention time window data (Table S4) for the chromatographic conditions we describe here. Detailed

Table 1. Evaluation of Lipidomics Method Performance using IPL Standards

lipid class	origin of standard	moieties present in standard ^a	dominant positive mode adduct ion	ion exact <i>m/z</i>	observed <i>m/z</i> ^b	rel. mass uncertainty (ppm) ^c	correct LOBSTAHS ID?	confidence in assignment after adduct hierarchy screening ^d	structural isomers or isobars present after screening?
MGDG	natural	34:0	[M + NH ₄] ⁺	776.6246	776.6248	0.2	yes	high	no
		36:0	[M + NH ₄] ⁺	804.6559	804.6561	0.3	yes	high	no
DNP-PE	synthetic	32:0	[M + NH ₄] ⁺	875.5505	875.5507	0.2	yes	high	no
SQDG	natural	34:3	[M + NH ₄] ⁺	834.5396	834.5398	0.2	yes	high	no
		34:2	[M + NH ₄] ⁺	836.5552	836.5554	0.2	yes	high	no
PG	synthetic	32:0	[M + NH ₄] ⁺	740.5436	740.5438	0.3	yes	high	no
PE	synthetic	32:0	[M + H] ⁺	692.5225	692.5227	0.3	yes	high	no
PC	synthetic	32:0	[M + H] ⁺	734.5694	734.5696	0.2	yes	high	no
DGDG	natural	34:2	[M + NH ₄] ⁺	934.6462	934.6463	0.1	yes	high	yes
		36:4	[M + NH ₄] ⁺	958.6462	958.6463	0.1	yes	high	yes
mean						0.2			

^aMultiple moieties were present in glycolipid standards purified from natural samples; only predominant moieties are shown. ^bMean observed *m/z* ratio in 5 independent samples. ^cSee the following equation:

$$\left| \frac{\text{measured exact mass} - \text{calculated exact mass}}{\text{calculated exact mass}} \right| \times 10^6$$

^d“High confidence”: Assignment fully satisfied all adduct hierarchy rules and other screening criteria.

instructions for application of retention time data from other chromatographic methods are included in the [Supporting Information](#). A third filter can then be applied to exclude assignments of IPL, ox-IPL, FFA, and PUA that contain an odd total number of acyl carbon atoms. We envision that this filter would be applied to data of exclusively eukaryotic origin: Since nonacetogenic fatty acid synthesis is confined almost exclusively to bacteria and archaea,⁴⁸ FA synthesized by eukaryotes will be composed of an even total number of carbon atoms.

After applying these initial optional criteria, LOBSTAHS then screens each assignment using adduct ion hierarchy data ([Scheme 1](#), “Application of core adduct ion hierarchy screening rules”; [Table S2](#)). This screening serves as the primary orthogonal filter to eliminate any confounding secondary isotopes and unassigned lipid-extractable features still remaining in the data set. During this process, LOBSTAHS uses a series of rules to compare the relative abundance ranks of sets of adduct ion assignments that have the same parent compound. The package makes several annotations using simple codes that indicate the degree to which the assignment complies with the hierarchy rules ([Scheme 1](#), in bold). Assignments that fail the adduct ion hierarchy screening criteria are excised from the data set, and all remaining assignments in the data set are then pooled.

Additional rules-based screening is then performed on the pooled data to identify and annotate possible isomers and isobars ([Scheme 1](#), “Isomer & isobar detection, annotation”). Codes can be applied to identify positional or regioisomers (code C3r), functional structural isomers (code C3f), or isobars (code C3c). LOBSTAHS can apply several of these different codes to a given assignment as long as the criterion for each is satisfied. Upon completion of screening, LOBSTAHS produces an R object containing the annotated data set. Statistical analysis can then be performed in R on the final matrix of compound assignments, or the results can be exported to a .csv file for external analysis.

EXPERIMENTAL SECTION

Model Data Set Used To Demonstrate the Workflow.

To demonstrate LOBSTAHS, we applied the workflow in

[Scheme 1](#) to examine the effect of oxidative stress on a model algal lipidome. For the analysis, we used lipid data collected from cultures of a mutant strain of the marine diatom *Phaeodactylum tricornutum*, which was designed for studies of oxidative stress. In the study,⁴⁰ a strain of *P. tricornutum* (CCMP2561; Provasoli-Guillard National Center for Marine Algae and Microbiota) was genetically modified⁴⁹ to express a reduction–oxidation sensitive green fluorescent protein (roGFP) at different locations within the cell.^{50,51} Cultures of the transformants were treated with three concentrations of H₂O₂ (0, 30, and 150 μmol L^{−1}) to evaluate the effects of peroxidation; culture conditions are described in van Creveld et al.⁴⁰

Sample Collection and Extraction. In the experiment, duplicate samples for lipid analysis were collected from each treatment at 4, 8, and 24 h time points. Two procedural blanks were also collected. Sample material was collected by vacuum onto 0.7 μm pore size glass fiber filters (GF/F), which were snap frozen in liquid nitrogen and then stored at −80 °C until thawed for extraction. Extraction was performed using a modified Bligh and Dyer⁵² method described in Pendorff et al.;⁴² an internal standard (dinitrophenyl-phosphatidylethanolamine, DNP-PE) and a synthetic antioxidant (butylated hydroxytoluene, BHT) were added at the time of extraction. Lipid extracts were transferred to 2 mL HPLC vials, topped with argon, and stored at −80 °C prior to analysis. All chemicals used in sample extraction and chromatography were LC/MS grade or higher. Where used, water was obtained from a Milli-Q system without further treatment (EMD Millipore, Billerica, MA, USA).

HPLC-ESI-MS Analysis. Samples from the *P. tricornutum* data set were analyzed by HPLC-ESI-MS using a modification of the method described in Hummel et al.⁵³ Lipid extracts were evaporated to near dryness and reconstituted in a similar volume of 7:3 acetonitrile/isopropanol. Headspace was filled with argon to minimize further oxidation. For HPLC analysis, an Agilent 1200 system (Agilent, Santa Clara, CA, USA) comprising temperature-controlled autosampler (4 °C), binary pump, and diode array detector was coupled to a Thermo Exactive Plus Orbitrap mass spectrometer (ThermoFisher

Scientific, Waltham, MA, USA). Chromatographic conditions, electrospray ionization source settings, MS acquisition settings, and procedures used for calibration of the mass spectrometer are described in the [Supporting Information](#). Using authentic standards and two independent methods for MS feature detection, we determined the average relative mass uncertainty of the exactive was <0.2 ppm ([Tables 1 and S6](#)); evaluation of these standards is discussed below.

Analysis of *P. tricornutum* Data Using LOBSTAHS. xcms, CAMERA, and LOBSTAHS were then used to identify and annotate lipidome components in the positive ionization mode data. The R package IPO⁵⁴ was used to optimize settings for several xcms functions, and a 2.5 ppm mass uncertainty tolerance was used to obtain database matches in LOBSTAHS. Using the annotated output we obtained from LOBSTAHS, we then calculated the relative abundances of lipidome constituents present in the 0 and 150 μM H_2O_2 treatments at 24 h. Statistical techniques were used to identify biomarkers of oxidative stress. Unless otherwise noted, we restricted our analysis to only “high confidence” assignments; these were assignments without structural isomers or isobars given codes of C1 or C2a according to the logic in [Scheme 1](#). The specific settings used in xcms, CAMERA, and LOBSTAHS, details of statistical methods, and links to scripts we used to obtain results and figures are included in the [Supporting Information text and Table S5](#).

RESULTS AND DISCUSSION

Screening and Annotation of *P. tricornutum* Data in LOBSTAHS. Using LOBSTAHS, we identified 21 869, or 6.4%, of the 340 991 mass spectral features initially detected in the data set using xcms. Sequential application of the various screening criteria allowed us to exclude features from the data set based on specific characteristics ([Table 2](#)). Of these initial features, 177 053, or 52%, were immediately eliminated as likely

secondary isotope peaks identified by CAMERA. The 163 938 remaining features were then matched at 2.5 ppm against entries in the default positive mode database. We then used LOBSTAHS to perform screening based on feature retention time and assignment total acyl carbon number. LOBSTAHS excluded 7792 features because the retention time fell outside the range expected for the assignment’s parent lipid class. An additional 7733 features were eliminated because the compound assignment did not contain an even total number of acyl carbon atoms; this optional restriction was applied given the known eukaryotic origin of the data. Adduct ion hierarchy screening was then applied to the remaining 52 337 features. Application of this final orthogonal filter yielded a data set containing 2056 compound assignments; these assignments represented 1969 unique parent compounds ([Table 2](#)).

The identities of 1163, or 57%, of these final database assignments were unique within the scope of our database, meaning the underlying features were matched in the final data set to only one possible parent compound. 1149 of these assignments were IPL, ox-IPL, or TAG ([Figure 1a](#)); the remainder were photosynthetic pigments. We classified 1056 of these identifications as “high confidence,” indicating that the distribution of adducts present in the constituent features perfectly satisfied the adduct hierarchy rules ([Figure 1a](#) and symbols with darkest tones in [Figures S2–S10](#)); these were used in the analysis below.

Identification and Annotation of Isomers and Isobars.

The remaining 893 assignments (43.4%) were characterized by some degree of ambiguity, meaning the data set contained at least one isobar or structural functional isomer of the underlying features ([Table S7](#); symbols with lightest tones in [Figures S2–S10](#)). In 752 instances, the dominant adduct of the parent compound was a (functional) structural isomer of the dominant adduct of a different compound assigned from the database ([Chart S1](#), first example). In 195 cases, the dominant adduct ion of the parent compound was an isobar of the primary adduct ion of a different compound ([Chart S1](#), second example). Although these ambiguous assignments represented 43.4% of all assignments in the screened data set, they belonged to just 25% of retained features (27% of peak groups; [Table S7](#)). The difference was due to the presence of a small number of features (793) whose 54 assignments were doubly ambiguous, i.e., having both isobars and functional structural isomers (symbols with two-tone shading in [Figures S2–S10](#); [Chart S1](#), third example). The number of competing assignments for each identified compound varied largely by lipid class. For example, LOBSTAHS found no functional structural isomers for compounds identified in several lipid classes: digalactosyldiacylglycerol (DGDG), phosphatidylethanolamine (PE), and sulfoquinovosyldiacylglycerol (SQDG) ([Figures S3, S7, and S9](#)). Doubly ambiguous assignments were confined to only four classes: diacylglyceryl carboxyhydroxymethylcholine (DGCC), diacylglyceryl trimethylhomoserine and diacylglyceryl hydroxymethyl-trimethyl- β -alanine (DGTS and DGTA), phosphatidylcholine (PC), and phosphatidylglycerol (PG) ([Figures S2, S4, S6, and S8](#)).

Annotation of Potential Regioisomers. LOBSTAHS also identified regioisomers for 352 unique parent compounds in the *P. tricornutum* lipidome ([Table S7](#); symbols with black dots in [Figures S2–S10](#)). These were instances in which the same assignment was applied to two or more features appearing at different retention times in the same sample. Many of these assignments were oxylipins and ox-IPL, indicating the presence

Table 2. Progressive Screening and Annotation of the *P. tricornutum* Dataset using xcms, CAMERA, and LOBSTAHS

operation(s) applied	no. present in data set			
	peaks	peak groups	database assignments ^a	unique parent compounds
xcms and CAMERA				
initial feature detection; preprocessing	340 991	18 314		
LOBSTAHS				
eliminate secondary isotope peaks	163 938	12 146		
apply initial compound assignments from database	67 862	5077	15 929	14 076
apply RT screening criteria	60 070	4451	13 504	11 779
exclude IP-DAG/TAG with odd total no. of acyl C atoms	52 337	3871	7458	6283
adduct ion hierarchy screening	21 869	1595	2056 ^b	1969

^aFigure reflects all assignments from database, including photosynthetic pigments. ^b1163, or 57%, of these had no competing assignments such as functional structural isomers or isobars; these 1163 assignments represented 990 unique parent compounds.

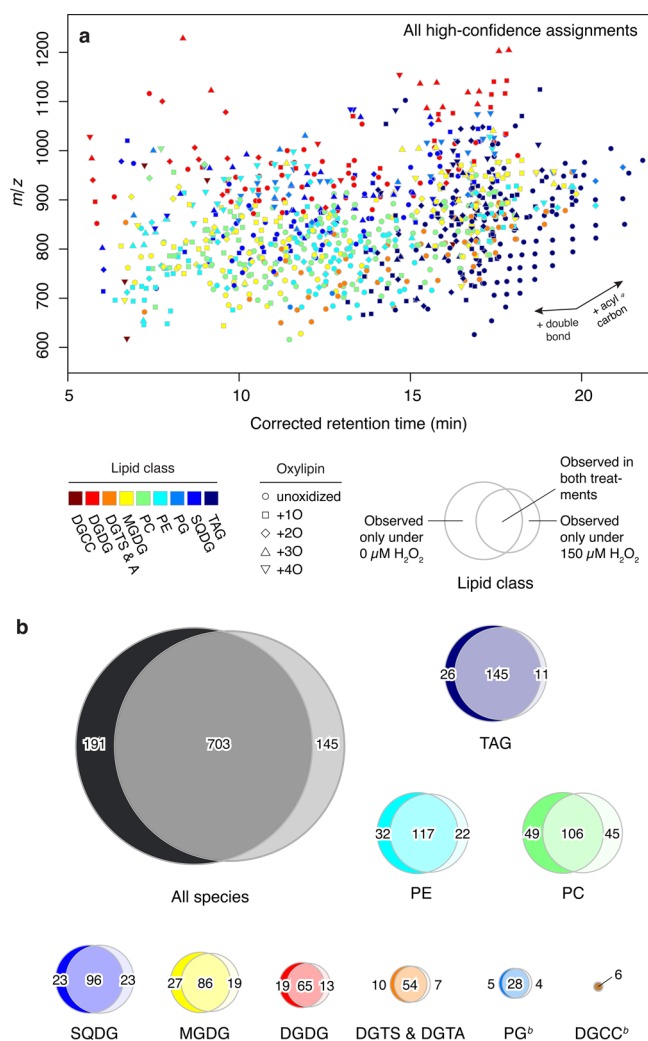


Figure 1. (a) All IPL, ox-IPL, and TAG identified in the *P. tricornutum* data set with high confidence ($N = 1039$; figure excludes pigments). (b) Distribution by lipid class of high-confidence assignments made in the 0 and 150 μM H_2O_2 treatments at 24 h ($N = 894$ and $N = 848$, respectively). Ellipse size in (b) reflects the number of compounds identified within each class and treatment. The assignments presented in (a) and (b) fully satisfied the LOBSTAHS adduct hierarchy screening criteria (i.e., annotated “C1” or “C2a” according to the logic in Scheme 1) and had no competing assignments, such as possible structural isomers, identified in the data set. Excluded are those compounds having an odd total number of acyl carbon atoms. ^aGeneral direction of movement within m/z versus RT plot, for a given lipid class and oxidation state. The direction of movement that results from addition or removal of additional oxygen atom(s) varies by lipid class. ^bNot to scale.

of multiple oxidized isomers of the same parent IPL that could be used as biomarkers for oxidative stress. Without further analysis, we were unable to determine whether these isomers represented the oxidation of a fatty acid by the same mechanism at a different acyl carbon position or instead the presence of different oxidized functional groups that yielded equivalent exact masses (e.g., a dihydroxy-, hydroperoxy-, or α - or γ -ketol acid). The level of identification and annotation provided by LOBSTAHS supports a wide range of possible molecular structures for each assignment; an example from the data set is presented in Chart S1. The data were consistent with studies in both model plant^{7,8} and animal³⁴ systems that

demonstrated the coexistence of a diversity of ox-IPL with both their parent IPL and smaller, traditional oxylipin degradation products.

Evaluation of Screening and Identification Performance Using Two Methods. As a means of validating the accuracy and reliability of our approach, we asked LOBSTAHS to identify and annotate all species present in 5 quality control (QC) samples of known composition that were interspersed randomly with samples from the *P. tricornutum* data set prior to analysis on the mass spectrometer (Table 1; Table S6). The samples contained a mixture of authentic IPL standards that has been used extensively in other work in our laboratory.^{42,55}

Because the choice of preprocessing software can have a significant impact on feature detection,⁵⁶ we conducted parallel analyses with both xcms/CAMERA and an alternative program, MAVEN.^{47,57} In both cases, LOBSTAHS correctly identified all components of the standard mixture without ambiguity (Table 1; Table S6). Because purified standards do not exist for long-chain ox-IPL, we were unable to directly evaluate the Type 1 error rate for their identification. As a second means of validation, we compared assignments in the screened data set to two independent inventories of the *P. tricornutum* lipidome.^{58,59} LOBSTAHS found and identified with high confidence 13 of the 16 most abundant IPL and TAG species in one inventory⁵⁸ and nearly all those in the other.⁵⁹ While our approach also correctly identified the three remaining components in the former study (PG 32:1, PG 36, and DGTS and DGTA 40:10), functional structural isomers or isobars made unambiguous identification impossible without further manual inspection of ms^2 spectra.

Resilience of Core *P. tricornutum* Lipidome under Oxidative Stress. Evidence of the effect of oxidative stress on the lipidome of *P. tricornutum* was observed through comparison of compounds identified in 0 and 150 μM H_2O_2 treatments at 24 h (Figures 1 and 2a,b and Table S9). That the two treatments produced only subtle differences in molecular diversity (Figure 1b) suggests much of the core lipid inventory remained robust to the imposed oxidative stress. The vast majority of the 949 oxidized and unoxidized lipid moieties we identified in the healthy organism (879, or 92.6%) could still be identified in the lipidome of the stressed cultures (Figure 1b). On the basis of peak area, oxidized lipid moieties accounted for 5–7% of the *P. tricornutum* lipidome across nearly all treatments and time points. On the basis of the relatively consistent size of this oxidized lipid fraction and its persistence in even the 0 μM H_2O_2 treatment, we consider it a quantitative constraint on the baseline level of lipid peroxidation associated with metabolic processes in photosynthetic organisms.^{1,2}

Differences in Degree of Remodeling between Lipid Classes and Functional Groupings. We used similarity profile analysis of the scaled LOBSTAHS data⁶⁰ to place the annotated features into 181 groups of components which clustered significantly according to their behavior (Figures 2 and S11). The components of each group are given in Table S8. We further examined the up- and downregulation of lipidome components under oxidative stress by dividing potential biomarkers into classes based on their molecular headgroups (Table S9). This allowed us to examine class-specific differences in the number of acyl carbon atoms, acyl carbon-to-carbon double bonds, and oxidation states (i.e., additional oxygen atoms) of component lipids under the two treatments. Differential expression of chemical properties within several classes (Figure 2; Table S9) suggested the

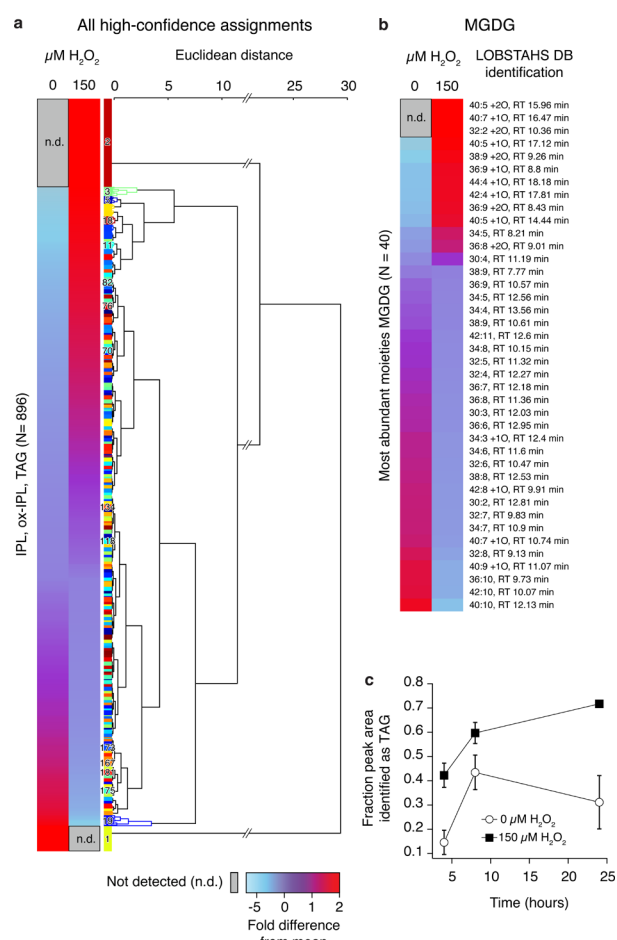


Figure 2. Remodeling of the *Phaeodactylum tricornutum* lipidome after 24 h, as visualized from data analyzed with LOBSTAHS. (a) Heatmap showing relative abundances across two treatments (0 and 150 μM H_2O_2) of all IPL, ox-IPL, and TAG identified with high confidence. Each row ($N = 896$) represents a different compound identified from the database; Figure S11 contains an expanded version of the plot that includes labels for each individual compound. (b) Heatmap detail, showing changes in the most abundant ($N = 40$) moieties of monogalactosyldiacylglycerol (MGDG), a lipid typically localized to the chloroplast. (c) Fraction of total peak area identified as triacylglycerol (TAG) at three time points during the experiment. Error bars are \pm SD of two replicates. In (a) and (b), shading shows the relative abundance of each compound as a fold difference of the mean peak area observed in that treatment from the mean peak area of the compound observed across all treatments. Dendrogram clustering and group definitions were determined by similarity profile analysis.⁶⁰ The numbers and identities of the components assigned to each group in (a) are given in Table S5 and Figure S11. Solid black lines in the dendrogram indicate branching that was statistically significant ($P \leq 0.01$).

P. tricornutum lipidome was remodeled in subtle but pervasive ways.

Fatty Acid Chain Elongation Is an Apparent Response to Oxidative Stress in the Chloroplast. Oxidative stress appeared to induce elongation of fatty acids throughout the *P. tricornutum* lipidome (Table S9). Lipid moieties upregulated by oxidative stress had longer fatty acid chains than those that were downregulated. We observed the greatest breadth of structural change in monogalactosyldiacylglycerol (MGDG), a lipid typically localized to the chloroplast (Table S9; Figure 2b). Moieties of MGDG upregulated in the 150 μM H_2O_2

treatment had significantly longer fatty acid chains and were more oxidized than those downregulated under oxidative stress; oxidation and elongation were also accompanied by a statistically significant decrease in acyl chain unsaturation (Table S9). The MGDG moieties responsible for these shifts in class structural properties were confined largely to groups 1, 2, 4, 5, 7, 9, 12, 166, 167, and 180 in our similarity profile analysis (Figures 2b and S11; Table S8). Lipid oxidation has been previously linked in the diatom *Skeletonema costatum* to lipolytic cleavage of MGDG and phospholipids within the chloroplast, resulting in oxylipin production from free fatty acids.⁶¹ While intact oxidized MGDG species have not been previously observed in algae, ox-MGDG have been documented in terrestrial plants upon wounding.⁸ The production of ox-IPL in *Arabidopsis thaliana* may be a means of binding ROS within the cell membrane to limit damage elsewhere.⁶²

Significant Enrichment Observed in TAG. Whereas the impact of oxidative stress within most lipid classes was confined to relatively modest changes in structural properties, treatment with 150 μM H_2O_2 induced a very significant enrichment in the fraction of peak area we identified as triacylglycerols (TAG; Figure 2c). Enigmatically, the TAG moieties upregulated in the 150 μM treatment were significantly less oxidized than those downregulated (Table S9). We hypothesize that the growth of this chemically reduced TAG pool may be evidence of enhanced de novo production of unoxidized TAG as a response to oxidative stress. Increased TAG synthesis is a known response to nutrient starvation in virtually all algae,^{63,64} including *P. tricornutum*.^{58,59} While increased TAG production in algae has not been previously linked directly to oxidative stress, increased production has been observed as a response to viral infection in the haptophyte alga *Emiliania huxleyi*.⁶⁵

CONCLUSIONS

Using a model data set, LOBSTAHS allowed us to identify differences in lipid speciation between treatments that could be used as potential indicators of oxidative stress in *P. tricornutum*. This potential extended beyond individual oxylipins and ox-IPL to sets of highly interrelated compounds that represented different stages of degradation and oxidation within the same lipidome (Figures 2 and S11; Table S8). While we demonstrated our approach using culture data from a single marine diatom, we designed LOBSTAHS so it can be used with exact-mass HPLC/MS data from virtually any experiment or natural system.

For a minority of features in the screened data set that had isobars and/or structural isomers (Table S7; Figures S2–S10), we were unable to make unambiguous identifications using LOBSTAHS alone. Many of these could be identified more rigorously through comparison with results from alternative commercial software or by manual inspection using authentic standards or diagnostic MS fragmentation patterns. Our objective, however, was not to definitively characterize a few individual compounds with absolute certainty but instead to putatively identify a broad range of possible biomarkers for further analysis and discovery. In doing so, LOBSTAHS achieves a level of certainty between Level 3 (“Putatively characterized compound classes”) and Level 2 (“Putatively annotated compounds”), using the scheme of Sumner et al.⁶⁶ The overall results we obtained with the *P. tricornutum* data set demonstrate the ability of LOBSTAHS to assist in this task. The screening and annotation process allowed us to assess a range of levels of confidence on the putative assignments,

producing an ample subset of high-confidence compound identifications sufficient to facilitate a detailed statistical analysis. Future integration of methods such as chiral HPLC⁶⁷ or matching of diagnostic ms² fragmentation spectra, with screening tools such as the one we present here, will assist in discovery of biomarkers in larger data sets and with even greater confidence.

■ ASSOCIATED CONTENT

● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b01260.

Methodological details, instructions for obtaining all software and data, supplementary discussion, supplementary data figures, supplementary data tables, and a supplementary chart showing examples of the types of isomers that can be identified with LOBSTAHS (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: james.r.collins@aya.yale.edu.

Present Address

[§]B.R.E.: Department of Oceanography, University of Hawai'i at Manoa, 1000 Pope Road, Honolulu, Hawaii, United States.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Elizabeth Kujawinski, Winn Johnson, and Krista Longnecker for discussions on the analysis of metabolomics datasets and assistance with xcms; Marian Carlson and Matthew Johnson for discussions on the effects of oxidative stress in microorganisms; Eugene Melamud for answering our many questions about MAVEN. Liz Kujawinski provided thoughtful feedback on an earlier version of the manuscript. Finally, we thank two anonymous reviewers for critical comments that improved the manuscript significantly. This research was supported by the Gordon and Betty Moore Foundation through Grant GBMF3301 to B.A.S.V.M. This research was also funded in part by a grant to B.A.S.V.M. from the Simons Foundation and is a contribution of the Simons Collaboration on Ocean Processes and Ecology (SCOPE). J.R.C. acknowledges support from a U.S. Environmental Protection Agency (EPA) STAR Graduate Fellowship (Fellowship Assistance Agreement No. FP-91744301-0).

■ REFERENCES

- (1) Apel, K.; Hirt, H. *Annu. Rev. Plant Biol.* **2004**, *55*, 373–399.
- (2) Crastes de Paulet, A.; Douste-Blazy, L.; Paoletti, R. *Free Radicals, Lipoproteins, and Membrane Lipids*; Plenum Publishing Corporation: New York, 1988.
- (3) Lesser, M. P. *Annu. Rev. Physiol.* **2006**, *68*, 253–278.
- (4) Strassburg, K.; Huijbrechts, A. M.; Kortekaas, K. A.; Lindeman, J. H.; Pedersen, T. L.; Dane, A.; Berger, R.; Brenkman, A.; Hankemeier, T.; van Duynhoven, J.; Kalkhoven, E.; Newman, J. W.; Vreeken, R. J. *Anal. Bioanal. Chem.* **2012**, *404*, 1413–1426.
- (5) Kuhn, H.; Borngraber, S. In *Lipoxygenases and Their Metabolites*; Nigam, S., Pace-Asciak, C. R., Eds.; Kluwer Academic: New York, 1999; pp 5–28.
- (6) Wenk, M. R. *Cell* **2010**, *143*, 888–895.
- (7) Buseman, C. M.; Tamura, P.; Sparks, A. A.; Baughman, E. J.; Maatta, S.; Zhao, J.; Roth, M. R.; Esch, S. W.; Shah, J.; Williams, T. D.; Weltri, R. *Plant Physiol.* **2006**, *142*, 28–39.
- (8) Vu, H. S.; Tamura, P.; Galeva, N. A.; Chaturvedi, R.; Roth, M. R.; Williams, T. D.; Wang, X.; Shah, J.; Weltri, R. *Plant Physiol.* **2012**, *158*, 324–339.
- (9) Andreou, A.; Brodhun, F.; Feussner, I. *Prog. Lipid Res.* **2009**, *48*, 148–170.
- (10) Thomas, A.; Patterson, N. H.; Marcinkiewicz, M. M.; Lazaris, A.; Metrakos, P.; Chaurand, P. *Anal. Chem.* **2013**, *85*, 2860–2866.
- (11) Haller, E.; Stübiger, G.; Lafitte, D.; Lindner, W.; Lämmerhofer, M. *Anal. Chem.* **2014**, *86*, 9954–9961.
- (12) Carini, P.; Van Mooy, B. A. S.; Thrash, J. C.; White, A.; Zhao, Y.; Campbell, E. O.; Fredricks, H. F.; Giovannoni, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7767–7772.
- (13) Van Mooy, B. A. S.; Fredricks, H. F.; Pedler, B. E.; Dyhrman, S. T.; Karl, D. M.; Koblizek, M.; Lomas, M. W.; Mincer, T. J.; Moore, L. R.; Moutin, T.; Rappe, M. S.; Webb, E. A. *Nature* **2009**, *458*, 69–72.
- (14) Fulton, J. M.; Fredricks, H. F.; Bidle, K. D.; Vardi, A.; Kendrick, B. J.; DiTullio, G. R.; Van Mooy, B. A. S. *Environ. Microbiol.* **2014**, *16*, 1137–1149.
- (15) Vardi, A.; Van Mooy, B. A. S.; Fredricks, H. F.; Popendorf, K. J.; Ossolinski, J. E.; Haramaty, L.; Bidle, K. D. *Science* **2009**, *326*, 861–865.
- (16) Ianora, A.; Miralto, A. *Ecotoxicology* **2010**, *19*, 493–511.
- (17) Vardi, A. *Commun. Integr. Biol.* **2008**, *1*, 134–136.
- (18) Pohnert, G. In *Algal Chemical Ecology*; Amsler, C. D., Ed.; Springer-Verlag: Berlin, 2008; pp 195–202.
- (19) Casotti, R.; Mazza, S.; Brunet, C.; Vantrepotte, V.; Ianora, A.; Miralto, A. *J. Phycol.* **2005**, *41*, 7–20.
- (20) Miralto, A.; Barone, G.; Romano, G.; Poulet, S. A.; Ianora, A.; Russo, G. L.; Buttino, I.; Mazzarella, G.; Laabir, M.; Cabrini, M.; Giacobbe, M. G. *Nature* **1999**, *402*, 173–176.
- (21) Balestra, C.; Alonso-Saez, L.; Gasol, J. M.; Casotti, R. *Aquat. Microb. Ecol.* **2011**, *63*, 123–131.
- (22) Ribalet, F.; Intertaglia, L.; Lebaron, P.; Casotti, R. *Aquat. Toxicol.* **2008**, *86*, 249–255.
- (23) Edwards, B. R.; Bidle, K. D.; Van Mooy, B. A. S. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 5909–5914.
- (24) Brügger, B. *Annu. Rev. Biochem.* **2014**, *83*, 79–98.
- (25) Holčapek, M. *Anal. Bioanal. Chem.* **2015**, *407*, 4971–4972.
- (26) Andreou, A.; Feussner, I. *Phytochemistry* **2009**, *70*, 1504–1510.
- (27) Lamari, N.; Ruggiero, M. V.; d'Ipollito, G.; Kooistra, W. H. C. F.; Fontana, A.; Montresor, M. *PLoS One* **2013**, *8*, e73281.
- (28) Girotti, A. W. *J. Lipid Res.* **1998**, *39*, 1529–1542.
- (29) Triantaphylides, C.; Krischke, M.; Hoeberichts, F. A.; Ksas, B.; Gresser, G.; Havaux, M.; Van Breusegem, F.; Mueller, M. J. *Plant Physiol.* **2008**, *148*, 960–968.
- (30) Sparvero, L. J.; Amoscato, A. A.; Kochanek, P. M.; Pitt, B. R.; Kagan, V. E.; Bayir, H. J. *Neurochem.* **2010**, *115*, 1322–1336.
- (31) Spickett, C. M.; Pitt, A. R. *Antioxid. Redox Signaling* **2015**, *22*, 1646–1666.
- (32) Balvers, M. G.; Verhoeckx, K. C.; Bijlsma, S.; Rubingh, C. M.; Meijerink, J.; Wortelboer, H. M.; Witkamp, R. F. *Metabolomics* **2012**, *8*, 1130–1147.
- (33) Bruins, M. J.; Dane, A. D.; Strassburg, K.; Vreeken, R. J.; Newman, J. W.; Salem, N., Jr.; Tyburczy, C.; Brenna, J. T. *J. Lipid Res.* **2013**, *54*, 1598–1607.
- (34) Ni, Z.; Milic, I.; Fedorova, M. *Anal. Bioanal. Chem.* **2015**, *407*, 5161–5173.
- (35) R Core Team. R Foundation for Statistical Computing: Vienna, Austria, 2015.
- (36) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
- (37) Tautenhahn, R.; Boettcher, C.; Neumann, S. *BMC Bioinf.* **2008**, *9*, 504.
- (38) Benton, H. P.; Want, E. J.; Ebbels, T. M. D. *Bioinformatics* **2010**, *26*, 2488–2489.

- (39) Kuhl, C.; Tautenhahn, R.; Bottcher, C.; Larson, T. R.; Neumann, S. *Anal. Chem.* **2012**, *84*, 283–289.
- (40) van Creveld, S. G.; Rosenwasser, S.; Schatz, D.; Koren, I.; Vardi, A. *ISME J.* **2015**, *9*, 385–395.
- (41) Husen, P.; Tarasov, K.; Katafiasz, M.; Sokol, E.; Vogt, J.; Baumgart, J.; Nitsch, R.; Ekroos, K.; Ejsing, C. S. *PLoS One* **2013**, *8*, e79736.
- (42) Pependorf, K. J.; Fredricks, H. F.; Van Mooy, B. A. S. *Lipids* **2013**, *48*, 185–195.
- (43) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534–2536.
- (44) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7*, 234.
- (45) Ejsing, C. S.; Duchoslav, E.; Sampaio, J.; Simons, K.; Bonner, R.; Thiele, C.; Ekroos, K.; Shevchenko, A. *Anal. Chem.* **2006**, *78*, 6202–6214.
- (46) Layre, E.; Sweet, L.; Hong, S.; Madigan, C. A.; Desjardins, D.; Young, D. C.; Cheng, T. Y.; Annand, J. W.; Kim, K.; Shamputa, I. C.; McConnell, M. J.; Debono, C. A.; Behar, S. M.; Minnaard, A. J.; Murray, M.; Barry, C. E., 3rd; Matsunaga, I.; Moody, D. B. *Chem. Biol.* **2011**, *18*, 1537–1549.
- (47) Clasquin, M. F.; Melamud, E.; Rabinowitz, J. D. *Curr. Protoc. Bioinformatics* **2012**, *37*, 14.11.1–14.11.23.
- (48) Pearson, A. In *Treatise on Geochemistry*; Holland, H. D., Turekian, K. K., Eds.; Elsevier: Oxford, 2014; pp 291–336.
- (49) Rosenwasser, S.; Graff van Creveld, S.; Schatz, D.; Malitsky, S.; Tzfadia, O.; Aharoni, A.; Levin, Y.; Gabashvili, A.; Feldmesser, E.; Vardi, A. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 2740–2745.
- (50) Dooley, C. T.; Dore, T. M.; Hanson, G. T.; Jackson, W. C.; Remington, S. J.; Tsien, R. Y. *J. Biol. Chem.* **2004**, *279*, 22284–22293.
- (51) Hanson, G. T.; Aggeler, R.; Oglesbee, D.; Cannon, M.; Capaldi, R. A.; Tsien, R. Y.; Remington, S. J. *J. Biol. Chem.* **2004**, *279*, 13044–13053.
- (52) Bligh, E. G.; Dyer, W. J. *Can. J. Biochem. Physiol.* **1959**, *37*, 911–917.
- (53) Hummel, J.; Segu, S.; Li, Y.; Irgang, S.; Jueppner, J.; Giavalisco, P. *Front. Plant Sci.* **2011**, *2*, 54.
- (54) Libiseller, G.; Dvorzak, M.; Kleb, U.; Gander, E.; Eisenberg, T.; Madeo, F.; Neumann, S.; Trausinger, G.; Sinner, F.; Pieber, T.; Magnes, C. *BMC Bioinf.* **2015**, *16*, 118.
- (55) Van Mooy, B. A. S.; Fredricks, H. F. *Geochim. Cosmochim. Acta* **2010**, *74*, 6499–6516.
- (56) Cajka, T.; Fiehn, O. *Anal. Chem.* **2016**, *88*, 524–545.
- (57) Melamud, E.; Vastag, L.; Rabinowitz, J. D. *Anal. Chem.* **2010**, *82*, 9818–9826.
- (58) Abida, H.; Dolch, L. J.; Mei, C.; Villanova, V.; Conte, M.; Block, M. A.; Finazzi, G.; Bastien, O.; Tirichine, L.; Bowler, C.; Rebeille, F.; Petroustos, D.; Jouhet, J.; Marechal, E. *Plant Physiol.* **2015**, *167*, 118–136.
- (59) Levitan, O.; Dinamarca, J.; Zelzion, E.; Lun, D. S.; Guerra, L. T.; Kim, M. K.; Kim, J.; Van Mooy, B. A. S.; Bhattacharya, D.; Falkowski, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 412–417.
- (60) Clarke, K. R.; Somerfield, P. J.; Gorley, R. N. *J. Exp. Mar. Biol. Ecol.* **2008**, *366*, 56–69.
- (61) d'Ippolito, G.; Tucci, S.; Cutignano, A.; Romano, G.; Cimino, G.; Miralto, A.; Fontana, A. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2004**, *1686*, 100–107.
- (62) Mene-Saffrane, L.; Dubugnon, L.; Chetelat, A.; Stolz, S.; Gouhier-Darimont, C.; Farmer, E. E. *J. Biol. Chem.* **2009**, *284*, 1702–1708.
- (63) Goncalves, E. C.; Wilkie, A. C.; Kirst, M.; Rathinasabapathi, B. *Plant Biotechnol. J.* **2015**.
- (64) Merchant, S. S.; Kropat, J.; Liu, B.; Shaw, J.; Warakanont, J. *Curr. Opin. Biotechnol.* **2012**, *23*, 352–363.
- (65) Hunter, J. E.; Frada, M. J.; Fredricks, H. F.; Vardi, A.; Van Mooy, B. A. S. *Front. Mar. Sci.* **2015**, *2*, 81.
- (66) Sumner, L.; Amberg, A.; Barrett, D.; Beale, M.; Beger, R.; Daykin, C.; et al. *Metabolomics* **2007**, *3*, 211–221.
- (67) Senger, T.; Wichard, T.; Kunze, S.; Gobel, C.; Lerchl, J.; Pohnert, G.; Feussner, I. *J. Biol. Chem.* **2005**, *280*, 7588–7596.