Examining inputs of biogenic and oil-derived hydrocarbons in surface waters following the *Deepwater Horizon* oil spill

Helen K. White[1], Charles T. Marx[1], David L. Valentine[2], Charles Sharpless[3], Christoph Aeppli[4], Kelsey M. Gosselin[5], Veronika Kivenson[6], Rachel M. Liu [7], Robert K. Nelson[5], Sean P. Sylva[5], Christopher M. Reddy[5]*

[1]Department of Chemistry, Haverford College, 370 Lancaster Avenue, Haverford PA, 19041, USA. [2]Department of Earth Science and Marine Science Institute, University of California, Santa Barbara, CA, 93106, USA. [3]Department of Chemistry, University of Mary Washington, Fredericksburg, Virginia, 22401, USA. [4]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, 04544, USA. [5]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA. [6]Interdepartmental Graduate Program in Marine Science, University of California, Santa Barbara, CA, 93106, USA. [7]College of Creative Studies, University of California, Santa Barbara, CA, 93106, USA.

*Corresponding author E-mail: hwhite@alum.mit.edu

ABSTRACT

Aliphatic hydrocarbons in the surface ocean are derived from phytoplankton and oil, but the coexistence and cycling of these two sources is not well defined. Moreover, phytoplankton and oil can create thin layers of a non-miscible phase that appear as sheens on surface water, and are visually difficult to distinguish. Here we examine the co-occurrence of hydrocarbon compounds in surface water samples to determine the inputs from phytoplankton and oil using pentadecane (C_{15}-*n*-alkane) and heptadecane (C_{17}-*n*-alkane) as molecular markers. Surface water sheens collected from a 2015 field survey in the Northern Gulf of Mexico contained hydrocarbons from natural oil seepage, phytoplankton blooms (i.e., biogenic) and mixtures of the two. Microbial communities examined in surface water sheen samples were dominated by cyanobacteria of the Genus *Trichodesmium*. The hydrocarbon content of the field-collected surface sheens was used to inform the categorization 2,171 samples collected in 2010 during the *Deepwater Horizon* (DWH) oil spill. Of the water samples categorized, a small fraction (<1%) contained only biogenic hydrocarbons, and ~10% contained a biogenic hydrocarbon input mixed with oil. This study provides a method for identifying biogenic inputs in oil slicks and surface sheens, and highlights a molecular approach to distinguish the two sources.

KEYWORDS Hydrocarbon, Pentadecane, Heptadecane, Phytoplankton, Oil, Seep, Sheen, Cyanobacteria

INTRODUCTION

Phytoplankton (including cyanobacteria and green algae) and oil are both sources of hydrocarbon compounds to the ocean, and in the vicinity of natural oil seeps, water samples have been found to contain hydrocarbons from both phytoplankton and oil[1]. Laboratory studies suggest that the production of aliphatic hydrocarbons by cyanobacteria in the surface ocean can produce an input of hydrocarbons ~100-fold greater than the combined inputs from oil spills and natural oil seeps[2,3]. The predominant hydrocarbons produced by cyanobacteria as well as a wide range of other phytoplankton in surface waters are pentadecane (C_{15}-*n*-alkane) and heptadecane (C_{17}-*n*-alkane)[2, 4-7]. Hydrocarbon production is widely distributed in cyanobacteria[8] where it is thought to influence membrane flexibility and curvature[9] and other properties such as permeability[10]. The relative abundance of C_{15} and C_{17} *n*-alkanes observed in surface waters is determined by their biogenic origin, but can also reflect physicochemical differences that influence their environmental fate[5].

It has been suggested that microbial processes involved in the degradation of aliphatic hydrocarbons from oil and phytoplankton are similar and that the short-term cycling of biogenic hydrocarbons can prime the degradation of more complex and diverse oil-derived hydrocarbons, which are cycled on longer time scales[2,3]. Depending on the physical state (part of recently deceased biological detritus, dissolved in oil phase, dissolved in water phase, dispersed into small droplets, sorbed on sediment), hydrocarbon compounds including C_{15}-*n*-alkane and C_{17}-*n*-alkane can persist in the environment for extended periods of time[11] or can be degraded within days[12]. However, there is no evidence that biogenic hydrocarbons accumulate in the surface ocean, challenging the connection between the microbial cycling of biogenic and oil-derived hydrocarbons[3]. The cycling of biogenic and oil-derived hydrocarbons in surface waters may be

dissimilar because of the different physical disposition of the two sources of hydrocarbons as well as the presence of other compounds with which they co-occur. For example, oil is a complex mixture of compounds with varying structures that are biodegraded to different extents[13]. Considering this, it is not clear if a short-term cycle that degrades biogenic hydrocarbons can realistically prime the ocean's microbiome to manage hydrocarbon influxes from spills and seeps.

To characterize the influence of biogenic hydrocarbon cycling on both chronic and acute releases of oil hydrocarbons to the oceans, a more comprehensive understanding of sources of hydrocarbons is needed[3]. The relative abundance of the different *n*-alkanes could be used to distinguish these two sources. C_{15} and C_{17}-*n*-alkane are, however, susceptible to air-sea gas exchange to different extents[5] and numerous studies show that there is rapid depletion of oil-derived *n*-alkanes $<C_{23}$ in surface waters due to this process[14]. The environmental partitioning of oil-derived hydrocarbons compared to biogenic hydrocarbons between the surface water, atmosphere, and cells, combined with varying relative rates of biogenic production of C_{15} and C_{17}-*n*-alkanes between different algae and phytoplankton, further confounds efforts to understand the relationship between the two sources and how they are cycled[1, 3].

Distinguishing these two sources of hydrocarbons is of interest because phytoplankton blooms and petroleum have been observed to co-occur following oil spills including the IXTOC-I oil spill in 1979[15] and the *Deepwater Horizon* (DWH) oil spill in 2010[16] as well as in association with natural oil seeps[17]. Following the DWH spill, phytoplankton communities were exposed to oil-derived carbon, which entered the planktonic food web[18, 19]. The co-occurrence of oil and plankton garnered further interest in relation to the DWH spill as phytoplankton and other particulate matter including marine snow and clay particles were implicated in transporting oil vertically from the surface ocean to the seafloor[20, 21]. Besides their co-occurrence, phytoplankton blooms can be

visually mistaken for dispersed oil and oil slicks (see Figure 1), which can in turn complicate oil spill response efforts[22, 23]. Exploring knowledge-gaps in the relationship between oil and phytoplankton is of interest as it can enhance understanding of oil's impact on phytoplankton, which in turn has implications for oil spill response efforts and the fate of oil in the marine environment.

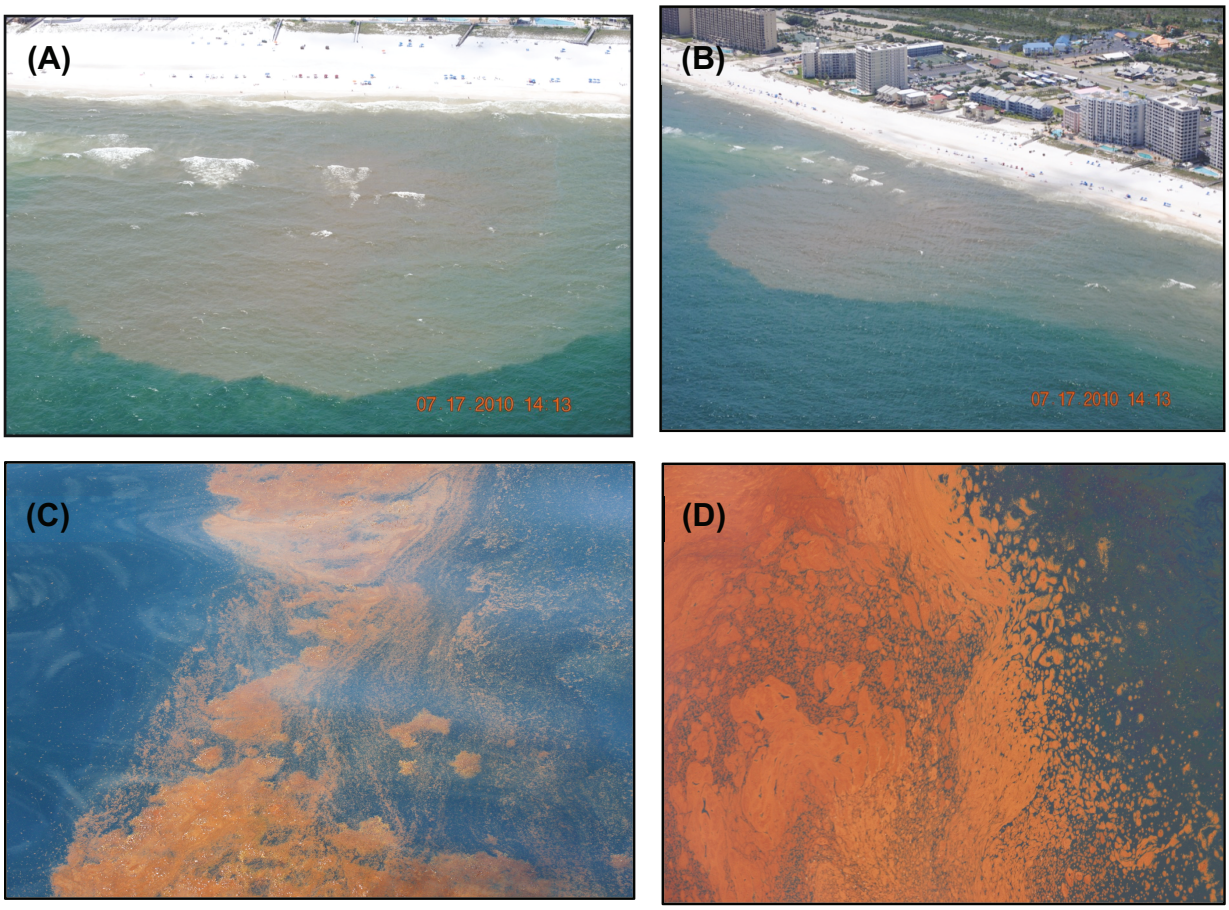

Figure 1. Example images of phytoplankton blooms (A, B) that could be mistaken for oil slicks taken from Gulf Shores, AL and Orange Beach, AL in July 2010, and a phytoplankton bloom (C) and oil emulsion (D) taken approximately three meters from the Ocean's surface during the Deepwater Horizon oil spill.

This study combines chemical, taxonomic and data-mining approaches to expand our understanding of biogenic hydrocarbon occurrence in the surface Ocean, and explore the concurrence of biogenic and oil-derived hydrocarbons in the Gulf of Mexico. In particular, this study aims to examine the utility of C_{15} and C_{17}-*n*-alkanes as molecular markers of biogenic inputs to distinguish oil from biogenic sheens, which appear as thin layers of a non-miscible phase resting on surface waters. Determining an appropriate molecular marker to apportion biogenic sheens from oil inputs is beneficial because even though phytoplankton have long been known to create sheens, challenges remain for deciphering sheens created by oil and phytoplankton as observed by overflights or satellite. A data set of 41 sheen samples collected from Gulf of Mexico surface waters in 2015 was analyzed via chemical methods to determine the presence of hydrocarbons from phytoplankton, oil, and mixtures of the two. Taxonomic analysis of the microbial community composition in sheen samples was performed to obtain information about the specific source of any biogenic hydrocarbons present. Based on characteristic patterns for petroleum-derived and biogenic oil, publicly available hydrocarbon data related to the *Deepwater Horizon* Natural Resource Damage Assessment (NRDA), collected between 2010-2013 was then analyzed to examine biogenic and oil-derived hydrocarbon inputs in the context of an accidental oil release. This study provides a synthesis of the presence and coexistence of biogenic (specifically cyanobacteria-produced) and oil-derived hydrocarbons in Gulf of Mexico waters, and the implications of these findings for oil spill response efforts and the fate of oil in the marine environment.

EXPERIMENTAL

Sample Collection, Extraction, and Analysis. Sheens were collected from Gulf of Mexico waters in June 2015 from the R/V *Atlantis* using Teflon screens (250-micron mesh size) according to the procedure previously described[24] (see Table S-1 for details of sample collection dates and locations and images shown in Figure S-1 of the Supporting Information (SI)). Briefly, the thin layer of non-miscible phase that appeared as a sheen on the surface of the water was sampled by passing the screen through the top few centimeters of surface water. Sampling continued until the screens were at least light-brown, which took approximately 10-20 min per sample. Via this method, free floating microbes would pass through the screen, but microbes attached to the sheen (oil or biogenic) would sorb to the net along with the sheen. Samples containing oil sheens were sub-sampled for microbial analysis by cutting a quarter of the Teflon screen using a sterilized blade and then placing immediately on ice in sterile containers, and then transferred for storage at -80°C prior to DNA extraction. The remainder of the Teflon screen and all other samples collected were placed in pre-combusted aluminum foil and kept frozen at -20°C prior to analysis. Two pure cultures of cyanobacteria were also examined including a Prochlorococcus culture grown in the lab, and a Trichodesmium bloom collected from the western tropical North Atlantic (15°48'N 45°48'W) using a Teflon screen on May 13, 2014 from the R/V *Atlantic Explorer*. For chemical analysis, pieces of the Teflon screen samples (4x4 cm^2) were extracted in 50-mL glass centrifuge vials with 10 mL dichloromethane/methanol (DCM/MeOH; 80/20 v/v) by vortexing for 2 minutes and collecting the organic phase. The procedure was repeated two times with 5 mL DCM/MeOH, and the combined organic phases (20 mL) were dried with 2g anhydrous Na_2SO_4. PAHs and *n*-alkanes were quantified using gas chromatography coupled to a mass spectrometer (GC/MS) in single ion monitoring mode as previously described[11]. The resulting quantities reported were normalized to the total extractable

mass (µg/g extractable material; measured gravimetrically). Procedural blanks comprised of Teflon screens analyzed alongside the samples did not contain any hydrocarbon compounds.

Microbial community analysis A small portion (approximately 0.5x2 cm) of the Teflon screen was aseptically cut and used for microbial analysis. The V4 region of the 16S rRNA gene was amplified using a previously-described method[25] with updated V4 primers[26] for sequencing on the Illumina MiSeq platform. Briefly, DNA was extracted from the Teflon screen using a MoBio PowerSoil kit with phenol chloroform. Extracts were purified and concentrated with ethanol precipitation. Amplicon PCR reactions contained 1μl template DNA (5ng/μl), 2μl forward primer, 2μl reverse primer, and 17μl AccuPrime™ Pfx SuperMix. Thermocycling consisted of 95˚C for 2min, 30 cycles of 95˚C for 20s, 55˚C for 15s, 72˚C for 5min, and a final elongation at 72˚C for 10min. Sample concentrations were normalized using SequelPrep Normalization Kit and visualized on an Agilent Tapestation (California NanoSystems Institute). Samples were sequenced at the UC Davis Genome Center on the Illumina MiSeq platform with 250nt, paired end reads. A PCR-grade water sample was included in extraction, amplification, and sequencing as negative control to assess for DNA contamination. Raw sequences were quality checked and analyzed using the open-source, mothur pipeline[25] and the SILVA bacterial reference database (Release 128)[27]. Phyloseq[28] was used to examine community diversity and operational taxonomic unit abundance. Sequencing analysis was performed on the Bridges high performance computing system[29, 30].

Processing of publicly available data. Additional water chemistry data used in this study were taken from the DIVER database, which is the NOAA repository hosting environmental data related to the *Deepwater Horizon* NRDA (https://www.diver.orr.noaa.gov/). The data consists of concentrations of hydrocarbons measured in water samples that may be present in the dissolved phase or as liquid oil (e.g. droplets or sheens). To reduce variability in the analytical methods

used to determine these concentrations, the data were limited to measurements made by Alpha Analytical (Mansfield, MA), a primary contract lab for the NRDA efforts. Samples were presumed to contain oil predominantly from the DWH spill due to the date (5/5/10 to 3/1/13), location (between latitude 24.25 to 30.74 N, and longitude: -97.28 to -80.53W), and intent of collection. It is possible that oil from natural seeps in the region may also be present in the water samples as natural seepage was active during this time in this region[1, 31]. Data with quality codes of "U" or "U, NSR", "J" or "J, NSR" and (defined in Table S-2) were removed from the analysis similar to previously described analysis[32]. A total of 10,700 multi-contaminant water samples contained measurements for C_6-C_{40} *n*-alkanes. A sub-set of 2,171 observations in which C_6-C_{40} *n*-alkanes were all detected were selected for downstream analysis without regard to whether concentrations surpassed reporting limits. Data analysis was performed in R (v 3.4.0) using the package tidyr. The data and code used in the analysis and additional details related to data filtering and quality control have been deposited in Figshare (DOI 10.6084/m9.figshare.7963586).

Data Analysis. The distribution and relative abundance of *n*-alkanes as well as oil biomarker compounds (pristane, phytane, $TPAH_{50}$ as previously defined[33], and hopane) with concentrations above the reporting limits were used to categorize samples from the DIVER database into six different categories: 1) oil, 2) oil and biogenic, 3) weathered oil, 4) weathered oil and biogenic, 5) biogenic, and 6) uncertain (see flowchart in Figure 2). For categorization, a continuous distribution of *n*-alkanes was considered to be indicative of the presence of oil. This continuous distribution of *n*-alkanes was defined as being the presence of *n*-alkanes anywhere in the range C_6-*n*-alkane to C_{40}-*n*-alkane where seven out of nine *n*-alkanes had concentrations above the reporting limit. Seven out of nine *n*-alkanes was chosen as a way to compensate for the potential

of one to two compounds in the chosen range being below reporting limits. Samples were categorized into the oil group if they contained a continuous distribution of *n*-alkanes with no carbon range preference (e.g. with a carbon preference index near 1, indicating no preference). Samples were categorized into the oil and biogenic group if they contained a continuous distribution of *n*-alkanes, and a biological overprint. Evidence of a biological overprint was determined by comparing the sum of concentrations of C_{15}-*n*-alkane and C_{17}-*n*-alkane to the sum of concentrations of C_{14}-*n*-alkane, C_{16}-*n*-alkane, and C_{18}-*n*-alkane. This value is 0.71 in the source oil from the Macondo Well[34] and ranges from 0.37-0.79 in oil that has been weathered yet still has detectable *n*-alkanes in this carbon range (values based on quantities previously determined[33]). For this study, a value of >1.5 was used to indicate the presence of a biogenic overprint evidenced by elevated C_{15}-*n*-alkane and C_{17}-*n*-alkane concentrations. Samples that did not contain a continuous distribution of *n*-alkanes were further categorized to see if they contained C_{15}-*n*-alkane and/or C_{17}-*n*-alkane and no other *n*-alkanes in the range C_6-*n*-alkane to C_{40}-*n*-alkane, and then if they contained other oil-derived compounds including pristane, phytane, PAHs ($TPAH_{50}$), and hopane. Samples that only contained C_{15}-*n*-alkane and/or C_{17}-*n*-alkane were categorized as biogenic if they did not contain other oil-derived compounds, and as weathered oil and biogenic if they did. Samples that did not contain C_{15}-*n*-alkane and/or C_{17}-*n*-alkane, but did contain other oil-derived compounds were categorized as weathered oil. These criteria prioritized categorization with a high degree of certainty over characterizing all samples present. All other samples that fell outside of the categorization were labeled as uncertain (Figure 2).

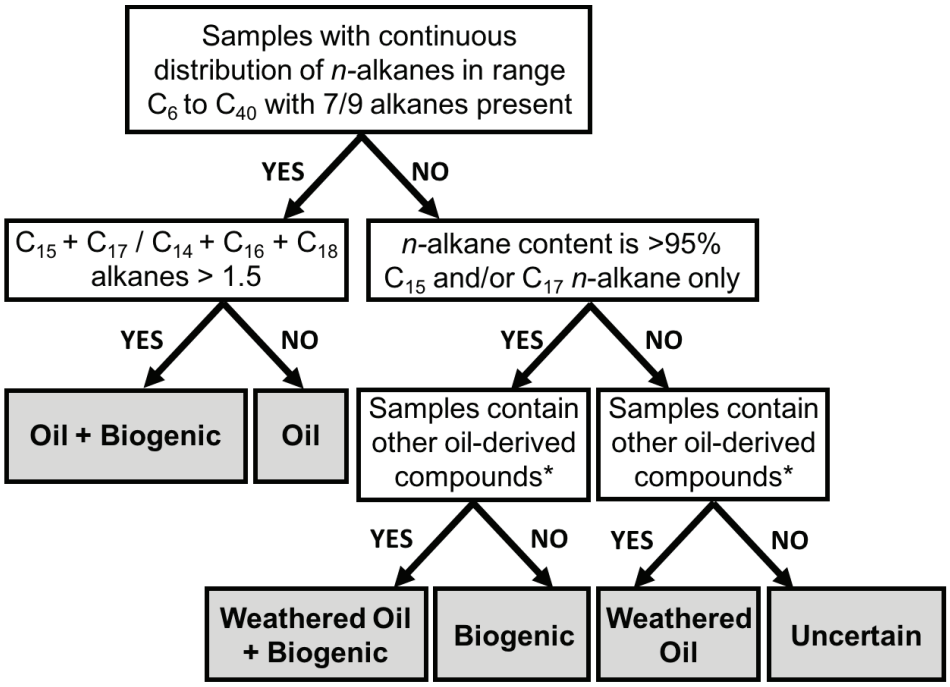

Figure 2. Flowchart for sample categorization. *Other oil-derived compounds include pristane, phytane, PAHs ($TPAH_{50}$), and hopane.

RESULTS AND DISCUSSION

Biogenic, oil-derived hydrocarbons and mixtures of the two were identified in surface water sheens. Surface sheens collected from Gulf of Mexico waters during the 2015 cruise could not be categorized visually from the bridge of the ship, but upon collection, were given preliminary categorizations into four groups: background, biogenic, oil, and mixed oil/biogenic (details provided in Table 1, example images shown in Figure S-2, and hydrocarbon distributions shown in Figure 3). Subsequent chemical analysis verified these categorizations. Background samples contained no hydrocarbons with the exception of two samples that contained C_{17}-*n*-alkane only. Biogenic samples were dominated by C_{17}-*n*-alkane and contained no isoprenoids or other oil-derived compounds such as pristane, phytane, hopane or PAHs. Oil samples and those containing mixtures of oil and biogenic inputs contained varying distributions of *n*-alkanes, the isoprenoids

pristane and phytane and PAH compounds (Tables 1, S-1, S-2 and S-3). For samples in these two groups there is considerable variability in the amount of C_{17}-*n*-alkane, which is apparent in the increased ratios of C_{17}-*n*-alkane to pristane and C_{17}-*n*-alkane/C_{18}-*n*-alkane in the mixed oil/biogenic samples compared to oil samples (Table 1). The larger range and values observed, indicates the presence of an additional and variable input of C_{17}-*n*-alkane from biogenic sources.

***Trichodesmium* can dominate the microbial community present in 2015 oil sheens.** For the oil sheen samples examined, *Trichodesmium* is the most abundant identified taxon in seven of nine samples except TN-23 and TN-25 (Table 2 and Figure S-3). Sheen TN-23 is dominated by *Candidatus* Actinomarina and sheen TN-25 is dominated by Rhodobacteraceae (Table 2). Since *Trichodesmium* is a known producer of biogenic hydrocarbons[6], this data suggests that *Trichodesmium* is likely a major source of biogenic hydrocarbons in Gulf of Mexico surface waters. Analysis of the pure cultures of *Trichodesmium* and *Prochlorococcus* indicates that they both produce C_{17}-*n*-alkane and *Prochlorococcus* also produces C_{15}-*n*-alkane (the latter was also observed previously[2]). The presence of Trichodesmium in oil sheens containing a mixture of oil-derived and biogenic hydrocarbons, verifies previous data indicating the coexistence of hydrocarbons from both phytoplankton and oil in the vicinity of natural oil seeps[1]. Following from this observation, we proceeded to determine whether hydrocarbons from phytoplankton coexisted in the context of samples collected from the DWH spill.

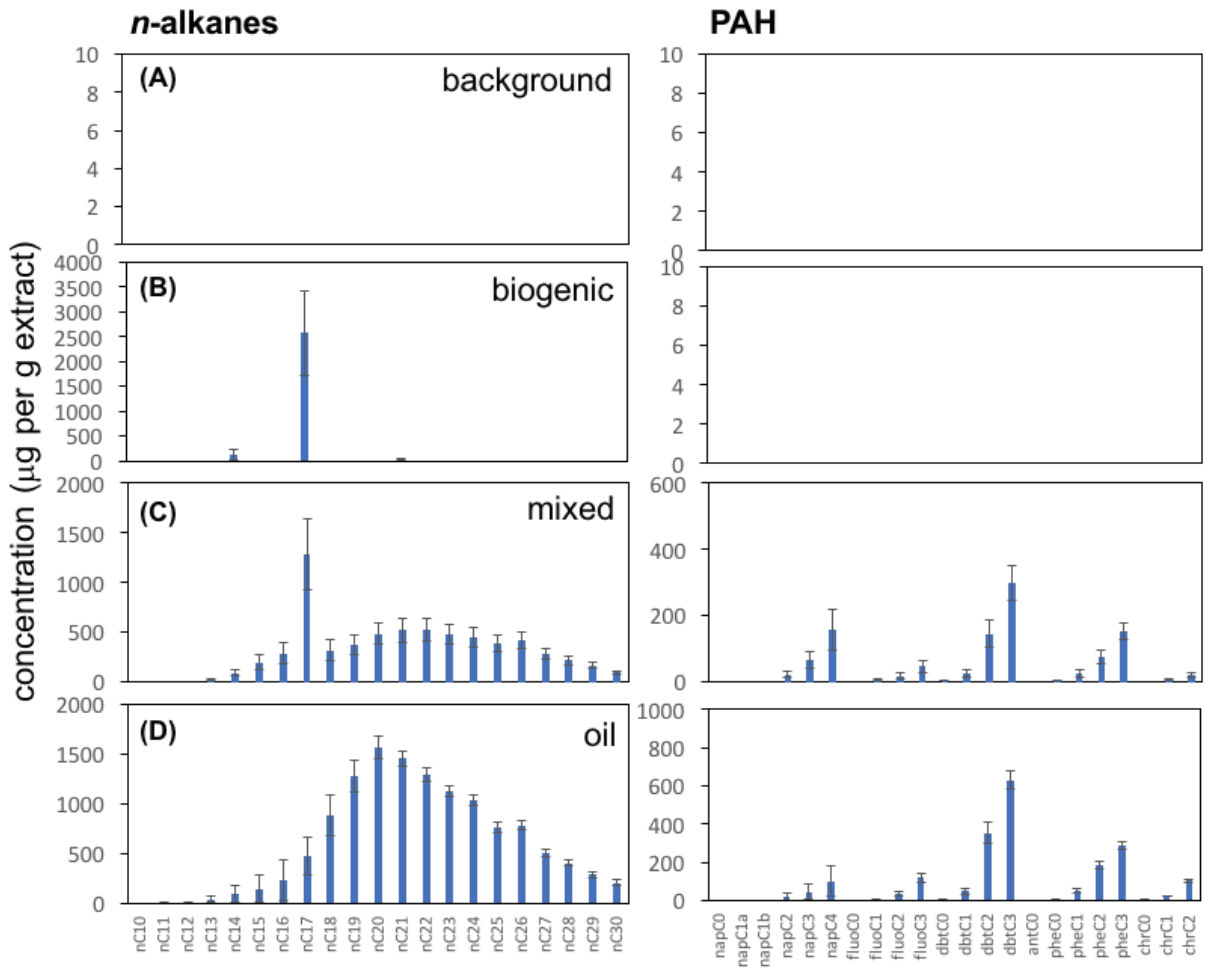

Figure 3. Mean hydrocarbon distributions present in the sheens from four sample types (A) background, (B) biogenic, (C) oil/biogenic mixture, (D) oil. Mean values are shown with error bars representing the standard error.

Table 1. Select ratios for hydrocarbon compounds present in surface sheen samples collected from Gulf or Mexico surface waters in June 2015 and grouped by sample type.[a]

Sample Type[b]	**Number of samples**	**C_{17}-*n*-alkane/ pristane**	**C_{18}-*n*-alkane/ phytane**	**C_{15}-*n*-alkane /C_{17}-*n*-alkane**	**C_{17}-*n*-alkane /C_{18}-*n*-alkane**
Background[c]	4	C_{17} only[g]	nd[h]	nd	C_{17} only
Biogenic[d]	5	C_{17} only	nd	nd-0.01 [na[i], na]	C_{17} only
Oil[e]	27	C_{17} only -12.78 [3.60, 2.59]	phytane only-1.38 [1.13, 0.20]	nd-0.66 [0.27, 0.21]	C_{17} only-4.97 [1.37, 1.24]
Oil / biogenic mixture[f]	5	C_{17} only-6.91 [6.32, 0.84]	nd-1.22 [1.13, 0.08]	nd-0.14 [na, na]	C_{17} only-4.57 [4.18, 0.41]

[a] Range of values measured is shown where applicable with mean and standard deviation in square brackets. These calculations exclude samples with values below the detection limits. Data for all samples including location and time collected is provided in the Supporting Information, Table S-1.
[b] Determined visually in the field when the samples were collected.
[c] Background samples were collected away from visible sheens. This group includes TN-01, TN-07, TN-11, and TN-18 samples and contained no other oil-derived compounds such as pristane, phytane, hopane or PAHs.
[d] Biogenic samples were collected when no visible oil or odor present. These samples are purely biological in appearance, contained no more than two other oil-derived compounds such as PAHs or hopane, and include TN-08, TN-09, TN-10, TN-20, TN-21. Sample TN-38 was also categorized as biogenic, but is not included in the above table as all compounds measured were below detection limits.
[e] Samples were collected from oil sheens and include TN-2, TN-3, TN-4, TN-5, TN-6, TN-14, TN-15, TN-16, TN-17, TN-19, TN-22, TN-23, TN-24, TN-25, TN-26, TN-27, TN-28, TN-29, TN30, TN-31, TN-32, TN-33, TN-34, TN-39, TN-40, TN-41, TN-42 and contain some PAHs detailed in the Supporting Information, Table S-3.
[f] Oil and biogenic samples were collected from dispersed sheens where oil was intermittently observable along with biological material including TN-12, TN-13, TN-35, TN-36, TN-37 and contain some PAHs detailed in the Supporting Information, Table S-4.
[g] C_{17} only = ratio cannot be determined because only C_{17} *n*-alkane has a concentration above the detection limit.
[h] nd = not detected noted when neither compound in the ratio is above the detection limit.
[i] na = not applicable due to sample size.

Table 2. Top three most abundant bacterial taxa present in oiled sheen samples collected from Gulf or Mexico surface waters in June 2015.

Oiled sheen	Phylum	Class	Order	Genus	Percent abundance
TN-22	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	10
	Proteobacteria	Alphaproteobacteria	Rhodobacterales	*Rhodobacteria*	6
	Proteobacteria	Alphaproteobacteria	SAR11 clade	Surface 1 unclassified	6
TN-23	Actinobacteria	Acidimicrobiia	Acidimicrobiales	*Actinomarina*	7
	Proteobacteria	Alphaproteobacteria	SAR11 clade	Surface 1 unclassified	7
	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Unclassified	6
TN-24	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	26
	Cyanobacteria	Cyanobacteria	Subsection I	*Synechococcus*	11
	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Unclassified	5
TN-25	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Unclassified	16
	Chloroflexi	Unclassified	Unclassified	Unclassified	11
	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	9
TN-26	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	44
	Chloroflexi	Unclassified	Unclassified	Unclassified	7
	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Unclassified	6
TN-27	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	67
	Proteobacteria	Alphaproteobacteria	Rhodobaterales	Unclassified	4
	Bacteroidetes	Cytophagia	Cytophagales	*Microschilla*	2
TN-28	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	62
	Bacteroidetes	Unclassified	Unclassified	Unclassified	4
	Chloroflexi	Unclassified	Cytophagales	Unclassified	3
TN-29	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	12
	Bacteroidetes	Flavobaceriia	Flabobacteriales	Unclassified	8
	Proteobacteria	Alphaproteobacteria	Rhodobaterales	*Leisingera*	7
TN-34	Cyanobacteria	Cyanobacteria	Subsection III	***Trichodesmium***	20
	Proteobacteria	Alphaproteobacteria	Rhodospiralles	AEGEAN-169 marine group	6
	Proteobacteria	Alphaproteobacteria	SAR11 clade	Surface 1 unclassified	5

Variable biogenic and oil hydrocarbon inputs measured in water samples collected in response to the *Deepwater Horizon* spill. Similar to the 2015 field data, different hydrocarbon distributions representing biogenic, oil, and mixed oil/biogenic inputs were observed in data from the water samples collected as part of the response to the DWH spill in 2010. A further

distinction was made to consider the presence of oil that had been extensively weathered not only via dissolution, evaporation and photo-oxidation, but also biodegradation, all of which contribute to the preferential removal of lower molecular weight oil compounds and *n*-alkanes. As such, this oil did not contain *n*-alkanes, but did contain other oil-derived compounds including pristane, phytane, PAHs, and hopane (outlined in Figure 2). This contributed two additional groups – 'weathered oil', and 'weathered oil + biogenic' – to the categorization of samples (Figure 2). Of the 2,171 water samples remaining after data reduction, 11 samples (0.5%) contained biogenic inputs only, 239 (11%) contained a mixture of biogenic and oil inputs and 1,628 (75%) contained oil or weathered oil (Table 3). The remaining 294 (13%) samples were uncertain and were not categorized. Water samples containing hydrocarbons solely from biogenic sources have the lowest concentration of total *n*-alkanes (mean and median = 0.8 ppm), compared to those containing oil (mean = 3.6 E03 ppm, median = 4.8 ppm; Table 3). PAH concentrations are higher in oil (mean = 5.4 E02 ppm, median = 3.5 E-01 ppm; Table 3) compared to weathered oil (mean = 3.1 E01 ppm, median = 3.7 E-02 ppm) samples. PAH concentrations are the lowest in samples containing biogenic and oil or weathered oil mixtures (Table 3). When present, a biogenic overprint is most evident in the C_{15}-*n*-alkane to C_{17}-*n*-alkane and the C_{17}-*n*-alkane to C_{18}-*n*-alkane ratios, which have higher values when compared to oil and weathered oil alone due to the elevated presence of biogenic C_{15} and C_{17}-*n*-alkanes (Table 3).

Table 3. Categorization and alkane concentrations and distribution of water samples examined from the DIVER database[a]. Range of values are shown with mean and median values in square brackets.

Sample Type	Number of samples	Total *n*-alkanes (ppm)	C_{17}-*n*-alkane /pristane[b]	C_{18}-*n*-alkane /phytane[b]	C_{15}-*n*-alkane /C_{17}-*n*-alkane[b]	C_{17}-*n*-alkane/C_{18}-*n*-alkane[b]	$TPAH_{50}$ (ppm)[c]
Biogenic	10	1.7 E-01 – 1.6 [7.6 E-01, 7.8 E-01]	na[d]	na	2.4 E-01 – 2.3 [7.8 E-01, 2.7 E-01]	na	na
Oil	225	8.0 E-01 – 6.1 E05 [3.6 E03, 4.8]	2.2 E-02 – 1.9 [1.1, 1.2]	7.0 E-02 – 4.6 [1.9, 1.9]	7.8 E-02 – 1.8 [8.1 E-01, 8.0 E-01]	4.3 E-01 – 4.7 E01 [9.1 E-01, 8.6 E-01]	nd[e] – 9.1 E04 [5.4 E02, 3.5 E-01]
Oil and biogenic mixture	55	6.3 E-01 – 4.5 E01 [7.1, 4.4]	2.2 E-01 – 9.4 E-01 [6.9 E-01, 8.0 E-01]	1.1 – 2.1 [1.4, 1.4]	2.0 E-01 – 4.6 [1.8, 1.4]	6.6 E-01 – 1.1 E01 [2.7, 1.0]	nd – 2.6 [3.7 E-01, 7.7 E-02]
Weathered oil and biogenic mixture	179	6.9 E-02 – 5.5 [1.2, 8.7 E-01]	na	na	1.8 E-01 – 7.5 [1.9, 1.5]	na	1.1 E-03 – 3.6 E-01 [5.5 E-02, 3.7 E-02]
Weathered oil	1408	nd – 3.5 E01 [9.9 E-01, 1.5 E-01]	na	1.4 E-01 – 4.6 [1.5, 1.1]	1.9 E-01 – 4.9 [1.2, 8.4 E-01]	1.9E-01 – 1.3 E01 [1.9, 3.4 E-01]	nd – 4.4E+03 [3.1 E01, 3.7 E-02]

[a] DIVER database is publicly available and serves as NOAA's repository for environmental data, including data related to the *Deepwater Horizon* NRDA, collected between 2010-2013. [b] Ranges, medians and means of ratios for each of the sample categories were calculated for samples only when both compounds were present in the sample. Samples without both compounds present are not included in any of the reported values shown. The number of samples excluded from the analysis for each sample type is summarized in Table S-5. [c] $TPAH_{50}$ is the sum of 50 oil-derived PAHs as described previously[33]. [d] na = not available because the compounds were not selected in the categorization or in the case of ratios, the compounds were not present in any of the samples for the calculation to be made. [e] nd = not detected and below reporting limits.

Water samples with biogenic inputs are spatially spread. Water samples analyzed from the DIVER database that contain biogenic inputs either with or without oil and weathered oil are found both along coastlines as well as offshore (Figure 4), and there is no clear spatial separation between samples of any particular category. The mixtures of biogenic and oil inputs observed could support previous findings that oil released from the DWH spill may have stimulated phytoplankton growth, in particular in the northeastern GOM and off of the Southwest Pass[35]. There is, however, no evidence that the biogenic component of these samples was alive at the time of collection, and it is not clear from this data how oil from the DWH spill impacted phytoplankton growth. Overall this data indicates that biogenic and oil mixtures are relatively widespread (identified in 11% of the investigated samples). In the absence of chemical analysis, and from a distance, delineating oil sheens from biogenic sheens, and mixtures of the two is visually and spatially challenging.

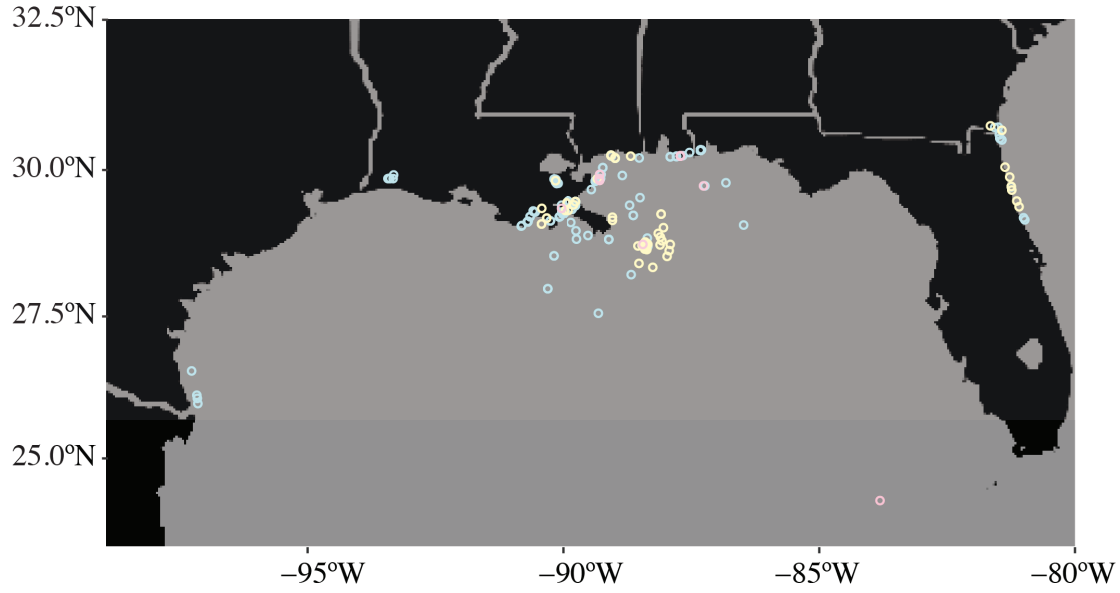

Figure 4. Water samples from the DIVER database containing biogenic inputs, categorized as biogenic (pink circles; n=10), oil and biogenic (yellow circles; n=55), weathered oil and biogenic (blue circles; n=179).

CONCLUSION

It is evident from the field samples and those in the DIVER database that inputs of biogenic and oil-derived hydrocarbons coexist in Gulf of Mexico surface waters. The utilization of publicly available datasets obtained as part of the DWH NRDA has proven to have significant scientific value[32, 36] serving as a valuable resource to compare field data collected by individual research groups. While unable to provide specific insight into the complex relationship between oil hydrocarbons and phytoplankton previously described[37] and summarized[38], this study does lend support for continued examination of this important topic. Previous lab and field studies indicate that both an increase and decrease in phytoplankton production can occur in the presence of oil[16], and more recently, connections between oil spills and red tides or harmful algal blooms (HABs) have been described[39], providing evidence that oil can impact plankton food webs. In this study the co-existence of phytoplankton, in particular *Trichodesmium*, implies potential for the export of oil-derived carbon from spills and seeps to the seafloor. Future studies could examine whether *Trichodesmium* death, which can rapidly terminate blooms, facilitate aggregation and expedite the vertical flux to depths[40, 41], could play a role in transporting oil to the seafloor as occurred in the cases of IXTOC and DWH.

This study highlights that some water samples collected in response to the DWH spill, and that are captured in the DIVER database, contain phytoplankton bloom-derived hydrocarbons and no oil biomarkers. Our estimates indicate that this quantity is low (<1% of all samples categorized), which is to be expected considering these samples were collected in response to an oil spill. However, visual inspection of surface sheens alone can be misleading and sheens suspected to be oil-derived could instead have a biogenic or mixed oil and biogenic origin. This study highlights the utility of examining surface sheens at a molecular-level, with a particular focus on C_{15}-*n*-alkane

and C_{17}-*n*-alkane, to provide a robust assessment for the contribution of biogenic hydrocarbons. This approach is particularly important in biologically productive regions of the ocean where phytoplankton blooms and associated biogenic sheens exist. Overall these findings contribute to an improved understanding of the presence of mixed sources of hydrocarbons in the surface ocean and recommend caution when using optical properties alone to define slicks and spilled oil.

Supporting Information. The Supporting Information (SI) contains Tables S-1, S-2, S-3, S-4 and S5, and Figures S-1, S-2 and S-3. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgments. This research was made possible by funding from a Gulf Research Program Early-Career Research Fellowship and a Henry Dreyfus Teacher-Scholar Award to HKW; NSF OCE-1333162 and OCE-1635562to DLV; NSF OCE-133148 and OCE-1634478 to CMR and CA; NSF DGE-1650114 to VK, and support from the Beckman foundation to CTM. This work used the Extreme Science and Engineering Discovery Environment (XSEDE; supported by NSF ACI-1548562) specifically, the Bridges system, supported by NSF ACI-1445606, at the Pittsburgh Supercomputing Center (PSC) through allocation TG-DEB170007. Computational resources were provided by Center for Scientific Computing at UCSB and NSF Grant CNS-0960316. RML would like to acknowledge Eleanor Arrington for assistance with microbial community analysis. The authors thank Edward Overton for the images provided in Figure 1 A and B.

The authors declare no competing financial interest.

References.

(1) Talukder, A. R.; Ross, A.; Crooke, E.; Stalvies, C.; Trefy, C.; Qi, X.; Fuentes, D.; Armand, S.; Revill, A. Natural hydrocarbon seepage on the continental slope to the east of Mississippi Canyon in the northern Gulf of Mexico. *Geochem., Geophys., Geosyst*. **2013**, *14,* 1940-1956. DOI: 10.1002/ggge.20130

(2) Lea-Smith, D. J.; Biller, S. J.; Davey, M. P.; Cotton, C. A.; Sepulveda, B. M.; Turchyn, A.V.; Scanlan, D. J.; Smith, A. G.; Chisholm, S. W.; Howe, C. J. Contribution of cyanobacterial alkane production to the ocean hydrocarbon cycle. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 13591–13596. DOI: 10.1073/pnas.1507274112

(3) Valentine, D. L.; Reddy, C. M. Latent hydrocarbons from cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 13434–13435. DOI: 10.1073/pnas.1518485112

(4) Blumer, M.; Guillard, R. R.; Chase, T. Hydrocarbons of marine phytoplankton. *Mar. Biol.* **1971**, *8*, 183-9. DOI: 10.1007/BF00355214

(5) Schwarzenbach, R. P.; Bromund, R. H.; Gschwend, P. M.; Zafirou, O. C. Volatile organic compounds in seawater. *Org. Geochem.* **1978**, *1,* 93-107. DOI:10.1016/0146-6380(78)90006-2

(6) Carpenter, E. J.; Harvey, H. R.; Fry, B.; Capone, D. G. Biogeochemical tracers of the marine cyanobacterium Trichodesmium. *Deep Sea Res., Part I*. **1997**, *44,* 27-38. DOI: 10.1016/S0967-0637(96)00091-X

(7) Martí, S.; Bayona, J. P.; Albaigés, J. A potential source of organic pollutants into the Northeasetern Atlantic: The outflow of the Mediterranean deep-lying waters through the Gibraltar Strait. *Environ. Sci. Technol.* **2001**, *35*, 2682-2689. DOI: 10.1021/es000258p

(8) Coates, R. C.; Podell, S.; Korobeynikov, A.; Lapidus, A.; Pevzner, P.; Sherman, D. H.; Allen, E. E.; Gerwick, L.; Gerwick, W. H. Characterization of cyanobacterial hydrocarbon

composition and distribution of biosynthetic pathways. *PloS one*. **2014**, *9*, e85140. DOI: 10.1371/journal.pone.0085140

(9) Lea-Smith, D. J.; Ortiz-Suarez, M. L.; Lenn, T.; Nürnberg, D. J.; Baers, L. L.; Davey, M. P.; Parolini, L.; Huber, R. G.; Cotton, C. A. R.; Mastroianni, G.; Bombelli, P.; Ungerer, P.; Stevens, T. J.; Smith, A. G.; Bond, P. J.; Mullineaux, C. W.; Howe, C. J. Hydrocarbons are essential for optimal cell size, division and growth of cyanobacteria. *Plant*. *Physiol*. **2016**, pp-01205. DOI: 10.1104/pp.16.01205

(10) Valentine, R. C.; Valentine, D. L. *Omega-3 fatty acids and the DHA principle,* CRC: Boca Raton, FL, 2009.

(11) Aeppli, C.; Swarthout, R.F.; O'Neil, G.W.; Katz, S.D.;, Nabi, D.; Ward, C.P.; Nelson, R.K.; Sharpless, C.M.; Reddy, C.M. How persistent and bioavailable are oxygenated Deepwater Horizon Oil transformation products? *Environ*. *Sci*. *Technol*. **2018**, *52,* 7250–7258. DOI: 10.1021/acs.est.8b01001

(12) Prince, R.C.; Butler, J.D.; Redman, A.D. The rate of crude oil biodegradation in the sea. *Environ*. *Sci*. *Technol*. **2017**, *51*, 1278–1284. DOI: 10.1021/acs.est.6b03207

(13) Aeppli, C.; Nelson, R. K.; Radović, J. R.; Carmichael, C. A.; Valentine, D. L.; Reddy, C. M. Recalcitrance and degradation of petroleum biomarkers upon abiotic and biotic natural weathering of Deepwater Horizon oil. *Environ*. *Sci*. *Technol*. **2014**, *48*, 6726-6734.

(14) MacDonald, I. R.; Leifer, I.; Sassen, R.; Stine, P.; Mitchell, R.; Guinasso Jr, N. Transfer of hydrocarbons from natural seeps to the water column and atmosphere. *Geofluids*. **2002**, 95-107. DOI: 10.1046/j.1468-8123.2002.0023.x

(15) Jernelöv, A.; Lindén, O. Ixtoc I: A case study of the World's largest oil spill. *Ambio*. **1981**, *10,* 299-306.

(16) Hu, C.; Weisberg, R. H.; Liu, Y.; Zheng, L.; Daly, K. L.; English, D. C.; Zhao, J.; Vargo, G. A. Did the northeastern Gulf of Mexico become greener after the Deepwater Horizon oil spill? *Geophys. Res. Lett.* **2011**, *38,* L09601. DOI: 10.1029/2011GL047184

(17) D'souza, N. A.; Subramaniam, A.; Juhl, A. R.; Hafez, M.; Chekalyuk, A.; Phan, S.; Yan, B.; MacDonald, I. R.; Weber, S. C.; Montoya, J. P. Elevated surface chlorophyll associated with natural oil seeps in the Gulf of Mexico. *Nat. Geosci.* **2016**, *9*, 215-218. DOI: 10.1038/ngeo2631

(18) Graham, W. M.; Condon, R. H.; Carmichael, R. H.; D'Ambra, I.; Patterson, H. K.; Linn, L. J.; Hernandez Jr, F.J. Oil carbon entered the coastal planktonic food web during the Deepwater Horizon oil spill. *Environ. Res. Lett.* **2010**, *5*, 045301. DOI: 10.1088/1748-9326/5/4/045301

(19) Chanton, J. P.; Cherrier, J.; Wilson, R. M.; Sarkodee-Adoo, J.; Bosman, S.; Mickle, A.; Graham, W. M. Radiocarbon evidence that carbon from the Deepwater Horizon spill entered the planktonic food web of the Gulf of Mexico. *Environ. Res. Lett.* **2012**, *7*, 045303. DOI: 10.1088/1748-9326/7/4/045303

(20) Passow, U.; Ziervogel, K.; Asper, V.; Diercks, A. Marine snow formation in the aftermath of the Deepwater Horizon oil spill in the Gulf of Mexico. *Environ. Res. Lett.* **2012**, *7,* 035301. DOI: 10.1088/1748-9326/7/3/035301

(21) Kinner, N. E.; Belden, L.; Kinner, P. Unexpected sink for Deepwater Horizon oil may influence future oil spill response. *EOS. Trans. Am. Geophys. Union.* **2014**, *95,* 176. DOI: 10.1002/2014EO210005

(22) National Oceanic and Atmospheric Administration (NOAA). *Dispersant application observer job aid*. NOAA Hazardous Materials Response Division, Seattle, WA, 1999.

(23) U.S. Coast Guard, National Oceanic and Atmospheric Administration, Environmental Protection Agency, Centers for Disease Control and Prevention and Minerals Management Service. *Special monitoring of applied response technologies (SMART)*. U.S. Coast Guard, Washington, D.C., 2006 [Online] Available at: https://response.restoration.noaa.gov/oil-and-chemical-spills/oil-spills/resources/smart.html [August, 2006].

(24) 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. DOI: 10.1021/es4024139

DOI: 10.1021/es500825q

(25) 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. DOI: 10.1128/AEM.01043-13

(26) 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. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal*. **2012**, *8,* 1621-1624. DOI: 10.1038/ismej.2012.8

(27) Quast, C.; Pruesse, E.; Yilmaz. P.; Gerken, J.; Schweer, T.; Yarza, P.; et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids. Res.* **2013**, *41*, D590–D596 DOI: 10.1093/nar/gks1219

(28) McMurdie, P. J.; Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, **2013**, *8*, p.e61217. DOI: 10.1371/journal.pone.0061217

(29) Nystrom, N. A., Levine, M. J., Roskies, R. Z., and Scott, J. R. 2015. Bridges: A uniquely flexible HPC resource for new communities and data analytics. In Proceedings of the 2015 Annual Conference on Extreme Science and Engineering Discovery Environment (St. Louis, MO, July 26-30, **2015**). XSEDE15. ACM, New York, NY, USA. http://dx.doi.org/10.1145/2792745.2792775.

(30) Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., Roskies, R., Scott, J.R.; Wilkens-Diehr, N. **2014**, XSEDE: Accelerating scientific discovery. Computing in Science & Engineering. 16(5):62-74. http://doi.ieeecomputersociety.org/10.1109/MCSE.2014.80

(31) Stout, S.A. Oil spill fingerprinting method for oily matrices used in the Deepwater Horizon NRDA. *Environ. Forensics*. **2016***, 17*, 218-243. DOI: 10.1080/15275922.2016.1177759

(32) Bagby, S. C.; Reddy, C. M.; Aeppli, C.; Fisher, G. B.; Valentine, D. L. Persistence and biodegradation of oil at the ocean floor following Deepwater Horizon. *Proc. Natl. Acad. Sci. U. S. A.* **2017**,*114*, E9-18. DOI: 10.1073/pnas.1610110114

(33) Stout, S. A.; Payne, J. R.; Emsbo-Mattingly, S. D.; Baker, G. Weathering of field-collected floating and stranded Macondo oils during and shortly after the Deepwater Horizon oil spill. *Mar. Pollut. Bull.* **2016***b*, *105,* 7–22. DOI: 10.1016/j.marpolbul.2016.02.044

(34) Reddy, C. M.; Arey, J. S.; Seewald, J. S.; Sylva, S. P.; Lemkau, K. L.; Nelson, R. K.; Carmichael, C. A.; McIntyre, C. P.; Fenwick, J.; Ventura, G. T.; et al. Composition and fate of gas and oil released to the water column during the Deepwater Horizon oil spill. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20229-20234. DOI: 10.1073/pnas.1101242108

(35) Ozhan, K.; Parsons, M. L.; Bargu, S. How were phytoplankton affected by the Deepwater Horizon Oil Spill? *BioScience*, **2014**, *64*, 829-836. DOI: 10.1093/biosci/biu117

(36) Valentine, D.L.; Fisher, G.B.; Bagby, S.C.; Nelson, R.K.; Reddy, C.M.; Sylva, S.P.; Woo, M.A. Fallout plume of submerged oil from Deepwater Horizon. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15906–15911. DOI: 10.1073/pnas.1414873111

(37) Teal, J. M.; Howarth, R. W. Oil spill studies: A review of ecological effects. *Environ. Manage.* **1984**, *8,* 27-43. DOI: 10.1007/BF01867871

(38) Ozhan, K.; Parsons, M.L.; Bargu, S. How were phytoplankton affected by the Deepwater Horizon oil spill? *BioScience*, **2014**, *64*, 829-836. DOI: 10.1093/biosci/biu117

(39) Almeda, R.; Cosgrove, S.; Buskey, E. J. Oil spills and dispersants can cause the initiation of potentially harmful dinoflagellate blooms ("Red Tides"). *Environ. Sci. Technol.* **2018**, *52*, 5718-5724. DOI: 10.1021/acs.est.8b00335

(40) Bar-Zeev, E.; Avishay, I.; Bidle, K. D.; Berman-Frank, I. Programmed cell death in the marine cyanobacterium Trichodesmium mediates carbon and nitrogen export. *The ISME journal*. **2013**, *7,* 2340-2348. DOI: 10.1038/ismej.2013.121

(41) Spungin, D.; Pfreundt, I.; Berthelot, H.; Bonnet, S.; AlRoumi, D.; Natale, F.; Hess, W. R.; Bidle, K. D.; Berman-Frank, I. Mechanisms of Trichodesmium demise within the New Caledonian lagoon during the VAHINE mesocosm experiment. *Biogeosciences*. **2016**, *13,* 4187-4203. DOI: 10.5194/bg-13-4187-2016

For TOC only