Production of Two Highly Abundant 2-Methyl-Branched Fatty Acids by Blooms of the Globally Significant Marine Cyanobacteria *Trichodesmium erythraeum*

Kelsey M. Gosselin, Robert K. Nelson, Amanda C. Spivak, Sean P. Sylva, Benjamin A.S. Van Mooy, Christoph Aeppli, Charles M. Sharpless, Gregory W. O'Neil, Eleanor C. Arrington, Christopher M. Reddy, and David L. Valentine*
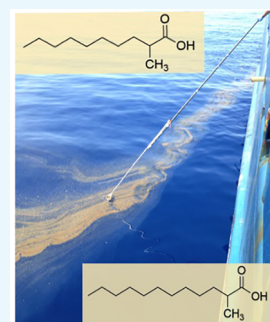
ACCESS | 📊 Metrics & More | 📄 Article Recommendations | 🆂🅸 Supporting Information

ABSTRACT: The bloom-forming cyanobacteria *Trichodesmium* contribute up to 30% to the total fixed nitrogen in the global oceans and thereby drive substantial productivity. On an expedition in the Gulf of Mexico, we observed and sampled surface slicks, some of which included dense blooms of *Trichodesmium erythraeum*. These bloom samples contained abundant and atypical free fatty acids, identified here as 2-methyldecanoic acid and 2-methyldodecanoic acid. The high abundance and unusual branching pattern of these compounds suggest that they may play a specific role in this globally important organism.

INTRODUCTION

The planktonic cyanobacterium *Trichodesmium* inhabits the tropical and subtropical Atlantic, Pacific, and Indian Oceans. This genus occurs throughout the upper, sunlit reaches of the ocean, as individual filaments or complex colonies, capable of diurnal nitrogen fixation in fully oxygenated seawater. *Trichodesmium* fixes an estimated 60 to 80 Tg of nitrogen (N) per year contributing up to 30% of the 100−200 Tg N fixed annually in the ocean.[1−5] This addition of bioavailable nitrogen from the atmosphere into the oceans underpins *Trichodesmium's* role as a major driver of productivity and thus the carbon cycle.[6] However, the contribution of *Trichodesmium* to total N fixation is not well constrained due to gaps in knowledge regarding its blooming behavior.[1,7] A clearer understanding of the fatty acids produced by *Trichodesmium* can provide insight into how they assemble their associated bacterial community or their mechanism for bloom formation. *Trichodesmium* blooms are infrequent but may contribute two times more N than nonblooms.[8] When these blooms occur, they seemingly suppress capillary waves and can create oil-like slicks at the sea surface that are visible to the naked eye.[9,10] Understanding the blooming behavior of *Trichodesmium* can help to better constrain their contribution to global N cycling.
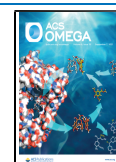
While sampling sea-surface oil sheens during a 2015 summer expedition to the oil-seep-prone Gulf of Mexico, we observed and investigated slicks that did not appear to be derived from oil. The visual similarities between hydrocarbon sheens and phytoplankton blooms often make these slicks difficult to distinguish by eyes without a closer inspection.[11] Slicks form at the sea surface in areas where compounds such as fatty acids or hydrocarbons dampen capillary waves. The observed slicks exhibited both dampening of capillary waves and also had the appearance of "sea sawdust" (Figure 1), a term frequently used to describe the aggregation of *Trichodesmium* colonies just beneath the sea surface.[1] We sampled several such *Trichodesmium* blooms by passing a precleaned Teflon screen through the sea surface until the screens showed a light brown color.[12] These samples were collected over the course of 8 days across multiple transects in the northern Gulf of Mexico, an area rich in natural oil seepage, in June 2015. The blooms often coincided with the location of oil sheens and many of our samples contained a mixture of both *Trichodesmium* colonies and oil-derived hydrocarbons, likely due to surface currents acting on both floating masses, leading to accumulation at the same location. Our present understanding is that the occurrence of *Trichodesmium* and oil sheens can be attributed to physical processes rather than a biogeochemical relationship. The presence and relative abundance of *Trichodesmium erythraeum* (*T. erythraeum*) (most closely related to strain
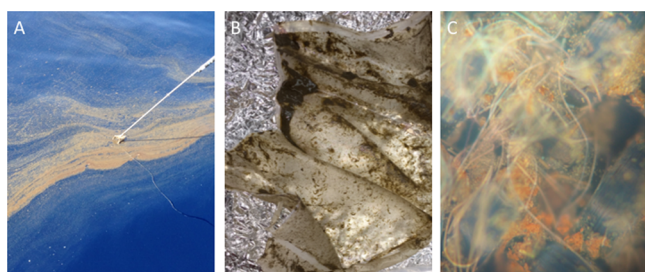
Figure 1. Images of *Trichodesmium*. (A) Bloom sample TN-10 collected from the deck of the R/V Atlantis on June 20, 2015, with a 10 cm × 10 cm Teflon screen (General Oceanics, Miami, FL), (B) material collected on the Teflon screen, and (C) colonies of individual *Trichodesmium* filaments scraped from the Teflon screen recorded using a photomicrograph (individual filaments are approximately 10 μm in diameter). Due to the sampling method, the colony structure was not well preserved in the samples.

Figure 2. GC-FID trace of the transesterified sample TN-10 with FAME carbon number along the *x*-axis. Peaks labeled * and ** are the two most abundant peaks in the sample confirmed as 2-methyl-C_{10} and 2-methyl-C_{12}. FAME carbon numbers were assigned from analysis of pure saturated FAME standards (Supelco 37 component FAME mix, CRM47885, Bellefonte, PA). Other unlabeled peaks are identified in Figure 4A.

IMS101) were confirmed by the 16S rRNA community composition on a subset of the samples using both previously published data and new analyses from our samples (Table S1).[11]

RESULTS AND DISCUSSION

Historical Results. In previous studies of *T. erythraeum* in the Gulf of Mexico, decanoic acid (C_{10}) and dodecanoic acid (C_{12}) were found to be, by far, the most abundant fatty acids representing up to 50% of the fatty acid pool. This is in contrast to other cyanobacteria that are typically dominant in the presence of hexadecanoic and octadecanoic acids (C_{16} and C_{18}).[13,14] In a 1967 study describing fatty acids in blue-green algae, Parker et al. sampled surface blooms of *T. erythraeum* near Port Aransas in the Gulf of Mexico. They found that one bloom contained 50% C_{10} fatty acid and noted that this pattern was striking because it was a relatively rare fatty acid.[13]

In 1997, when Carpenter et al. were searching for possible tracers of *Trichodesmium* in the Sargasso and Caribbean Seas, they expected to see a similar pattern in the fatty acids to what Parker et al.[13] found. Instead, they did not detect C_{10} and C_{12} in their field samples of *Trichodesmium thiebautii* (*T. thiebautii*) but did identify saturated C_{10} in their laboratory cultures of *T. erythraeum* IMS101, which were presumably more closely related to *T. erythraeum* that Parker et al.[13] sampled.[15]

Fatty Acid Composition of Teflon Screen Samples. Accordingly, we expected that gas chromatography (GC) analysis of the acid-catalyzed transesterified extracts of our Teflon screens containing *T. erythraeum* would include the fatty acid methyl esters (FAMEs) of C_{10} and C_{12}. However, when analyzed by one-dimensional GC with flame ionization detection (FID) and mass spectrometry (MS), the retention times and mass spectra of the two major peaks were offset and eluted later than the pure standards C_{10} FAME and C_{12} FAME (Figure 2). These now unknown compounds were present, respectively, at concentrations 2.1- and 1.2-fold higher than the sum of the three next most abundant fatty acids (C_{14}, C_{16}, and C_{18}) in a representative sample (Figure 2).

Subsequent application of comprehensive two-dimensional GC with time-of-flight mass spectrometry (GC × GC–ToF–MS) revealed the two major peaks to also be offset in both dimensions compared to the pure standards and they again had mismatched mass spectra. The spectra of the two major peaks in a representative field sample lacked the dominant *m/z* 74
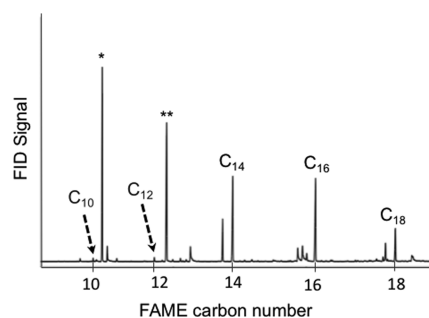
ion, typical of unbranched FAMEs, arising from a McLafferty rearrangement.[16] The mass spectra of the now unknown compounds instead showed an *m/z* 88 fragment, indicative of branching on the second carbon. A comparison of normal and 2-methyl-branched FAMEs along with their mass spectra fragmentation patterns is shown in Figure 3.

C_{10} FAME and C_{12} FAME were also identified, which provided additional confidence that the more abundant peaks resulted from unique compounds (Table 1).

Analysis of Unknown Compounds by GC−FID−MS and GC × GC−ToF−MS. The mass spectra of the unknown peaks, though not matched most closely by library searches, matched in the fragmentation patterns with the methyl esters of 2-methyldecanoate and 2-methyldodecanoate. The apparent match of mass spectra with the 2-methyl-branched compounds led us to synthesize these compounds to confirm the identity of the unknown peaks.

The 2-methylated methyl esters were synthesized by alkylating C_{10} FAME and C_{12} FAME according to the study by Dickschat et al.[16] and were confirmed to be 2-methyldecanoate and 2-methyldodecanoate by NMR spectroscopy. The compounds were also base-hydrolyzed to their corresponding 2-methyl carboxylic acids. We analyzed the synthetic 2-methyl-FAMEs following the same methods used for the field extracts by GC−FID−MS and GC × GC−ToF−MS (Figure 4). Both of these methods revealed the same retention times between the synthetic methyl-branched compounds and the abundant peaks in the transesterified field extracts. We also identified heptadecane ($C_{17}H_{36}$) in addition to our internal and external standards in the total ion chromatogram (TIC).[17]

Next, we compared the mass spectra of the major peaks in the field samples to the mass spectra of the synthetic 2-methyldecanoic acid and 2-methyldodecanoic acid. The mass spectra of the synthesized branched compounds matched those of the two unknown compounds in the field samples, confirming that the compounds in our samples were branched, not a straight chain.

Finally, we performed a spiking experiment by injecting synthetic 2-methyl FAMEs into the representative transesterified field extract at concentrations that would noticeably increase the peak size of our unknown peaks if they were indeed the same compounds. The two abundant peaks in the field extract increased in height and area but maintained their
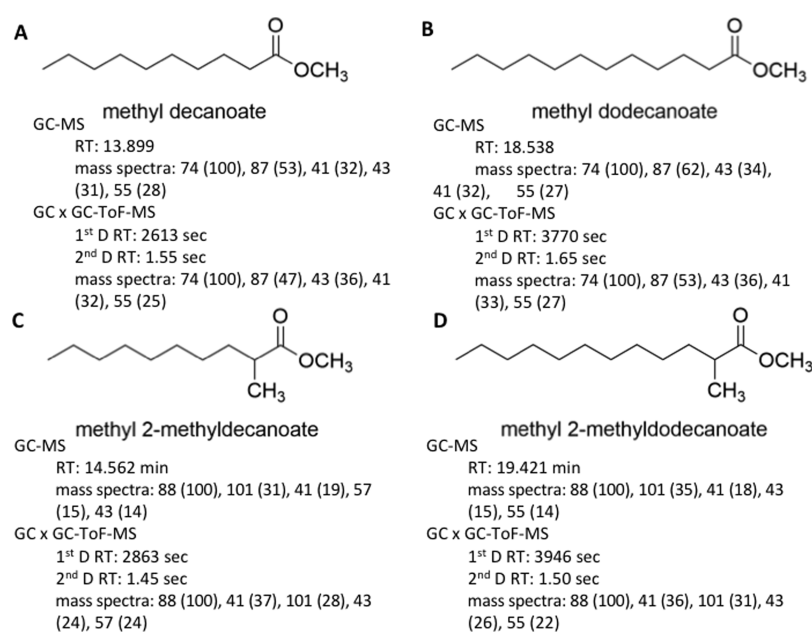
Figure 3. Chemical structure and mass spectra (GC−MS and GC × GC−ToF−MS) of (A) methyl decanoate (C_{10} FAME), (B) methyl dodecanoate (C_{12} FAME), (C) methyl 2-methyldecanoate (2-me-C_{10} FAME), and (D) methyl 2-methyldodecanoate (2-me-C_{12} FAME). The mass spectra are presented as the five most abundant ions with percent abundance in parenthesis.

peak shape following this addition (Figure 5). These changes indicate that the synthetic compounds align in retention time and structure to the compounds that were already in the extract, further confirming their identity as 2-methyldecanoate and 2-methyldodecanoate.

Distinguishing Ester-Bound versus Free Fatty Acids. Having concluded that the abundant fatty acids associated with *T. erythraeum* we sampled contained two relatively short and branched chain compounds, we sought to determine if they existed as free fatty acids (FFAs) or as an ester-bound form. We conducted a differential diagnosis with three analyses to identify the chemical state of the compounds in the field samples. Each treatment was analyzed by GC−FID−MS and GC × GC−ToF−MS (Figure 6).

A. Directly analyzed as a total lipid extract (TLE);
B. Transesterified in acidic methanol to convert any FFAs to FAMEs; and
C. Derivatized with *N,O*-bis(trimethylsilyl)-trifluoroacetamide and trimethylchlorosilane (BSTFA +1% TMCS) and pyridine to yield trimethylsilanes (TMS) derivatives of FFAs.

By comparing the retention time and mass spectra for the three treatments, we deduced the form of these compounds in nature. The TLE had two broad peaks that matched those of the FFA form of the synthetic standard. Relative to treatment A, the retention times of the peaks shifted with treatments B and C (transesterification and TMS derivatization, respectively), with the FAME eluting prior to the FFA and the TMS derivative after the FFA. The peaks for treatments B and C were much sharper, as expected for FAME and TMS derivatives. If the compounds existed in the samples as FAMEs, the retention times would not have shifted from treatment A to treatment B. Additionally, treatments A and C would have the same retention time because TMS would only derivatize FFAs and not FAMEs. These results indicate that some or all of the compounds are in the form of 2-methyldecanoic acid and 2-methyldodecanoic acid in the

field sample. We are presently unable to exclude the additional occurrence of these compounds bound to other biomolecules (e.g., lipids, glycolipids, or lipoproteins) as our methodology only accessed the free form of the fatty acids in the TLE. It is worth noting that Parker et al. base-saponified their extracts, which would have extracted both the free and bound fatty acids.[13] Further work is needed to determine if these compounds occur in the ester-bound form in addition to the FFA form.

Comparison of Results to Prior Findings. To our knowledge, these compounds were previously unreported in association with *Trichodesmium* or any other marine bacteria. Given the occurrence of these compounds in high concentrations in multiple field samples, we further investigated the potential for misidentification in previous studies based on chromatographic methods that identified straight-chain C_{10} and C_{12} in high abundance.

When analyzing the fatty acid composition of *Trichodesmium* in 1967, Parker et al were limited to the technology available at the time. They used retention times from a linear log plot for identification that did not include retention times for unusual compounds like 2-methyl-branched fatty acids. Additionally, they used packed diethylene glycol succinate and Apiezon L columns,[13] which may have made separation difficult, possibly contributing to misidentification. One of the few other studies that reported the fatty acid content of *Trichodesmium*[15] found only C_{12} in one of the five field samples. However, they did detect both C_{10} and C_{12} in their *T. erythraeum* laboratory cultures. This study references the work of Parker et al.[13] when discussing the presence of these compounds. If these compounds were the 2-methyl-branched counterparts, it could have been overlooked as the retention times would have been similar to those of the straight-chain compounds. Additionally, their presence was unanticipated based on previous research.

Although we cannot state with absolute certainty that the *T. erythraeum* sampled by Parker et al.[13] in 1967 and the *T.*

Table 1. Content (ng cm^{-2} of Teflon Net) of Compounds of Interest

sample name[a]	sample coordinates (decimal degrees)	C_{10}	2-methyl C_{10}	C_{12}	2-methyl C_{12}	C_{10} straight: branched	C_{12} straight: branched
control[b]							
061915TN-01	27.136, −90.451	5.1	5.9	6.0	6.4	0.86	0.94
062015TN-07	27.018, −90.581	3.1	11	2.4	2.9	0.28	0.83
062115TN-18	27.011, −90.381	1.8	4.8	1.4	1.4	0.37	1.00
no visible oil[c]							
062015TN-08	27.019, −90.580		31		21		
062015TN-09	27.019, −90.580	4.0	110	4.5	88	0.04	0.05
062015TN-10	27.020, −90.580	75	5400	58	3600	0.01	0.02
062515TN-21	27.323, −92.114	3.6	200	3.7	11	0.02	0.34
lightly oiled[d]							
062115TN-11	27.007, −90.276		9.6				
062115TN-12	27.003, −90.279	73	1800	64	1300	0.04	0.05
062115TN-13	27.003, −90.279	6.2	130	8.1	120	0.05	0.07
062315TN-20	26.984, −91.016	2.2	5.9	2.3		0.36	
062615TN-28	27.414, −91.823	37	670	40	690	0.06	0.06
062615TN-29	27.414, −91.823	7.0	29	4.9	13	0.24	0.38
062615TN-30	27.423, −91.833	3.7	64		15	0.06	
062615TN-31	27.423, −91.833	34	380	23	120	0.09	0.19
062615TN-32	27.418, −91.824	8.1	9.9	6.1	3.5	0.82	1.74
062615TN-33	27.418, −91.824	16	350	25	330	0.05	0.08
062715TN-35	27.015, −90.632	6.3	17		25	0.37	
062715TN-37	27.017, −90.629	16	120		39	0.13	
heavily oiled[e]							
061915TN-02	27.135, −90.453			4.7	3.0		1.57
061915TN-03	27.138, −90.457	5.3	24	12	12	0.22	1.00
061915TN-05	27.133, −90.452	3.4	4.2	6.5	2.5	0.81	2.60
061915TN-06	27.133, −90.452	9.3	4.2	21	6.4	2.21	3.28
062115TN-14	27.010, −90.818	5.3	15	4.8	5.8	0.36	0.83
062115TN-15	27.010, −90.818	12	230	22	190	0.05	0.12
062115TN-16	27.009, −90.814	6.9	16		17	0.43	
062115TN-17	27.009, −90.814	28	420	24	420	0.07	0.06
062615TN-22	27.395, −91.820	5.7	31	5.6	19	0.18	0.29
062615TN-23	27.395, −91.820	9.0	5.6		45	1.61	
062615TN-24	27.399, −91.822			4.9			
062615TN-25	27.399, −91.822			4.1			

[a]Sample name is in the format: date collected-sample material-sequence number. Sample names were assigned at the time of collection on the research vessel. [b]Samples are classified as: control samples collected in areas with no visible oil slick or *Trichodesmium* bloom material. [c]No visible oil—containing mostly biological materials. [d]Lightly oiled—containing some petroleum hydrocarbons and showing a light brown net color. [e]Heavily oiled—containing mostly petroleum hydrocarbons and showing a dark brown net color.

erythraeum we sampled in 2015 produce the same compounds, it is interesting to note the similarity in relative abundance, carbon chain length, and sampling location. However, it is also worth noting that we identified the straight chain analogues in our samples. Additional work is needed to determine if, for example, different ecotypes of *T. erythraeum* differentially produce straight chain versus branched compounds; if the 2-methyl-branched compounds are consistently found in association with *T. erythraeum*, it may suggest their misidentification in previous work.

Biological Functions of 2-Methyldecanoic and 2-Methyldodecanoic Acids. Given *Trichodesmium*'s cosmopolitan distribution and contribution to the nitrogen and carbon cycles, it is important to consider the role of these unusual fatty acids. It is likely that neither the straight-chain C_{10} and C_{12} nor their branched counterparts are conservative membrane components as noted by Carpenter et al.[15] regarding the former two compounds. They also identified the straight-chain C_{10} and C_{12} in their laboratory cultures of *T. erythraeum* IMS101 and not in their *T. thiebautii* field samples,

suggesting that perhaps only certain species of *Trichodesmium* produce these compounds or that their production only occurs under certain conditions.[15] A study of two strains of *Trichodesmium* (RS9602 and WH9601) in the Gulf of Aqaba[18] makes no mention of C_{10} or C_{12} fatty acids in their lipid analysis despite the use of C_{11} as an internal standard, indicating that these compounds were likely not present or perhaps overlooked in their analysis.

2-Methyldecanoic acid and 2-methyldodecanoic acid are atypical based on the position of the methyl group, likely requiring specialized biochemistry for their synthesis. These factors led us to investigate their potential biological roles. The available literature regarding the existence and roles of 2-methyl-branched short-chain fatty acids is sparse and limited to animal studies. Some studies have reported 2-methyldecanoic acid and 2-methyldodecanoic acid as "signaling compounds," but they are generally not the most abundant compound in the mixture. For example, Wood et al.[19] tentatively identified 2-methyldecanoic acid in the anal gland secretions of wolverines. The compound comprised 0.8−5% of the total ion current of
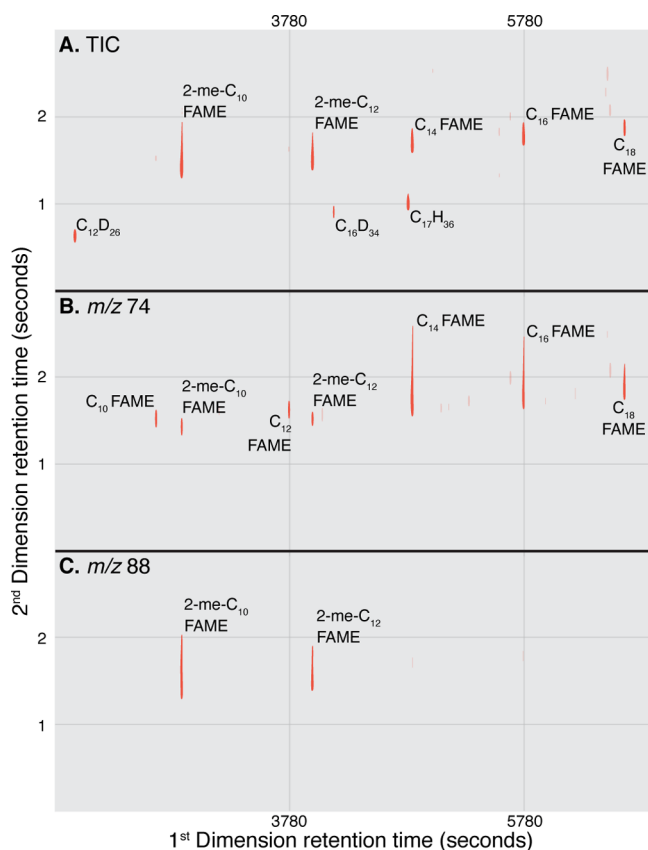
Figure 4. GC × GC−ToF−MS analysis of the transesterified field sample TN-10, (A) TIC within the C_{10}−C_{18} FAME window, (B) m/z 74 ion chromatogram, and (C) m/z 88 ion chromatogram. For reference, straight-chain FAME peaks were labeled based on the analysis of pure standards (Supelco 37 component FAME mix). In the TIC, the internal standard, dodecane-d_{26} ($C_{12}D_{26}$), and the external standard, hexadecane-d_{34} ($C_{16}D_{34}$), were also identified. The unknown compounds were identified based on the pure standards.

Figure 5. GC × GC−ToF−MS chromatograms of transesterified TN-10. (A) TN-10 extract, (B) TN-10 extract spiked with 2-methyl-C_{10} FAME, and (C) TN-10 extract spiked with 2-methyl-C_{12} FAME.

Figure 6. Sample TN-10 treated in three ways, (A) TLE, (B) acid-catalyzed TLE in methanol to produce FAME, and (C) TLE derivatized to produce TMS. The shift in retention time of our peaks of interest from their times in A to times in B and C indicates that in the TLE, the compounds exist as FFAs.

the volatile compounds in the anal gland secretion. These secretions were primarily used for "marking" or as a "fear/defense mechanism."[19] Additionally, the marking fluid (a mixture of urine and lipid fraction) collected from Bengal tigers contained a racemic mixture of 2-methyldecanoic acid and 2-methyldodecanoic acid.[20]

By analogy, our findings suggest the possible use of these fatty acids as a means of signaling for *Trichodesmium*. Fatty acid-based signaling could be especially useful for *Trichodesmium* during bloom formation because fatty acids are expected to form a slick at the sea surface, coinciding with the observed location of the blooming colonies. Effectively, the release of surface-active-signaling compounds would constrain the zone of signaling to the two dimensions of the ocean surface.

FFAs are also known potent energy uncouplers and could serve as a selective antibiotic.[21,22] 2-Methyldecanoic acid has been detected in the preen gland of the herring gull and mallard duck where it may function as an antibacterial agent that protects the feathers from feather-degrading bacteria.[23−25] Other studies on the preen gland have found that there is a symbiotic relationship between bacteria that produce the antibiotic substance in the preen gland and the host.[26] Based on these observations, these fatty acids might be synthesized and excreted by *Trichodesmium* for the purpose of structuring physically associated microbiota. Consistent with this
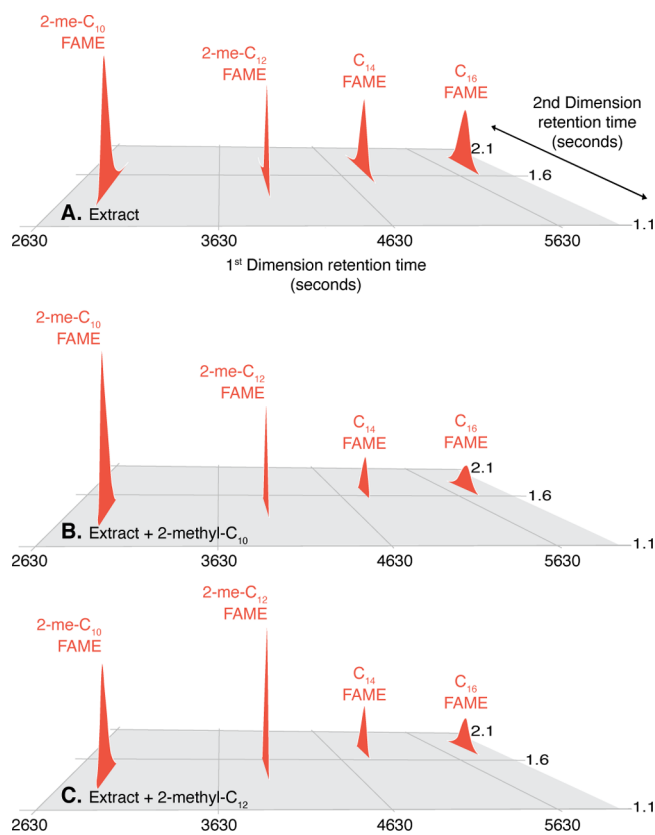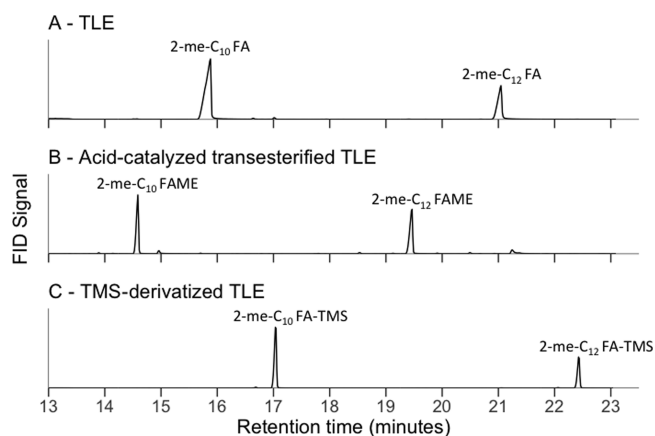
possibility, colonial forms of *Trichodesmium* are known to harbor a diverse assemblage of heterotrophic bacteria, perhaps associated with tolerance to or preference for FFAs.[27] Whatever the reason, *Trichodesmium* produces 2-methyldecanoic and 2-methyldodecanoic acid in high relative abundance, and it is likely not detrimental to growth. *Trichodesmium* growth is typically not limited by nitrogen or sunlight but by the availability of phosphorous.[28,29] As such, the synthesis of fatty acids can be achieved with carbon and energy in excess of the nutrient supply required for growth.

CONCLUSIONS

Using the methods described above, we identified two highly abundant 2-methyl-branched fatty acids, previously unreported in association with *Trichodesmium*, comprising up to 75% of the total fatty acid pool. Earlier studies report the presence of straight-chain analogues of 2-methyldecanoic and 2-methyl-dodecanoic acids but make no mention of these branched counterparts. In addition to identifying these compounds, we confirmed that some portion of them exists in nature as FFAs. Although the biological purpose(s) of these branched fatty acids is still uncertain, their atypical structures and high abundance in surface blooms indicate a possible role in cellular communication or interaction. Additional studies are needed to determine if these compounds are species-specific and to assess whether the straight-chain C_{10} and C_{12} are also found in high abundance as noted in a previous study.

EXPERIMENTAL SECTION

General Experimental Procedures. Initial compound identification was performed with an Agilent 7890 GC with an effluent split of 70:30 between a 5975C mass spectrometer and an FID. Peaks were separated on an Agilent DB-5ms column (60 m, 0.25 mm I.D., and 0.25 μm film). GC for all samples was performed on a Hewlett Packard 5890 Series II GC with an FID. GC × GC analysis was performed using a Leco Pegasus IV GC × GC−ToF system. The instrument was equipped with an Agilent 6890 GC and configured with a split/splitless autoinjector (7683B series) and a dual-stage cryogenic modulator (Leco, Saint Joseph, Michigan). The first-dimension column was a nonpolar Restek Rt-bDEX (30 m length, 0.25 mm I.D., and 0.25 μm df) and the second-dimension separations were performed using a 50% phenyl polysilphenylene-siloxane column (SGE BPX50, 1.0 m length, 0.10 mm I.D., and 0.1 μm df). Pure 2-methyldecanoic acid and 2-methyldodecanoic acid were prepared by alkylation of methyl decanoate and methyl dodecanoate followed by ester hydrolysis.

Field Samples. Samples were collected following the United States Coast Guard Marine Safety Laboratory protocol and other studies.[12,30] We used 10 cm diameter precleaned Teflon screens attached to a 3 m-long telescoping rod that we passed through the top few centimeters of the water surface. We sampled for approximately 30 min or until the screens showed a light brown color. The beginning and end locations of the sample were recorded. We collected 42 Teflon net samples from areas near oil seeps in the Gulf of Mexico (Table 1). Of the 42 nets, 4 were not oiled (purely biological material), 35 were oiled (combination of biological material and petroleum), and 3 were controls. The controls were collected using the same method as the samples in nonoiled water.

Extraction of Teflon Screens. Approximately, 3 cm^2 Teflon screens were spiked with undecanoic acid (C_{11}; internal standard), acidified with concentrated hydrochloric acid, and extracted three times with 4 mL of hexane. We refer to this extract as the TLE. The TLEs were acid-catalyzed trans-esterified and/or derivatized to produce trimethylsilane on any FFAs.

Transesterification was performed by adding 7 mL of 95:5 methanol/hydrochloric acid to the TLE after it was brought to complete dryness under N_2. The acidic methanol TLE was heated at 70 °C for 12 h to produce FAMEs. The FAMEs were extracted from the acidic methanol by adding hexane and water, vortexing, centrifuging, and removing the top (hexane) fraction to a new vial three times. The combined transesterified hexane extracts were dried over sodium sulfate and rotary-evaporated to 150 μL. Each extract was spiked with perdeuterated hexadecane to calculate the recovery of the internal standard and analyzed by GC with an FID.

To produce trimethyl silyl esters, the TLE was dried over sodium sulfate, rotary-evaporated to 150 μL, reacted with BSTFA +1% TMCS in pyridine, and heated at 70 °C for 30 min. Each extract was spiked with perdeuterated hexadecane to calculate the recovery of the internal standard and analyzed by GC with an FID.

Method blanks of only internal and external standards were characterized on precleaned Teflon screens and they contained no detectable analytes. The recovery of the internal standard ranged from 66 to 102%. The recoveries of C_{10} and C_{12} spiked onto precleaned Teflon screens and carried out using the method were 116 and 104%, respectively. A method detection limit following the protocol of Glaser et al. had a standard deviation of ±4.3 ng (low concentration) or ±29.9 ng (high concentration) for 2-methyl-C_{10} and ±5.9 ng (low concentration) or ±10.0 ng (high concentration) for 2-methyl-C_{12}, respectively.[31]

16S rRNA Gene Amplification and Analysis. To identify the cyanobacterial species present in our samples, we amplified and barcoded the V4 region of the 16S rRNA gene using previously described methods[32] with the 16Sf and 16Sr primers slightly modified.[33] Amplicon PCR reactions involved 1 μL of template DNA, 2 μL of a forward primer, 2 μL of a reverse primer, and 17 μL of an AccuPrime Pfx SuperMix. Thermocycling was carried out at 95 °C for 2 min; 30 cycles at 95 °C for 20 s, 55 °C for 15 s and 72 °C for 5 min; and a final elongation at 72 °C for 10 min. Sample DNA concentrations were normalized using the SequelPrep normalization kit, cleaned using the DNA Clean and Concentrator kit, visualized on an Agilent TapeStation, and quantified using a Qubit fluorometer. Samples were sequenced and demultiplexed at the University of California Davis Genome Center on an Illumina MiSeq platform with 250 nucleotide paired-end reads. A PCR-grade water sample was included in extraction, amplification, and sequencing as a negative control to monitor DNA contamination.

Trimmed fastq files were quality-filtered using the fastqPairedFilter command within the DADA2 R package v.1.14.1 (ref 34) with the following parameters: truncLen = c(160,160), maxN = 0, maxEE = c(2,2), truncQ = 2, rm.phix = TRUE, compress = TRUE, and multithread = TRUE. Quality-filtered reads were dereplicated using the derepFastq command. Paired dereplicated fastq files were joined using the mergePairs function with the default parameters. The makeSequenceTable command was used to construct a single-nucleotide variant table, and potential chimeras were removed de novo using removeBimeraDenovo. Taxonomic assignment of the sequences was done with the assignTaxonomy command using the Silva taxonomic training data set formatted for DADA2 v.138. If sequences were not assigned, they were left as not applicable.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c03196.

16S rRNA sequence abundance and equivalent taxonomy for the subset of samples (XLS)

■ AUTHOR INFORMATION

Corresponding Author

David L. Valentine − *Department of Earth Science and Marine Science Institute, University of California, Santa Barbara, California 93106, United States;* ● orcid.org/0000-0001-5914-9107; Email: valentine@ucsb.edu

Authors

Kelsey M. Gosselin − *Interdepartmental Graduate Program in Marine Science, University of California, Santa Barbara, California 93106, United States;* ● orcid.org/0000-0002-9327-8954

Robert K. Nelson − *Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, United States;* ● orcid.org/0000-0003-0534-5801

Amanda C. Spivak − *Department of Marine Sciences, University of Georgia, Athens, Georgia 30602, United States*

Sean P. Sylva − *Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, United States*

Benjamin A.S. Van Mooy − *Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, United States;* ● orcid.org/0000-0002-2804-6508

Christoph Aeppli − *Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine 04544, United States;* ● orcid.org/0000-0002-3758-0163

Charles M. Sharpless − *Andlinger Center for Energy and the Environment, Princeton University, Princeton, New Jersey 08544, United States*

Gregory W. O'Neil − *Department of Chemistry, Western Washington University, Bellingham, Washington 98225, United States;* ● orcid.org/0000-0002-1695-1906

Eleanor C. Arrington − *Interdepartmental Graduate Program in Marine Science, University of California, Santa Barbara, California 93106, United States*

Christopher M. Reddy − *Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, United States;* ● orcid.org/0000-0002-7814-2071

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.1c03196

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Capone, D.; Zehr, J.; Paerl, H.; Bergman, B.; Carpenter, E. *Trichodesmium*, a Globally Significant Marine Cyanobacterium. *Science* **1997**, *276*, 1221−1229.

(2) Walworth, N. G.; Fu, F.-X.; Lee, M. D.; Cai, X.; Saito, M. A.; Webb, E. A.; Hutchins, D. A. Nutrient-Colimited *Trichodesmium* as a Nitrogen Source or Sink in a Future Ocean. *Appl. Environ. Microbiol.* **2018**, *84*, e02137−e02117.

(3) Carpenter, E. J.; Capone, D. G.; Rueter, J. *Marine Pelagic Cyanobacteria : Trichodesmium and Other Diazotrophs*; NATO ASI series. Series C, Mathematical and physical sciences; 362; Kluwer Academic Publishers: Dordrecht; Boston: Norwell, MA, U.S.A., 1992.

(4) Zehr, J. P. Nitrogen Fixation by Marine Cyanobacteria. *Trends Microbiol.* **2011**, *19*, 162−173.

(5) Sohm, J. A.; Webb, E. A.; Capone, D. G. Emerging Patterns of Marine Nitrogen Fixation.(Report). *Nat. Rev. Microbiol.* **2011**, *9*, 499−508.

(6) Karl, D.; Michaels, A.; Bergman, B.; Capone, D.; Carpenter, E.; Letelier, R.; Lipschultz, F.; Paerl, H.; Sigman, D.; Stal, L. Dinitrogen Fixation in the World's Oceans. *Biogeochemistry* **2002**, *57*, 47−98.

(7) Mahaffey, C.; Michaels, A. F.; Capone, D. G. The Conundrum of Marine N_2 Fixation. *Am. J. Sci.* **2005**, *305*, 546−595.

(8) Westberry, T. K.; Siegel, D. A. Spatial and Temporal Distribution of *Trichodesmium* Blooms in the World's Oceans. *Global Biogeochem. Cycles* **2006**, *20*, 1−13.

(9) Luo, Y.-W.; Lima, I. D.; Karl, D. M.; Deutsch, C. A.; Doney, S. C. Data-Based Assessment of Environmental Controls on Global Marine Nitrogen Fixation. *Biogeosciences* **2014**, *11*, 691−708.

(10) Luo, Y.-W.; Doney, S. C.; Anderson, L. A.; Benavides, M.; Berman-Frank, I.; Bode, A.; Bonnet, S.; Boström, K. H.; Böttjer, D.; Capone, D. G.; Carpenter, E. J.; Chen, Y. L.; Church, M. J.; Dore, J. E.; Falcón, L. I.; Fernández, A.; Foster, R. A.; Furuya, K.; Gómez, F.; Gundersen, K.; Hynes, A. M.; Karl, D. M.; Kitajima, S.; Langlois, R. J.; LaRoche, J.; Letelier, R. M.; Marañón, E.; McGillicuddy, D. J., Jr.; Moisander, P. H.; Moore, C. M.; Mouriño-Carballido, B.; Mulholland, M. R.; Needoba, J. A.; Orcutt, K. M.; Poulton, A. J.; Rahav, E.; Raimbault, P.; Rees, A. P.; Riemann, L.; Shiozaki, T.; Subramaniam, A.; Tyrrell, T.; Turk-Kubo, K. A.; Varela, M.; Villareal, T. A.; Webb, E. A.; White, A. E.; Wu, J.; Zehr, J. P. Database of Diazotrophs in Global Ocean: Abundance, Biomass and Nitrogen Fixation Rates. *Earth Syst. Sci. Data* **2012**, *4*, 47−73.

(11) White, H. K.; Marx, C. T.; Valentine, D. L.; Sharpless, C.; Aeppli, C.; Gosselin, K. M.; Kivenson, V.; Liu, R. M.; Nelson, R. K.; Sylva, S. P.; Reddy, C. M. Examining Inputs of Biogenic and Oil-Derived Hydrocarbons in Surface Waters Following the Deepwater Horizon Oil Spill. *ACS Earth Space Chem.* **2019**, *3*, 1329−1337.

(12) Aeppli, C.; Reddy, C. M.; Nelson, R. K.; Kellermann, M. Y.; Valentine, D. L. Recurrent Oil Sheens at the Deepwater Horizon Disaster Site Fingerprinted with Synthetic Hydrocarbon Drilling Fluids. *Environ. Sci. Technol.* **2013**, *47*, 8211−8219.

(13) Parker, P. L.; Van Baalen, C.; Maurer, L. Fatty Acids in Eleven Species of Blue-Green Algae: Geochemical Significance. *Science* **1967**, *155*, 707.

(14) Holton, R. W.; Blecker, H. H.; Stevens, T. S. Fatty Acids in Blue-Green Algae: Possible Relation to Phylogenetic Position. *Science* **1968**, *160*, 545−547.

(15) Carpenter, E. J.; Harvey, H. R.; Fry, B.; Capone, D. G. Biogeochemical Tracers of the Marine Cyanobacterium *Trichodesmium*. *Deep. Sea. Res. Part 1. Oceanogr. Res. Pap.* **1997**, *44*, 27−38.

(16) Dickschat, J. S.; Bruns, H.; Riclea, R.; Flitsch, S. Novel Fatty Acid Methyl Esters from the Actinomycete *Micromonospora Aurantiaca. Beilstein J. Org. Chem.* **2011**, *7*, 1697−1712.

(17) Love, C. R.; Arrington, E. C.; Gosselin, K. M.; Reddy, C. M.; Van Mooy, B. A. S.; Nelson, R. K.; Valentine, D. L. Microbial Production and Consumption of Hydrocarbons in the Global Ocean. *Nat. Microbiol.* **2021**, *6*, 489.

(18) Post, A. F.; Dedej, Z.; Gottlieb, R.; Li, H.; Thomas, D. N.; el-Absawi, M.; el-Naggar, A.; el-Gharabawi, M.; Sommer, U. Spatial and Temporal Distribution of *Trichodesmium* Spp. in the Stratified Gulf of Aqaba, Red Sea. *Mar. Ecol. Prog. Ser.* **2002**, *239*, 241−250.

(19) Wood, W. F.; Terwilliger, M. N.; Copeland, J. P. Volatile Compounds from Anal Glands of the Wolverine, Gulo Gulo. *J. Chem. Ecol.* **2005**, *31*, 2111−2117.

(20) Burger, B. V.; Viviers, M. Z.; Bekker, J. P. I.; le Roux, M.; Fish, N.; Fourie, W. B.; Weibchen, G. Chemical Characterization of Territorial Marking Fluid of Male Bengal Tiger, Panthera Tigris. *J. Chem. Ecol.* **2008**, *34*, 659−671.

(21) Desbois, A. P.; Smith, V. J. Antibacterial Free Fatty Acids: Activities, Mechanisms of Action and Biotechnological Potential. *Appl. Microbiol. Biotechnol.* **2010**, *85*, 1629−1642.

(22) Desbois, A. P.; Lebl, T.; Yan, L.; Smith, V. J. Isolation and Structural Characterisation of Two Antibacterial Free Fatty Acids from the Marine Diatom. Phaeodactylum Tricornutum. *Appl. Microbiol. Biotechnol.* **2008**, *81*, 755−764.

(23) Levy, E. M.; Strain, P. M. The Composition of the Preen Gland Waxes of Some Marine Birds: A Word of Caution for Chemo-taxonomists. *Comp. Biochem. Physiol. B: Comp. Biochem.* **1982**, *72*, 255−260.

(24) Kolattukudy, P. E.; Bohnet, S.; Sasaki, G.; Rogers, L. Developmental Changes in the Expression of S-Acyl Fatty Acid Synthase Thioesterase Gene and Lipid Composition in the Uropygial Gland of Mallard Ducks (*Anas Platyrhynchos*). *Arch. Biochem. Biophys.* **1991**, *284*, 201−206.

(25) Shawkey, M. D.; Pillai, S. R.; Hill, G. E. Chemical Warfare? Effects of Uropygial Oil on Feather-degrading Bacteria. *J. Avian Biol.* **2003**, *34*, 345−349.

(26) Soler, J. J.; Martín-Vivaldi, M.; Ruiz-Rodríguez, M.; Valdivia, E.; Martín-Platero, A. M.; Martínez-Bueno, M.; Peralta-Sánchez, J. M.; Méndez, M. Symbiotic Association between Hoopoes and Antibiotic-producing Bacteria That Live in Their Uropygial Gland. *Funct. Ecol.* **2008**, *22*, 864−871.

(27) Frischkorn, K. R.; Rouco, M.; Van Mooy, B. A. S.; Dyhrman, S. T. Epibionts Dominate Metabolic Functional Potential of *Trichodesmium* Colonies from the Oligotrophic Ocean. *ISME J.* **2017**, *11*, 2090−2101.

(28) van Mooy, B. A. S.; Fredricks, H. F.; Pedler, B. E.; Dyhrman, S. T.; Karl, D. M.; Koblížek, M.; Lomas, M. W.; Mincer, T. J.; Moore, L. R.; Moutin, T.; Rappé, M. S.; Webb, E. A. Phytoplankton in the Ocean Use Non-Phosphorus Lipids in Response to Phosphorus Scarcity. *Nature* **2009**, *458*, 69−72.

(29) Sañudo-Wilhelmy, S. A.; Kustka, A. B.; Gobler, C. J.; Hutchins, D. A.; Yang, M.; Lwiza, K.; Burns, J.; Capone, D. G.; Raven, J. A.; Carpenter, E. J. Phosphorus Limitation of Nitrogen Fixation by *Trichodesmium* in the Central Atlantic Ocean. *Nature* **2001**, *411*, 66.

(30) *Oil Sample Handling and Transmittal Guide*. New London, CT, 2013.

(31) Glaser, J. A.; Foerst, D. L.; McKee, G. D.; Quave, S. A.; Budde, W. L. Trace Analyses for Wastewaters. *Environ. Sci. Technol.* **1981**, *15*, 1426−1435.

(32) Kozich, J. J.; Westcott, S. L.; Baxter, N. T.; Highlander, S. K.; Schloss, P. D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the Miseq Illumina Sequencing Platform. *Appl. Environ. Microbiol.* **2013**, *79*, 5112−5120.

(33) Caporaso, J. G.; Lauber, C. L.; Walters, W. A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S. M.; Betley, J.; Fraser, L.; Bauer, M.; Gormley, N.; Gilbert, J. A.; Smith, G.; Knight, R. Ultra-High-Throughput Microbial Community Analysis on the Illumina HiSeq and MiSeq Platforms. *ISME J.* **2012**, *6*, 1621−1624.

(34) Callahan, B. J.; McMurdie, P. J.; Rosen, M. J.; Han, A. W.; Johnson, A. J. A.; Holmes, S. P. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nat. Methods* **2016**, *13*, 581−583.