Discrete Organic Phosphorus Signatures are Evident in Pollutant Sources within a Lake Erie Tributary

Brooker, MR[1,2]*; Longnecker, K[3]; Kujawinski, EB[3]; Evert, MH[1,2]; and Mouser, PJ[1,4].

[1]Department of Civil, Environmental, and Geodetic Engineering, Ohio State University, Columbus, Ohio 43210, United States

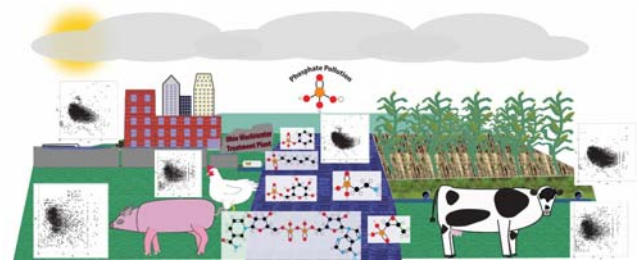
[2]Environmental Science Graduate Program, Ohio State University, Columbus, Ohio 43210, United States

[3]Woods Hole Oceanographic Institution, Department of Marine Chemistry and Geochemistry, Woods Hole, MA 02543, USA

[4]Department of Civil and Environmental Engineering, University of New Hampshire, Durham, New Hampshire 03824, USA

Keywords: phosphorus, Great Lakes, agriculture, mass spectrometry, harmful algal blooms

TOC ART



ABSTRACT

Phosphorus loads are strongly associated with the severity of harmful algal blooms in Lake Erie, a Great Lake situated between the United States and Canada. Inorganic and total phosphorus measurements have historically been used to estimate nonpoint and point source contributions, from contributing watersheds with organic phosphorus often neglected. Here, we used ultrahigh resolution mass spectrometry to characterize the dissolved organic matter and specifically dissolved organic phosphorus composition of several nutrient pollutant source materials and aqueous samples in a Lake Erie tributary. We detected between 23-313 organic phosphorus formulae across our samples, with manure samples having greater abundance of phosphorus- and nitrogen containing compounds compared to other samples. Manures also were enriched in lipids and protein-like compounds. The greatest similarities were observed between the Sandusky River and wastewater treatment plant effluent (WWTP), or the Sandusky River and agricultural edge of field samples. These sample pairs shared 84% of organic compounds and 59 to 73% of P-containing organic compounds, respectively. This similarity suggests that agricultural and/or WWTP sources dominate the supply of organic phosphorus compounds to the river. We identify formulae shared between the river and pollutant sources that could serve as possible markers of source contamination in the tributary.

INTRODUCTION

Freshwater lakes, such as the Great Lakes in North America, provide numerous economic opportunities to shoreline communities in the form of tourism, recreation, fisheries, manufacturing, and the transportation of goods across local and international boundaries. Lake Erie is one of five Great Lakes in the United States and Canada, and as with many freshwater resources worldwide, it has experienced recurrent harmful algal blooms that are believed to be propagated from anthropogenically-sourced nutrients from within its drainage basins[1, 2]. Primary productivity in freshwater systems is most often limited by phosphorus or nitrogen[1, 2], therefore changes in the abundance and form of these nutrients from upstream sources can have a profound effect on the downstream ecosystem. Since the early 2000s, increased nutrient loads have led to recurrent toxic cyanobacterial blooms along the southern coastline of the western Lake Erie basin, while hypoxia has developed in the hypolimnion of the central basin in the lake[1, 3, 4].

The magnitude of Lake Erie algae blooms in a given year is most strongly correlated to spring (May-June) phosphorus loads from its tributaries[5], with small blooms sometimes inflicting severe damage to the ecosystem. For example, although it was smaller in size compared to years past, the *Microcystis* bloom at the Toledo water treatment facility intake pipe in 2014 had a major impact on the shoreline community[4]. Microcystin concentrations in the treated water were twice as high as state guidelines (currently 1.6 μg/l in Ohio), causing the shutdown of the drinking water treatment plant serving over 400,000 Toledo residents and resulting in $65 million in economic damages to property values, tourism, recreation, and emergency water handling[4]. The impact of this and other phytoplankton blooms on the local economy has served

as a call-to-action for Ohio legislators to improve our understanding of nutrient pollutants contributing to this problem and develop best management practices to minimize discharge.

Phosphorus pollution in drainage basins is derived from both point (e.g. municipal/industrial wastewater effluents or combined sewer overflows) and nonpoint sources (sewage leaks, urban area runoff, or agricultural runoff/tile drainage)[6, 7], making individual pollutant sources difficult to isolate and manage. An extensive sampling network has been established in select Lake Erie tributaries to monitor loads to the lake[7], with an emphasis on reactive and total phosphorus. Reactive phosphorus, a term used interchangeably with orthophosphate (PO_4^{3-}), is readily assimilated by algae and simple to measure[7, 8]. Total phosphorus has been useful in forecasting harmful algal bloom severity[5]. Models have helped fill in the gaps between discrete sampling locations by considering local land usage to estimate spatial contributions[3, 7]. However, despite efforts made toward monitoring and modeling source contributions to Lake Erie, distinguishing between specific pollutant sources in order to mitigate the most impactful loads to the lake has proven difficult.

To gain further insight into pollutant sources and phosphorus pool dynamics, researchers are applying new mass spectrometry tools. One method, analysis of oxygen isotopic fractionation, has allowed for partial source tracking of phosphate entering Lake Erie from its tributaries[9]. Isotopic fractionation arises, in part, from the enzymatic hydrolysis of dissolved organic phosphorus (DOP). Analysis of oxygen isotopes in Lake Erie samples indicated a non-riverine source of phosphate was supplying the algal bloom, but the analysis could not establish whether organic or inorganic forms were the source of this phosphorus[9]. DOP is rarely analyzed in environmental samples because (1) concentrations are low [8, 10, 11], and (2) the low elemental abundance of phosphorus within dissolved organic matter (DOM) makes detection using mass

spectrometry difficult[12-14]. For example, DOP typically accounts for less than 15% of TDP, or around 5-15 μg P L^{-1} in the Lake Erie waterways[7]. Analysis of DOP has been disregarded in favor of measuring total dissolved phosphorus (TDP), as TDP effectively defines bioavailable phosphorus[6, 15]. However, TDP obscures the diversity of DOP formulae elucidated through mass spectrometry[14, 16] which may aid in source identification and provide a better understanding of biogeochemical controls in the system.

Electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry (ESI FT-ICR-MS) can provide new insight into the molecular composition of environmental samples through non-targeted identification of phosphorus in dissolved organic matter (DOM). To date, ESI FT-ICR-MS has rarely been used to investigate DOP, and, in some cases, phosphorus has been excluded from these analyses[17, 18] due in part to a low elemental abundance of phosphorus (~0.3%) in organic matter. In general, the negative mode of electrospray ionization has yielded a greater number of DOP formulae compared to positive mode[19, 20]. However, despite its low relative abundance, concentration through chemical precipitation and filtration prior to ESI FT-ICR-MS analysis identified a diversity of recalcitrant DOP formulae in Everglades samples[14]. Organophosphorus compounds can also be concentrated for mass spectrometry analysis using solid phase extraction (SPE), which removes background interferences (*i.e.,* salts) while retaining organic constituents that resemble the original sample[21, 22]. Even in the absence of selective concentration, ESI FT-ICR-MS analysis revealed an abundance of organic nitrogen-, sulfur-, phosphorus-containing compounds in Lake Superior and its tributaries[16]. Based on the frequency of harmful algal blooms and the results of the aforementioned study conducted in Lake Superior, we expected Lake Erie would be replete in unique organic phosphorus compounds that could be related to tributary sources.

The objective of this study was to characterize organic-bound phosphorus from select point and non-point pollutant sources in a Lake Erie tributary. Using ultrahigh resolution mass spectrometry, we analyzed samples from several nonpoint and point sources. The nonpoint sources included three fertilizers (hog, chicken, and dairy manures), as well as runoff from the edge of an agricultural field previously amended with mineral fertilizers. A sample of municipal wastewater treatment plant (WWTP) effluent was collected as a point source. Finally, a grab sample was taken from the Sandusky River. The following data describe the signatures of nutrient pollutants in these sources, including the understudied pool of organic phosphorus.

Methods

Site Description and Sample Collection

Sampling was performed in the Sandusky River tributary system, which drains into the Western Lake Erie Basin at Sandusky Bay. The Sandusky River is dominated by nonpoint phosphorus pollution (90%) with smaller contributions from point (9%) and atmospheric (1%) sources (Figure 1A)[23]. The primary land use in the watershed is agricultural, with the vast majority of fertilizer application derived from inorganic (66%) forms rather than from manure (27%) or biosolids (7%)[23]. Most of the manure applied in the Lake Erie basin originates from cattle (50%), hog (34%), and poultry (5%) sources (Figure 1B)[23]. Sampling was conducted on March 14, 2016 following a precipitation event (Figure 1C). At the time of collection, flows were high (>90th percentile) and corresponded with a high total phosphorus load (www.heidelberg.edu/NCWQR)[7]. Six samples were collected from the Sandusky River tributary network (Figure 1D), including (1) an edge of field site, (2) hog manure, (3) chicken (poultry) manure, (4) dairy (cattle) manure, and (5) wastewater treatment plant (WWTP) effluent.

Downstream of these sampling locations, another sample was collected from the (6) Sandusky River.

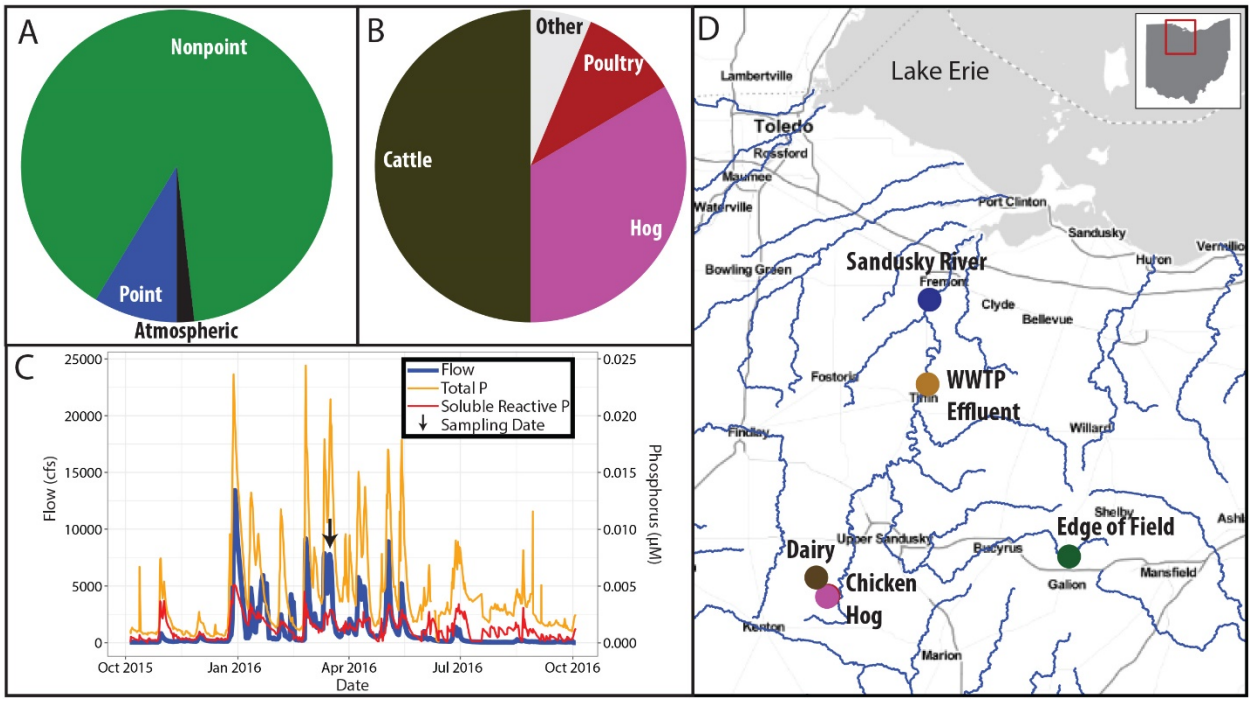


Figure 1. (**A**) The Ohio Lake Erie Task Force (Ohio EPA) has estimated the nonpoint contribution from point and nonpoint sources for the Sandusky River[23]. (**B**) This group has also detailed the contributions of various manures, as elemental P, to the Western Lake Erie basin[23]. (**C**) Flow, total phosphorus, and soluble reactive phosphorus were reported in the 2015-2016 water year by Heidelberg University (www.heidelberg.edu/NCWQR)[7]. The arrow shows the flow conditions at the time of sampling. (**D**) Six samples were collected from the Sandusky River watershed situated in north-central Ohio. The chicken and hog samples were collected on the same property.

Sampling equipment was pre-conditioned by triple rinsing sampling devices and storage containers with Milli-Q water. The chicken manure sample was retrieved from the center of an open-air stockpile following excavation by landowner equipment. The hog manure sample was sampled from a hog manure pit using a PVC sampling device. Dairy manure was collected from

a secondary lagoon using the PVC sampling device. An edge of field sample was collected from the mouth of a tile drainage pipe flowing into the connected stream. At this field, fertilizer was applied as monoammonium phosphate (11-52-0, NPK) in 2014 and anhydrous ammonia (82-0-0) in 2015. No fertilizer was added in 2016 for soybean planted that year. Wastewater effluent was collected following chlorination but prior to discharge from the Tiffin Water Pollution Control Center. Finally, the Sandusky River sample was collected from the faucet of the USGS station (USGS 04198000). All samples were collected in pre-rinsed (Milli-Q water) polyethylene containers, transported on ice to the OSU Environmental Biotechnology Laboratory, and held at 4°C. Wet samples were processed within 24 hours.

Sample Processing

The dry weight of manure samples was determined by weighing subsamples into porcelain dishes and heating for 24 hours at 70°C. Following the dry weight determination, duplicate manure samples were suspended at equivalent ratios of water to dry weight (15:1) using Milli-Q water[19]. The manure-water mixtures were equilibrated overnight at 4°C. Combusted glassware (30 min at 500°C) was used for the remainder of sample preparation. All samples were vacuum-filtered through pre-rinsed (methanol and DI water) 0.7 µm glass fiber filters (Whatman GF/F). The concentrations of dissolved organic carbon (NPOC) and nitrogen (TDN) were determined using a Shimadzu TOC-V/TNM-1 analyzer. Phosphorus (TDP) concentrations were measured using an Agilent ICP-AES. Samples were run as previously described for NPOC/TDN[24], while TDP was measured at wavelength 213.648 nm[25]. Each sample was prepared in duplicate at a concentration of 6.5 mg L^{-1} NPOC in preparation for solid-phase extraction.

We previously determined that the Plexa-PAX solid phase extraction columns were most efficient at retaining organic phosphorus compounds used as laboratory standards (Figures S1 and S2). Thus, Plexa-PAX SPE columns were used for the concentration of DOM. The 6 samples we collected were prepared in duplicate along with two reference standards (Pony Lake Fulvic Acid [PLFA], Suwannee River Fulvic Acid [SRFA]) for a total of 14 samples. Briefly, columns were prepared by wetting with 3 mL 100% HPLC grade methanol and were then rinsed with 2L DI water. While still wet, 275 mL of each sample were gravity filtered through the conditioned SPE columns to collect and concentrate the organic contents. The binding efficiency of samples was calculated from the concentrations of C, N, and P measured before and after SPE filtration (Table S1).

Samples were eluted from the columns with 5 mL of HPLC grade methanol, followed by 5 mL of methanol with 5% formic acid. These elutions were combined into amber glassware and stored at -20°C. The samples were shipped on dry ice to the Woods Hole Oceanographic Institution for ESI(-) FT-ICR MS analysis.

ESI FT-ICR-MS Data Analysis

Mass spectrometry data was collected as previously described (SI Methods)[16]. Peaks were detected in the range of 200-1000 Da. Molecular formula assignments were made with the Compound Identification Algorithm[17, 18]. A total of 14,637 unique peaks were detected under this analysis (SI Excel file). Quality controls were used to quality filter the dataset: peaks observed in DI water or solvent blanks, and singletons were removed (SI Table S2). Only *m/z* values with an assigned formula were considered for further analysis.

Additional data analyses were performed using R Statistics (version 3.1.1). The distributions of peak heights and *m/z* values were compared among replicates and samples. Then the peak heights were normalized to the sum of peaks for each replicate. Replicates were combined through averaging of these normalized peak heights. Sample similarity was compared based on presence/absence of the formula using Venn Euler diagrams, and based on relative peak heights using a Bray-Curtis dissimilarity matrix generated by the 'vegan' package. A list of organic phosphorus formulae, shared between the Sandusky River and at least one source material, were filtered as a subset from the data. Putative tracers were further screened from this list with the stipulation that the maximum relative peak height was observed in a source sample.

Results

The amount of carbon, nitrogen, and phosphorus varied across samples (SI Figure S3), with manure-extracted DOM having considerably higher concentrations relative to other aqueous samples. The manure samples had nutrient concentrations in the range of 30-76 mg C L^{-1}, 12-60 mg N L^{-1}, and 4.8-9.6 mg P L^{-1} as compared to WWTP effluent, edge of field, and Sandusky River samples (6.5-8.8 mg C L^{-1}, 2.6-11 mg N L^{-1}, and <0.03-0.09 mg P L^{-1}). The influent and effluent concentrations for these samples were measured to estimate the amounts that were retained by PAX columns. PAX extraction efficiency varied considerably, with 8-44% C, 6-41%N, and <0-100% P retained by the columns (SI Table S1). Despite a low recovery in a few replicates, average recoveries across our samples were within the range of those reported in other studies using solid phase extraction[20, 22]. Moreover, several P recoveries reported in S1 were derived from low concentration measurements near our analytical quantitation limits.

ESI FT-ICR-MS analysis was used to characterize the molecular properties of organic matter isolated from the six sample materials (Figure S4). A total of 7250 formulae were

identified in the dataset following QA/QC removal of contaminants and formula assignment (SI Table S2). Reproducibility between replicates were generally high (>80% shared formula) for all six samples (SI Table S2), which allowed for a combination of replicate data by averaging the normalized signal between duplicates as well as including any detected formulae in the final representative sample (SI Table S3). The number of identified formulae ranged from 1803 to 4522 across the samples (Figure 2A and SI Table S3). Despite the considerably higher concentration of C, N, and P in the manures, they held a significantly lower number of total DOM formulae (between 1803 and 2473) than the other samples (3136 to 4522). Within these data, the number of formulae containing a P atom ranged from 132 to 313 across the six sources, representing between 3.3% and 12.8% of detected formulae (Figure 2B). The manure samples contained the greatest proportion of DOP formulae (10.9% to 12.8%), 2 to 3-fold what was detected in the edge of field, WWTP effluent, and Sandusky River samples (3.3% to 5.3%). Manure samples also had a greater abundance of formulae with N or S atoms compared to the other samples, with CHON representing 30-40% of the manure formulae versus 16-21% in the other three samples.

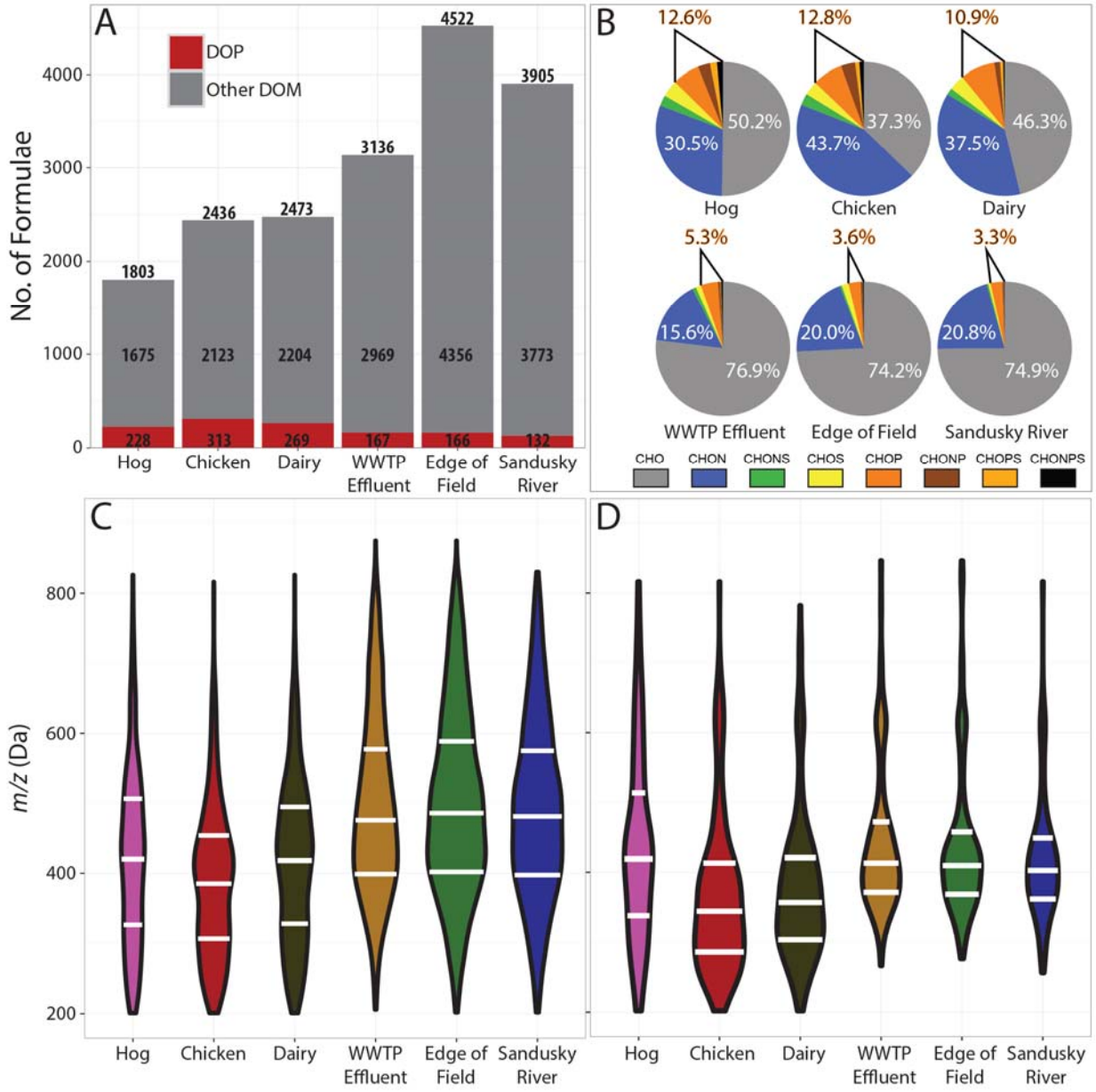


Figure 2. (**A**) The number of assigned formulae representing DOM (full bar) and DOP (red bar) varied across the six watershed samples. Actual values are printed within their respective bars, with the total noted above. Note that any formula containing a P atom was considered to be DOP. (**B**) The proportional distribution of major atom classes for each sample shown with pie charts, with percentages indicating the total proportion of DOP. The distribution of (**C**) DOM and (**D**) DOP *m/z* values were visualized using kernel-based cumulative density plots (violin plots). The width of each band indicates the kernel-based density of *m/z* values relative to total number, with white bands representing sample quartiles.

In our assessment of the distribution of molecular formula, we observed the manure samples to be composed of a greater number of low molecular mass formulas compared to the other samples (Figure 2C). Specifically, the median molecular mass of observed *m/z* values for hog (420 Da), chicken (387 Da), and dairy (419 Da) manures were on average 70 Da lower than that of the WWTP effluent (474 Da), edge of field (485 Da) and Sandusky River (481 Da). DOP formula generally followed this trend, with the chicken (341 Da) and dairy (355 Da) manures having a lower median mass than the WWTP effluent (414 Da), edge of field (412 Da), and Sandusky River (404 Da) samples (Figure 2C). The hog manure sample had the highest median molecular mass of DOP *m/z* values (424 Da). Furthermore, unlike the other 5 samples, which were shifted toward lower molecular mass of DOP relative to DOM, the hog manure DOP molecular mass distribution resembled that of its overall DOM.

Sample signatures were visualized using Van Krevelen diagrams, which relate the C:H to the O:C molar ratios for all observed formulas (Figure 3A). The relative placement of each formula provides an estimation of molecular class, which we refer to as "–like" types. The overall scatter had considerable differences across sources, with manure samples exhibiting a greater diversity of molecular type classes (*i.e.,* more scatter) compared to the Sandusky River, edge of field, and WWTP effluent samples, which more tightly clustered around the lignin-like features. To further highlight these differences in overall scatter, we tallied the relative abundance of formulae within the 7 different molecular classes (Figure 3B). Although the majority of formulae across all samples were lignin-like (51-80%), the WWTP effluent, edge of field, and Sandusky River samples were especially dominated by lignin- and tannin-like features (86-88%) compared to manures. In contrast, the three manure samples consisted of a higher

proportion of most other molecular classes, notably protein-, lipid, and carbohydrate-like features.

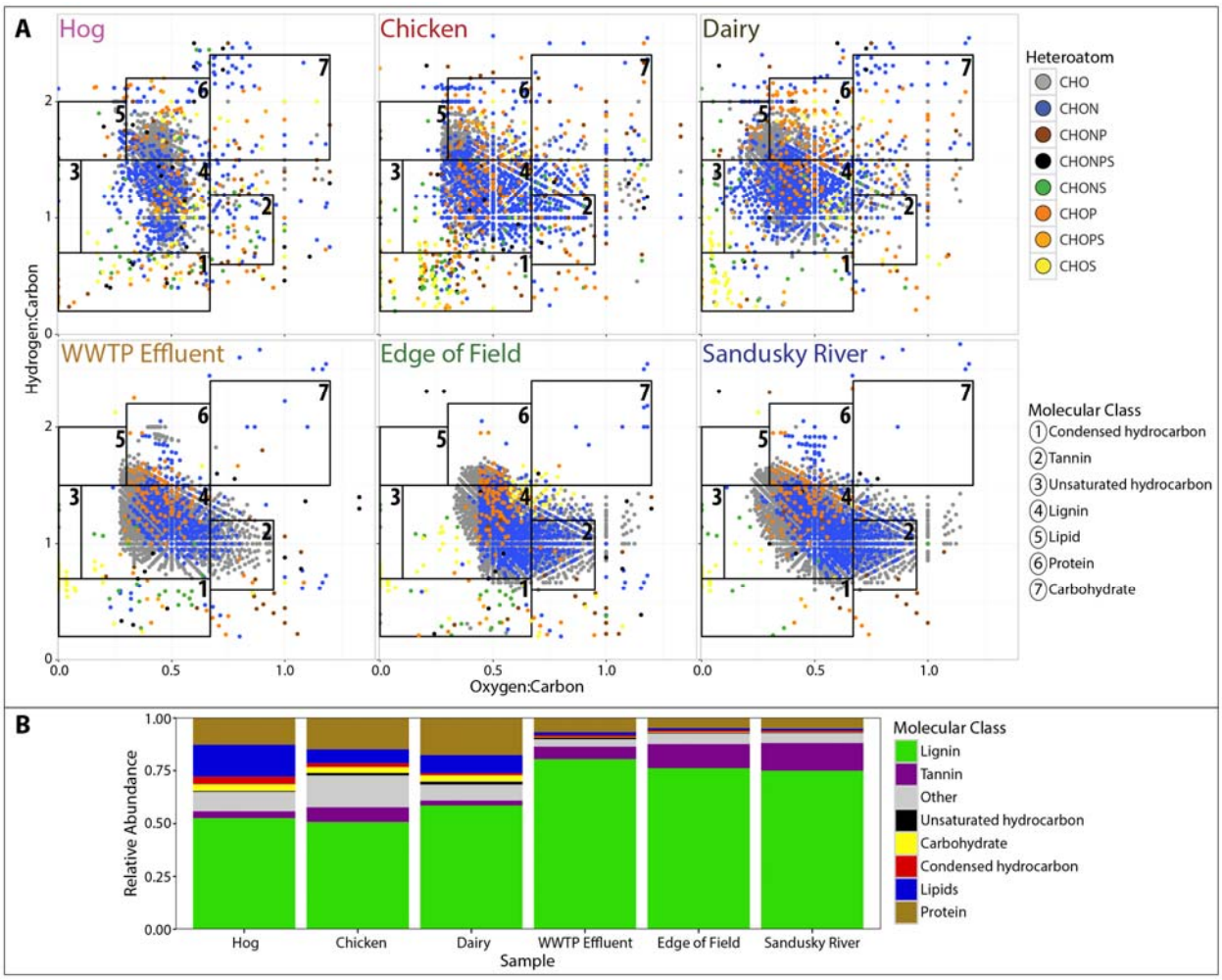


Figure 3. (**A**) Van Krevelen diagrams showing the molar ratio of hydrogen:carbon versus oxygen:carbon for each assigned formula, color-coded based on atomic composition. Lipid-, protein-, carbohydrate-, unsaturated hydrocarbon-, lignin-, tannin-, and uncondensed hydrocarbon-like molecular class ranges are represented by boxes. (**B**) The relative abundance of molecular classes was summarized for each sample.

To further probe sample similarity, we compared DOM and DOP signatures based upon the abundance of shared formulae using Euler diagrams (Figure S5). Although they were collected from 12 to 41 miles apart, the WWTP effluent, edge of field, and Sandusky River samples shared 54% of all assigned DOM formula. When we calculated the intersection between

the Sandusky River and either the WWTP effluent or the edge of field sample, over 84% of assigned formulae were shared for both data sets. This level of similarity is comparable to our replicates (81-90% shared formula, SI Table S2). Interestingly, the edge of field sample shared a considerably greater number (124) and percentage (75%) of DOP formula with the Sandusky River as compared to the WWTP effluent sample (98, 59%).

Across the three manure samples, only 33% of the formulae were shared with the Sandusky River. In pairwise comparisons, the intersection between the Sandusky River and individual manures ranged from 31-36%. Moreover, these shared formulae were not unique among manures and the Sandusky River; all but one were also present in the WWTP effluent and edge of field runoff samples. Only five *m/z* values were uniquely shared between the manures and Sandusky River sample.

We expanded our analysis to consider the similarity of two NOM standards (PLFA and SRFA) based on relative peak heights of all assigned formulae using hierarchal clustering analysis (Figure 4A-B). Dendrogram clustering patterns for DOM (Figure 4A) reaffirmed the calculated similarities between the Sandusky River, edge of field, and WWTP effluent samples based on the presence/absence analysis (Figure S5A). These three samples and the NOM standards formed a separate branch from the manure samples, which exhibited greater dissimilarity between one another and the rest of the samples (Figure 4A). When considering only DOP formulae, dissimilarity grew between all samples, although the two major branches remained the same (Figure 4B). We found that the SRFA clustered among our samples for DOM, yet when we only considered DOP, both SRFA and PLFA were separated from the WWTP effluent, edge of field, and Sandusky River samples. These two NOM standards are

primarily derived from a terrestrial origin, with the Pony Lake (Antarctica) more geographically remote and less anthropogenically-impacted than the Suwannee River (Georgia, USA).

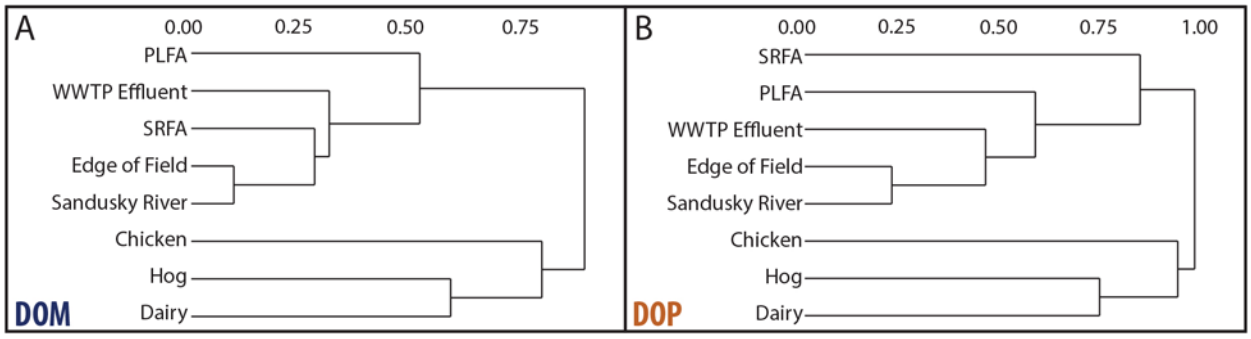


Figure 4. Hierarchal clustering dendrograms for (**A**) DOM and (**B**) DOP prepared from Bray-Curtis dissimilarity matrices generated from relative peak heights from *m/z* values with assigned formulae. Numbers along the top reflect the level of dissimilarity between samples at the branch point.

In an effort to identify phosphorus formulae originating from our point source (WWTP effluent) and non-point sources (all others) in the Sandusky River, we generated a list of DOP formulae present in at least one sample and the Sandusky River. The list was further screened to remove formulae that increased in peak abundance from the source to the Sandusky River, as this could indicate origination of these *m/z* values within the river. Our filtering resulted in 72 formulae, which we propose could serve as markers for detecting or tracking source-derived nutrients (SI Table S5). We next identified formulae from this list that were unique to the (1) edge of field, (2) WWTP effluent, or (3) the three manure samples (Figure 5). The relative peak height for the manure marker was an order of magnitude higher than was observed in the Sandusky River, while peak heights for edge of field and WWTP effluent markers were comparable between the source and Sandusky River sample. Marker compounds are one

potential tool for regulatory agencies to identify the sources of nutrient contamination in waterways.

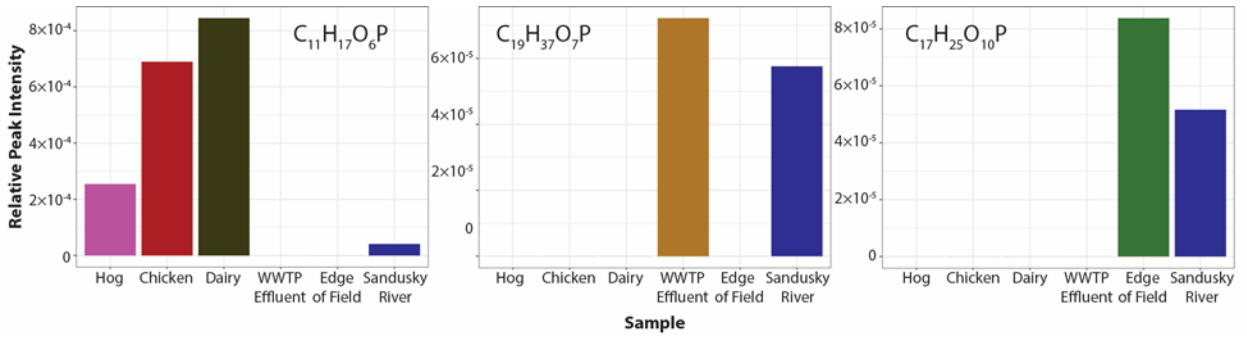


Figure 5. Relative peak heights for potential markers for detecting or tracking source-derived DOP nutrients shared uniquely by the Sandusky River and either the (1) three manures, (2) WWTP effluent, or (3) edge of field samples. Please note different scales on y-axis for each graph.

Discussion

Worldwide, eutrophication has been linked to human agricultural practices including dense animal operations and/or the application of inorganic fertilizers to crop fields[26, 27]. In Lake Erie tributaries where land use is dominated by agriculture, the majority of phosphorus is thought to be derived from inorganic fertilizer applied to fields[7, 23]. However, this finding relies upon bulk phosphorus analyses of total or dissolved reactive P, measurements that cannot be used to discriminate between point and non-point pollution sources within the watershed. Our ultrahigh resolution MS analysis showed that DOM and DOP signatures collected from drainage tiles at the edge of an agricultural field in the Sandusky River were highly similar (84% DOM, 75% DOP) to that of the river itself collected 41 miles downstream. This level of similarity is remarkable considering Sandusky River replicates shared 85% of *m/z* values. Closer in hydrologic proximity (12 miles between sampling locations), the Sandusky River and WWTP effluent sample also had similar DOM (84% shared formula), but were more dissimilar in their

DOP (59% shared *m/z* values). It is notable that the edge of field and Sandusky River are most alike in their DOP character, as this finding is consistent with the type of nutrient pollution, primary land use, and fertilizer form previously reported for the Sandusky River[23].

The signatures of the three manure samples were vastly different from all of the other samples. Manures account for 27% of total P applied as fertilizer to agricultural systems for the Lake Erie basin[23] serving as a rich source of natural fertilizer despite challenges associated with their handling. Our analysis shows manure samples are abundant in N- (> 30%) and P- (>10%) containing organic molecules that are easily liberated from the solids by water. The DOM that was extracted from these manures in our labs had higher relative phosphorus and nitrogen concentrations than our other samples, and this can likely explain the high abundance of DOP formulae. Manure DOM also consisted of lower molecular *m/z* values, relative to the other samples, which may represent more labile compounds that are easily assimilated into the ecosystem[28]. The lower molecular weights of these molecules may be due to mineralization or enzymatic degradation of complex materials within the digestive tracts of these animals[29, 30]. Future studies should consider the signatures associated with manure-applied field runoff.

DOP and DOM signatures from point and non-point sources would be altered by abiotic (*i.e.*, photodegradation) or biotic (*i.e.,* biodegradation) processes in soils, groundwater and surface waters as it moves through the watershed. In particular, the transport of organic compounds, derived from biosolids and manure, through porous media into the water column could be retarded by adsorption to solid materials[31, 32]. Sorption affinity of phosphorus is specific to each compound and is also affected by soil type[33] therefore we would expect hydrophilic compounds to be more prominent in manure-derived runoff. Although the molecular masses in

365 our analysis provide little information about hydrophobicity, the extraction method we used to
366 obtain our manure DOM was likely to have selected for the more hydrophilic compounds
367 The DOM of our samples consisted of between 3-13% DOP, with 132-313 P-formulae
368 detected in each sample. While the manures were loaded with these P-formulae, the values
369 observed in the edge of field and Sandusky River samples appeared low in comparison to those
370 reported in the literature. Specifically, between 3.8-12% DOP (89-481 formulae) were reported
371 in Lake Superior and its tributaries while New Zealand coastal waters consisted of 5-27% (82-
372 293 formulae) phosphorus[16,20]. Our Sandusky River and edge of field samples were in the lower
373 range of these reported values, both in terms of percent and by the number of P-formulae
374 detected. In general, Lake Erie tributaries are loaded with dissolved phosphorus primarily in the
375 form of inorganic orthophosphate[7]. That inorganic phosphate can be readily assimilated by most
376 plants and organisms, while organic phosphorus requires enzymatic cleavage[13]. Natural organic
377 phosphorus exists in the P(V) (organophosphates) or P(III) state (organophosphonates), with the
378 latter requiring enzymatic oxidation to phosphate following liberation of the phosphonate
379 groups[13, 34]. Conversely, organophosphates are directly hydrolyzed into inorganic phosphate by
380 enzymes such as alkaline phosphatase[13, 15]. Organophosphonates therefore require a greater
381 investment of activation energy that has been found to slow microbial growth, leading to a
382 buildup of these compounds in natural systems[13, 35]. Organophosphorus can be utilized
383 concurrently with inorganic P, but its nutrient value is greater when total phosphorus supplies are
384 limited[15, 36]. The Lake Erie tributary network has relatively high phosphate concentrations
385 compared to other aquatic systems suggesting organic phosphorus turnover will be slower
386 relative to inorganic forms in its rivers and streams[7]. Whether these compounds persist and
387 accumulate in the lake remains to be seen.

Certain organic molecules in these samples are more susceptible to chemical transformations and would be more readily assimilated by microorganisms. The WWTP effluent sample was enriched with microbially-derived features (e.g. lipid- and protein-like) from the activated sludge process, while the edge of field and Sandusky River, like the SRFA standard, were greatly dominated by lignin. Tannin- and lignin-like features are regarded to have a terrestrial (plant-derived) origin compared to protein-, carbohydrate-, and lipid-like features which instead originate from endogenous microorganisms[16, 37]. In addition to indicating the source material, these molecular classes also correlate to the nominal oxidation state of carbon (NOSC, Figure S6), which describes the redox potential of the formula. Specifically, tannin-like features are oxidized; lipid-like features are reduced; and lignin-like features have an average oxidation state around zero[38]. This suggests that the reduced lipid- and protein-like features more common to the manure formula can be expected to oxidize during the transport in aerobic surface waters. We would therefore expect manure derived DOM to undergo the greatest amount of signature change relative to other samples. A targeted analysis of similar samples (e.g. using LC MS/MS of authentic standards to validate these proposed compounds) would be useful in elucidating these structures of *m/z* values shared between our samples, which would provide greater insights to their molecular properties[39].

The ESI FT-ICR MS approach is largely qualitative in nature and was used here to discern organic phosphorus signatures in distinct point- and non-point pollutant sources. While pollutant concentrations are likely correlated to the strength of detected peaks (i.e., peak heights) [40, 41], the potential for multiply-charge ions, differences in ionization efficiency, and multiple formulae for each *m/z* value currently limit the application of these tools in a more quantitative way[14, 42]. Therefore, although this approach holds merit for identifying end members in a

hydrologically complex system, a more integrated and quantitative approach is necessary if these tools are to be applied in practice. Studies that better connect ESI FT-ICR MS data to pollutant concentrations and marker compounds would greatly advance the feasibility of this approach for land management purposes.

The similarity observed in our ultrahigh resolution MS analysis between an agricultural edge of field sample and a Sandusky River supports the previously reported data that nutrient loads are predominantly sourced from agricultural fields in this Lake Erie tributary. However, the edge of field, WWTP effluent, and Sandusky River samples are hydrologically connected and would be expected to share some background DOM signature received from rainwater, runoff, and/or groundwater in the watershed. Still, the DOP signatures were highly divergent between manures and other source materials, which should allow us to detect the presence of these nonpoint and point sources in the tributary network. One limitation of this study is the number of end members evaluated. We would not necessarily expect the hog, chicken, and dairy manure signatures reported here to encompass signatures for all animal manures used in the watershed. However, given the high degree of separation between manures and other samples presented here, we would expect to be able to differentiate between most manure and non-manure sources in a watershed. Future studies could consider a broader spectrum of manures and their processing in order to better characterize these highly variable nutrient pollutant sources. Despite our low sample number limitation, this study identified formulae shared between the Sandusky River and other samples that could serve as source markers in the watershed. Additionally, we elucidated unique DOM and DOP signatures which could be tested by regulatory agencies for detecting and monitoring the presence of nutrient pollutant sources in similar tributaries.

ASSOCIATED CONTENT

Supporting Information

The solid phase extraction trials; molecular formula assignment methodology; sample elemental compositions; Venn Euler diagrams of shared molecules; molecular class NOSC values; sample sorption efficiency; full ESI FT-ICR-MS results; quality filtering of samples; number of detected peaks across samples; Venn count data; and tracer formula are provided in supporting information available free of charge on the ACS Publication website.

AUTHOR INFORMATION

Corresponding Authors

*Phone: 614-247-4429 Email: mouser.19@osu.edu

Phone: 614-648-9876 Email: brooker.26@osu.edu

ACKNOWLEDGEMENTS

This research was supported by an Ohio State University Field to Faucet Institute award to P.J.M and by a Harmful Algal Bloom Research Initiative grant from the Ohio Department of Higher Education. We thank Dr. Laura Johnson and her research team from Heidelberg University; Drs. Kevin McCluney and Bob Midden of Bowling Green State University; Kendall Stuckey of USGS; and Dr. Kevin King of USDA for their help in designing a sampling plan and identifying sampling locations. We thank Melissa Kido Soule for analysis of the FT-ICR MS samples.

References

1. Conroy, J.D.; Boegman, L.; Zhang, H.; Edwards, W.J.; Culver, D.A. "Dead Zone" dynamics in Lake Erie: the importance of weather and sampling intensity for calculated hypolimnetic oxygen depletion rates. *Aquat. Sci.* **2011,** *73* (2), 289-304; http://dx.doi.org/10.1007/s00027-010-0176-1.

2. Conley, D.J.; Paerl, H.W.; Howarth, R.W.; Boesch, D.F.; Seitzinger, S.P.; Havens, K.E.; Lancelot, C.; Likens, G.E. Controlling eutrophication: nitrogen and phosphorus. *Science* **2009,** *123*, 1014-1015; http://dx.doi.org/10.1126/science.1167755.

3. Michalak, A.M.; Anderson, E.J.; Beletsky, D.; Boland, S.; Bosch, N.S.; Bridgeman, T.B.; Chaffin, J.D.; Cho, K.; Confesor, R.; Daloglu, I.; Depinto, J.V.; Evans, M.A.; Fahnenstiel, G.L.; He, L.; Ho, J.C.; Jenkins, L.; Johengen, T.H.; Kuo, K.C.; Laporte, E.; Liu, X.; McWilliams, M.R.; Moore, M.R.; Posselt, D.J.; Richards, R.P.; Scavia, D.; Steiner, A.L.; Verhamme, E.; Wright, D.M.; Zagorski, M.A. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (16), 6448-52; http://dx.doi.org/10.1073/pnas.1216006110.

4. Steffen, M.M.; Davis, T.W.; McKay, R.M.L.; Bullerjahn, G.S.; Krausfeldt, L.E.; Stough, J.M.A.; Neitzey, M.L.; Gilbert, N.E.; Boyer, G.L.; Johengen, T.H.; Gossiaux, D.C.; Burtner, A.M.; Palladino, D.; Rowe, M.D.; Dick, G.J.; Meyer, K.A.; Levy, S.; Boone, B.E.; Stumpf, R.P.; Wynne, T.T.; Zimba, P.V.; Gutierrez, D.; Wilhelm, S.W. Ecophysiological Examination of the Lake Erie Microcystis Bloom in 2014: Linkages between Biology and the Water Supply Shutdown of Toledo, OH. *Environ. Sci. Technol.* **2017,** *51* (12), 6745-6755; http://dx.doi.org/10.1021/acs.est.7b00856.

5. Stumpf, R.P.; Wynne, T.T.; Baker, D.B.; Fahnenstiel, G.L. Interannual variability of cyanobacterial blooms in Lake Erie. *PloS one* **2012,** *7* (8), e42444; http://dx.doi.org/10.1371/journal.pone.0042444.

6. *Ohio Lake Erie Phosphorus Task Force Final Report II*; Ohio EPA: Columbus, Ohio; http://www.epa.ohio.gov/portals/35/lakeerie/ptaskforce2/Task_Force_Report_October_2013.pdf.

7. Baker, D.B.; Confesor, R.; Ewing, D.E.; Johnson, L.T.; Kramer, J.W.; Merryfield, B.J. Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga Rivers: The importance of bioavailability. *J. Great Lakes Res.* **2014,** *40* (3), 502-517; http://dx.doi.org/10.1016/j.jglr.2014.05.001.

8. Baldwin, D.S. Reactive “organic” phosphorus revisited. *Water Res.* **1998,** *32* (8), 2265-2270; http://dx.doi.org/10.1016/S0043-1354(97)00474-0.

9. Elsbury, K.E.; Paytan, A.; Ostrom, N.E.; Kendall, C.; Young, M.B.; McLaughlin, K.; Rollog, M.E.; Watson, S. Using oxygen isotopes of phosphate to trace phosphorus sources and cycling in Lake Erie. *Environ. Sci. Technol.* **2009,** *43* (9), 3108-14; http://dx.doi.org/10.1021/es8034126.

10. Monaghan, E. J., Ruttenberg,K.C., Dissolved organic phosphorus in the coastal ocean: Reassessment of available methods and seasonal phosphorus profiles from the Eel River Shelf. *LNO Limnology and Oceanography* **1999,** *44* (7), 1702-1714; http://dx.doi.org/10.4319/lo.1999.44.7.1702.

11. Ruttenberg, K.C.; Dyhrman, S.T. Dissolved organic phosphorus production during simulated phytoplankton blooms in a coastal upwelling system. *Frontiers in Microbiology* **2012,** *3* (274), 1-13; https://dx.doi.org/10.3389%2Ffmicb.2012.00274.

12. Kruse, J.; Abraham, M.; Amelung, W.; Baum, C.; Bol, R.; Kühn, O.; Lewandowski, H.; Niederberger, J.; Oelmann, Y.; Rüger, C.; Santner, J.; Siebers, M.; Siebers, N.; Spohn, M.; Vestergren, J.; Vogts, A.; Leinweber, P. Innovative methods in soil phosphorus research: A review. *Journal of Plant Nutrition and Soil Science* **2015,** *178* (1), 43-88; http://dx.doi.org/10.1002/jpln.201400327.

13. Karl, D.M. Microbially mediated transformations of phosphorus in the sea: new views of an old cycle. *Annual review of marine science* **2014,** *6*, 279-337; http://dx.doi.org/10.1146/annurev-marine-010213-135046.

14. Cooper, W.T.; Llewelyn, J.M.; Bennett, G.L.; Salters, V.J.M. Mass spectrometry of natural organic phosphorus. *Talanta* **2005,** *66* (2), 348-358; http://dx.doi.org/10.1016/j.talanta.2004.12.028.

15. Ruttenberg, K.C.; Dyhrman, S.T. Temporal and spatial variability of dissolved organic and inorganic phosphorus, and metrics of phosphorus bioavailability in an upwelling-dominated coastal system. *Journal of Geophysical Research: Oceans* **2005,** *110* (C10), 1-22; https://dx.doi.org/10.1029/2004JC002837.

16. Minor, E.C.; Steinbring, C.J.; Longnecker, K.; Kujawinski, E.B. Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. *Org. Geochem.* **2012,** *43*, 1-11; https://dx.doi.org/10.1016/j.orggeochem.2011.11.007.

17. Kujawinski, E.B.; Longnecker, K.; Blough, N.V.; Del Vecchio, R.; Finlay, L.; Kitner, J.B.; Giovannoni, S.J. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochim. Cosmochim. Acta* **2009,** *73* (15), 4384-4399; https://dx.doi.org/10.1016/j.gca.2009.04.033.

18. Kujawinski, E.B.; Behn, M.D. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **2006,** *78* (13), 4363-4373; https://dx.doi.org/10.1021/ac0600306.

19. Ohno, T.; Sleighter, R.L.; Hatcher, P.G. Comparative study of organic matter chemical characterization using negative and positive mode electrospray ionization ultrahigh-resolution mass spectrometry. *Analytical and Bioanalytical Chemistry* **2016,** *408* (10), 2497-2504; http://dx.doi.org/10.1007/s00216-016-9346-x.

20. Gonsior, M.; Peake, B.M.; Cooper, W.T.; Podgorski, D.C.; D'Andrilli, J.; Dittmar, T.; Cooper, W.J. Characterization of dissolved organic matter across the Subtropical Convergence off the South Island, New Zealand. *Mar. Chem.* **2011,** *123* (1), 99-110; https://dx.doi.org/10.1016/j.marchem.2010.10.004.

21. Ohno, T.; Ohno, P.E. Influence of heteroatom pre-selection on the molecular formula assignment of soil organic matter components determined by ultrahigh resolution mass spectrometry. *Analytical and Bioanalytical Chemistry* **2013,** *405* (10), 3299-3306; https://doi.org/10.1007/s00216-013-6734-3.

22. Raeke, J.; Lechtenfeld, O.J.; Wagner, M.; Herzsprung, P.; Reemtsma, T. Selectivity of solid phase extraction of freshwater dissolved organic matter and its effect on ultrahigh resolution mass spectra. *Environmental Science: Processes & Impacts* **2016,** *18* (7), 918-927; http://dx.doi.org/10.1039/c6em00200e.

23. *Ohio Lake Erie Phosphorus Task Force Final Report I*; Ohio Environmental Protection Agency, Division of Surface Water: Columbus, Ohio; http://www.epa.ohio.gov/portals/35/lakeerie/ptaskforce/Task_Force_Final_Report_April_2010.pdf.

24. Kekacs, D.; Drollette, B.D.; Brooker, M.R.; Plata, D.L.; Mouser, P.J. Aerobic biodegradation of organic compounds in hydraulic fracturing fluids. *Biodegradation* **2015,** *26* (4), 271-87; http://dx.doi.org/10.1007/s10532-015-9733-6.

25. Bartos, J.M.; Boggs, B.L.; Falls, J.H.; Siegel, S.A. Determination of phosphorus and potassium in commercial inorganic fertilizers by inductively coupled plasma-optical emission spectrometry: single-laboratory validation. *J. AOAC Int.* **2014,** *97* (3), 687-699; http://dx.doi.org/10.5740/jaoacint.12-399.

26. Torrent, J.; Barberis, E.; Gil-Sotres, F. Agriculture as a source of phosphorus for eutrophication in southern Europe. *Soil Use Manage.* **2007,** *23* (s1), 25-35; https://dx.doi.org/10.1111/j.1475-2743.2007.00122.x.

27. Carpenter, S.R.; Caraco, N.F.; Correll, D.L.; Howarth, R.W.; Sharpley, A.N.; Smith, V.H. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.* **1998,** *8* (3), 559-568; https://doi.org/10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2.

28. Ohno, T.; Chorover, J.; Omoike, A.; Hunt, J. Molecular weight and humification index as predictors of adsorption for plant- and manure-derived dissolved organic matter to goethite. *European Journal of Soil Science* **2007,** *58* (1), 125-132; https://doi.org/10.1111/j.1365-2389.2006.00817.x.

29. Calderón, F.J.; McCarty, G.W.; Reeves, J.B. Pyrolisis-MS and FT-IR analysis of fresh and decomposed dairy manure. *J. Anal. Appl. Pyrolysis* **2006,** *76* (1), 14-23; https://doi.org/10.1016/j.jaap.2005.06.009.

30. Golovan, S.P.; Meidinger, R.G.; Ajakaiye, A.; Cottrill, M.; Wiederkehr, M.Z.; Barney, D.J.; Plante, C.; Pollard, J.W.; Fan, M.Z.; Hayes, M.A. Pigs expressing salivary phytase produce low-phosphorus manure. *Nat. Biotechnol.* **2001,** *19* (8), 741-745; https://doi.org/10.1038/90788.

31. Sharma, R.; Bella, R.W.; Wong, M.T.F. Dissolved reactive phosphorus played a limited role in phosphorus transport via runoff, throughflow and leaching on contrasting cropping soils from southwest Australia. *Sci. Total Environ.* **2017,** *577*, 33-44; http://dx.doi.org/10.1016/j.scitotenv.2016.09.182.

32. Dodd, R.J.; Sharpley, A.N. Recognizing the role of soil organic phosphorus in soil fertility and water quality. *Resources, Conservation and Recycling* **2015,** *105* (2), 282-293; http://dx.doi.org/10.1016/j.resconrec.2015.10.001.

33. Berg, A.S.; Joern, B.C. Sorption dynamics of organic and inorganic phosphorus compounds in soil. *J. Environ. Qual.* **2006,** *35* (5), 1855-1862; http://dx.doi.org/10.2134/jeq2005.0420.

34. Pasek, M.A.; Sampson, J.M.; Atlas, Z. Redox chemistry in the phosphorus biogeochemical cycle. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (43), 15468-15473; https://doi.org/10.1073/pnas.1408134111.

35. Adams, M.M.; Gómez-García, M.R.; Grossman, A.R.; Bhaya, D. Phosphorus deprivation responses and phosphonate utilization in a thermophilic *Synechococcus* sp. from microbial mats. *J. Bacteriol.* **2008,** *190* (24), 8171-8184; http://dx.doi.org/10.1128/JB.01011-08.

36. Björkman, K.M.; Karl, D.M. Bioavailability of dissolved organic phosphorus in the euphotic zone at Station ALOHA, North Pacific Subtropical Gyre. *Limnology and Oceanography* **2003,** *48* (3), 1049-1057; http://dx.doi.org/10.4319/lo.2003.48.3.1049.

37. Feng, L.; Xu, J.; Kang, S.; Li, X.; Li, Y.; Jiang, B.; Shi, Q. Chemical Composition of Microbe-Derived Dissolved Organic Matter in Cryoconite in Tibetan Plateau Glaciers: Insights from Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Analysis. *Environ. Sci. Technol.* **2016,** *50* (24), 13215-13223; https://doi.org/10.1021/acs.est.6b03971.

38. Boye, K.; Noel, V.; Bone, S.E.; Bargar, J.R.; Boye, K.; Fendorf, S.; Tfaily, M.M.; Williams, K.H. Thermodynamically controlled preservation of organic carbon in floodplains. *Nature Geoscience* **2017,** *10* (6), 415-419; https://doi.org/10.1038/NGEO2940.

39. Lee, M.S.; Kerns, E.H. LC/MS applications in drug development. *Mass Spectrom. Rev.* **1999,** *18* (3-4), 187-279; https://doi.org/10.1002/(SICI)1098-2787(1999)18:3/4%3C187::AID-MAS2%3E3.0.CO;2-K.

40. Kamga, A.W.; Behar, F.; Hatcher, P.G. Quantitative Analysis of Long Chain Fatty Acids Present in a Type I Kerogen Using Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Compared with BF_3/MeOH Methylation/GC-FID. *Journal of The American Society for Mass Spectrometry* **2014,** *25* (5), 880-890; http://dx.doi.org/10.1007/s13361-014-0851-x.

603 41. Lu, Y.H.; Li, X.; Mesfioui, R.; Bauer, J.E.; Chambers, R.M.; Canuel, E.A.; Hatcher, P.G.
604 Use of ESI-FTICR-MS to characterize dissolved organic matter in headwater streams draining
605 forest-dominated and pasture-dominated watersheds. *PloS one* **2015,** *10* (12), e0145639;
606 https://doi.org/10.1371/journal.pone.0145639.

607 42. Banerjee, S.; Mazumdar, S. Electrospray ionization mass spectrometry: a technique to access
608 the information beyond the molecular weight of the analyte. *International journal of analytical*
609 *chemistry* **2012,** *2012*; http://dx.doi.org/10.1155/2012/282574.

Supporting Information

Discrete Organic Phosphorus Signatures are Evident in Pollutant Sources within a Lake Erie Tributary

Brooker, MR[1,2]*; Longnecker, K[3]; Kujawinski, EB[3]; Evert, MH[1,2]; and Mouser, PJ[1,4]*.

[1]Department of Civil, Environmental, and Geodetic Engineering, Ohio State University, Columbus, Ohio 43210, United States

[2]Environmental Science Graduate Program, Ohio State University, Columbus, Ohio 43210, United States

[3]Woods Hole Oceanographic Institution, Department of Marine Chemistry and Geochemistry, Woods Hole, MA 02543, USA

[4]Department of Civil and Environmental Engineering, University of New Hampshire, Durham, New Hampshire 03824, USA

This document includes one Excel file, and 19 pages including additional experimental methodology, 5 figures, and 5 tables including: ESI FT-ICR-MS results (SanduskyMaterialResults.xlsx); solid phase extraction efficiency for wastewater (Figures S1); solid phase extraction of organic phosphorus standards (Figure S2); sample elemental compositions (Figure S3); sample spectra (Figure S4); Venn Euler diagrams of shared molecules between samples (Figure S5); molecular class NOSC values (Figure S6); sample sorption efficiency (Table S1); quality filtering of samples (Table S2); number of peaks per sample (Table S3); Venn count data (Table S4); and formula for potential markers (Table S5).

Methods

Prior to ESI(-) FT-ICR MS analysis, the protocols used for organic matter collection were tested to determine their ability to isolate organic phosphorus compounds. Several organic phosphorus compounds were purchased to be used as reference compounds: 2-aminoethyl phosphonate (2-AEP); fosfomycin (FOM); *n*-hexylphosphonic acid (HexP); glucose-6-phosphate (G6P); phenyl phosphate (PhP); nicotinamide dinucleotide phosphate (reduced, NADH); monopotassium phosphate (PO4); and sodium pyrophosphate (P2O7). Each standard was prepared as a stock 1 mg L^{-1} P solution in DI water. A sample of primary clarifier water was collected from the Southerly Wastewater Plant (Columbus, OH) following the methods described in the manuscript. An initial experiment was designed to determine the carbon, nitrogen and phosphorus retention efficiency of four SPE column types (Agilent Bond Elut): functionalized styrene divinylbenzene (PPL); hydrophobic, bonded silica (C18); polymer anionic exchange (PAX); strong anionic exchange (SAX). While the manufacturer instructions call for the adjustment of samples to a pH 2 for the PPL and C18 columns, the PAX and SAX columns recommend adjusting the sample to a pH 10. The primary clarifier water was used to determine the retention of phosphorus by all four columns, both at pH 2 and pH 10 with duplicates for each column (n=16). Following the determination of pH adjustments, a mixture of the reference organic phosphorus compounds were used to determine the retention of these compounds for each filter at pH 10 using duplicate columns. However, due to observed desorption, the SAX columns were excluded from this subsequent analysis (n=6).

Primary clarifier water was passed through the SPE columns using the methods described in the manuscript. The amount of carbon applied to each column type was determined to meet the maximum sorption capacity. The retention efficiency of NPOC and TDN were determined by

the change in concentration between the influent and effluent of samples as measured with the Shimadzu TOC-V/TN. The retention efficiency of TDP was determined by the change in concentration between the influent and effluent samples using an Agilent ICP-OES.

A 7.5 mg L^{-1} P concentration mixture using equal parts (0.9375 mg L^{-1} P) of the eight phosphorus reference compounds – six organic, and two inorganic – was prepared for further analysis of the SPE columns. The mixture was analyzed using ion chromatography with an AS-11HC column on a Dionex ICS-2100 ion chromatograph (Dionex Corporation, Sunnyvale, CA). The flow rate was set at 1.5 mL/min for 15 min a sample, eluted in a 1-60 mM gradient of KOH at 30°C[1]. This method allowed for the detection of seven out of the eight compounds, with the lone exception being 2-AEP. These samples were made basic (pH 10) using KOH and gravity filtered through three SPE column types in duplicate (n=6). The effluent was collected in combusted glassware and ion chromatography analysis was used to visually detect the presence/absence of the compounds following passage through the solid phase columns.

Collection of Mass Spectrometry Data and Peak Detection

The samples were analyzed with electrospray ionization under the negative ionization mode on a 7T FTICR mass spectrometer (Thermo Fisher Scientific, Waltham, MA USA). The instrument settings were optimized by tuning on the SRFA standard. The samples were infused into the ESI interface at 4 μL min^{-1}, and the instrumental and spray parameters were optimized for each sample. The capillary temperature was set at 250°C, and the spray voltage was between 3.7 and 4 kV. For each sample, 200 scans were collected spanning the 200-1000 *m/z* range. An external calibration mixture (Thermo Calibration Mix; Thermo Fisher Scientific) was used to calibrate the mass accuracy to <1.5 ppm. The target average resolving power was 400,000 at *m/z*

400 (where resolving power is defined as m/Δm 50% where Δm is the width at half-height of peak m).

Individual transients as well as a combined raw file were collected using xCalibur 2.0 (Thermo Fisher Scientific). Transients were co-added and processed with custom-written MATLAB code[2]. Only transients with a total ion current >20% of the maximum value observed in each sample were added, processed with Hanning apodization, and zero-filled prior to fast Fourier transformation. All *m/z* values with a signal:noise ratio > 10 were retained. Spectra were internally re-calibrated using a list of *m/z* values present in the majority of samples resulting in a mass accuracy of < 1 ppm[3]. Individual sample peak lists were then aligned in MATLAB[4]. Formula assignments were made through the custom-built Compound Identification Algorithm at the Wood Hole Oceanographic Institution, as previously described[5, 6].

The nominal oxidation state of carbon (NOSC) for each identified formula was calculated according to the equation of *Boye et al*. (2016). The equation is based on the count of individual atom counts according to equation 1. The distribution of the NOSC values were considered for each molecular classification, using only unique formula (no duplicates between ^{12}C and ^{13}C isotopologues).

$$NOSC = 4 - \frac{4C+H-2O-3N-2S+5P}{C} \qquad \text{(equation 1)}$$

Results & Discussion

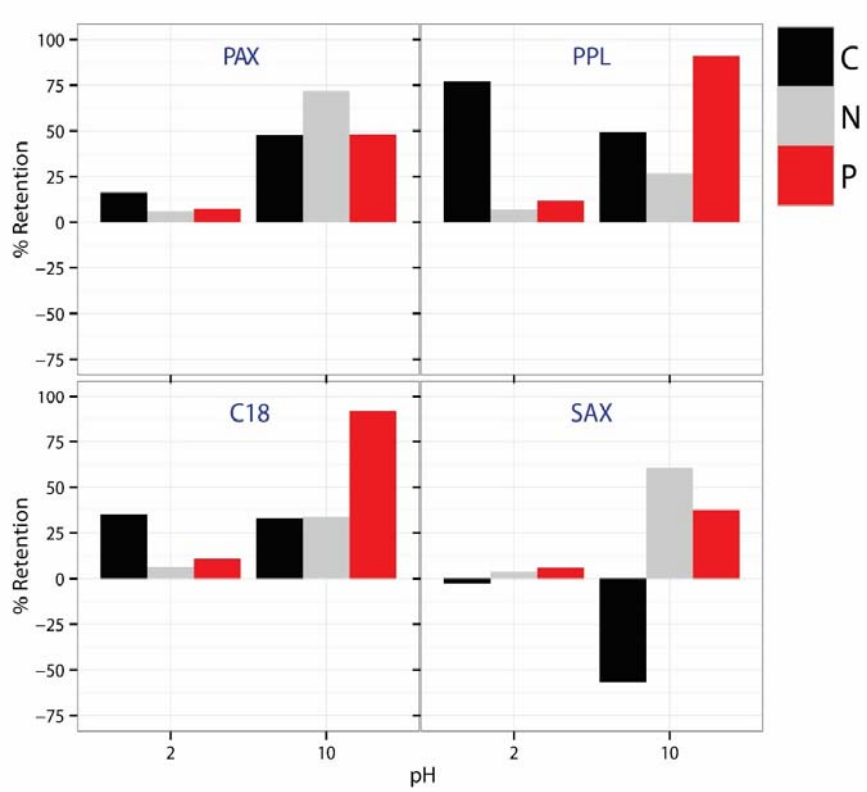


Figure S1. Wastewater primary clarifier water was used to assess the retention of dissolved organic carbon, total dissolved nitrogen, and total dissolved phosphorus by the solid phase extraction materials. Four different resins were tested: PAX, PPL, C18, and SAX. Samples of the wastewater were adjusted to pH 2 or 10 using hydrochloric acid or sodium hydroxide, respectively. The change in concentration was multiplied by the volume which was passed through the filter to estimate the % retention of these elements.

The selection of SPE materials has been principally chosen so that the resulting sample best reproduces the signature that would be observed in the original sample. Previous research has used PPL filters for its broad selectivity of carbon[7]. However, phosphorus represents a minor portion of dissolved organic matter pool. Selective concentration of organic phosphorus compounds enhances their detectability in the organic matter spectrum[8] Our objective was to determine which SPE material and methodology would best suit our needs to retain organic phosphorus compounds. The retention efficiencies of the all four SPE materials had enhanced P recovery when samples were adjusted to a pH 10 (SI Figure S1). Carbon retention displayed

some differences using this method with increased recovery for the PAX column, but a reduction in the carbon recovery for the other three columns. Most notably, the SAX column had an increased carbon concentration in the effluent, and therefore was removed from subsequent analyses. As the majority of phosphorus may have been inorganic in the primary clarifier water, it was important to demonstrate that these columns were retaining organic phosphorus compounds.

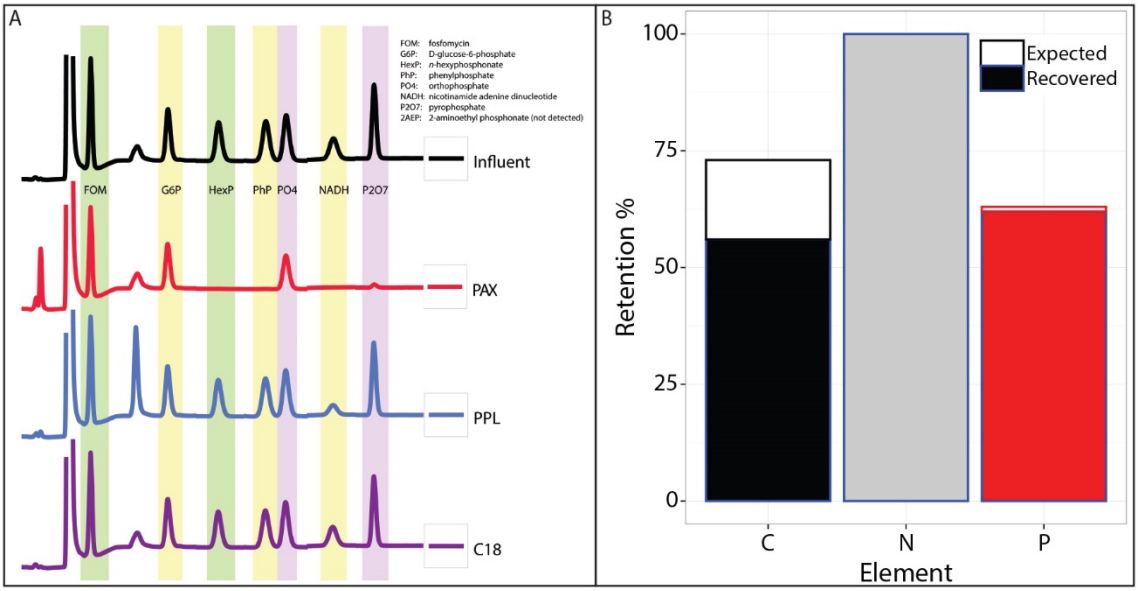


Figure S2. A standard solution consisting of equal parts phosphorus of: (inorganic) orthophosphate, pyrophosphate, (organophosphate) D-glucose-6-phosphate, phenylphosphate, NADH, (organophosphonate) fosfomycin, 2-aminoethyl phosphonate, and *n*-hexylphosphonate was prepared. The sample was basified to a pH 10 and passed through the Plexa-PAX, PPL, and C18 columns. The standard solution was read using ion chromatography before (influent) the eluent was collect from its respective column. The disappearance of a peak has been interpreted as the adsorption of that compound to the SPE column. The 2-aminoethyl phosphonate compound could not be detected using anionic IC. However, the expected retention % assuming complete recovery of 1-aminoethyl phosphonate, hexylphosphonate, phenylphosphate, NADH, and pyrophosphate by the Plexa-PAX filter indicated that this compound also adhered to this filter (e.g., 100% recovery of nitrogen).

The primary clarifier water was likely to contain minerals that could interfere with the interpretation of our results. For instance, the presence of magnesium in the water combined with the pH adjustment could lead to the precipitation of inorganic phosphates[9]. In fact, precipitates were visually observed in the samples prior to filtration. Therefore, using the laboratory phosphorus standards allowed us to detect their retention in the absence of interfering chemicals. Rather than quantifying the change in concentrations, the ion chromatographs were used to identify changes to the presence of particular compounds before and after SPE filtration (SI Figure S2A). The PAX column nearly lacked four of the compounds in its effluent chromatograph: HexP, PhP, NADH, and P2O7. These represented three organic and one inorganic compound. Notably, there was a near complete recovery of nitrogen – as determined by TDN analysis – that could indicate the recovery of the 2-AEP compound (Figure S2B). The determined recovery percent of nitrogen and phosphorus matched the results expected presuming complete recovery of 2-AEP, PhP, NADH, and P2O7. These results confirmed that the PAX column and methodology was adequate for organic phosphorus retention, and therefore this solid phase extraction resin was selected for future analyses.

The DOM of our samples ranged were composed of ≤12.8% DOP. Despite our efforts to enhance organic phosphorus recovery by using the anionic exchange SPE column, the non-manure samples were composed of less organic phosphorus than samples of Lake Superior tributaries[10]. It is noteworthy that we did not discern any retention of organic phosphorus standards by the C18 column, which had been used in the Lake Superior study[10]. Rather than retaining a greater number of phosphorus compounds, it is possible that our method simply enhanced the recovery amounts rather than isolating new compounds. ESI FT-ICR-MS does not measure concentrations so there is no valid way of determining this for our sample set[11].

Additionally, the formula algorithm also has an implicit bias against organic phosphorus in that it preferentially selects formula with the lowest non-oxygen (N+S+P) atom counts[5, 6]. For every phosphorus atom incorporated in a formula, it becomes less likely for that formula to be selected. Formula assignments are made within a 1 ppm error window, meaning that more options are available at higher molecular masses. Supporting this notion of an assignment bias, the organic phosphorus compounds were more often assigned in the lower molecular masses where there were fewer alternatives (data not shown). Our study is a rare instance in which organic phosphorus was the intended focal point of ESI FT-ICR-MS analysis. It would be useful to challenge the existing protocols if this technology is to be applied for other studies centering around organic phosphorus.

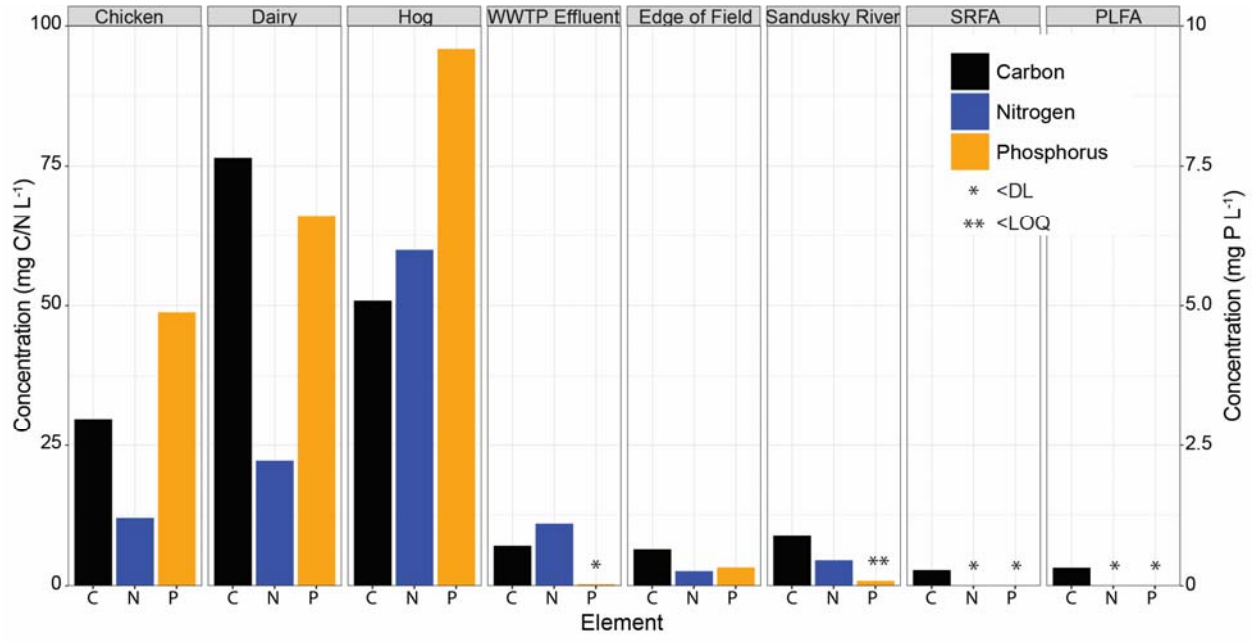


Figure S3. (A) The carbon, nitrogen and phosphorus concentrations were measured as non-purgeable carbon (NPOC), total dissolved nitrogen (TDN); and total dissolved phosphorus (ICP-OES). The detection limit (DL) for N was 0.01 mg N L^{-1}, while it was 0.03 mg P L^{-1} leading to a lower limit of quantification (LOQ) of 0.1 mg P L^{-1}. Concentrations were diluted prior to solid phase extraction.

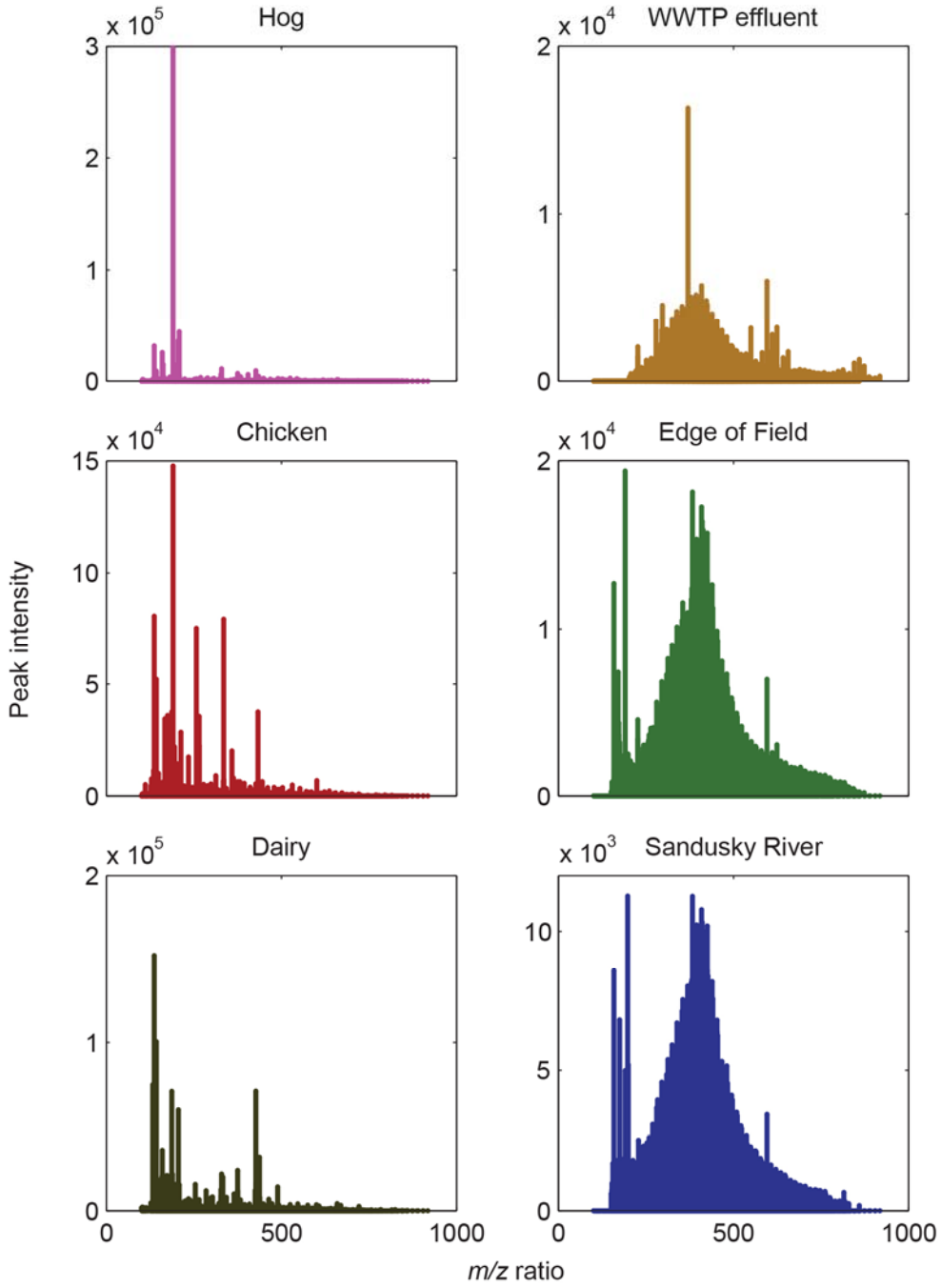


Figure S4. Negative ion mode spectra from DOM from the six watershed samples. The data have been blank-corrected and represent the average peak heights across the two replicates from each sample type.

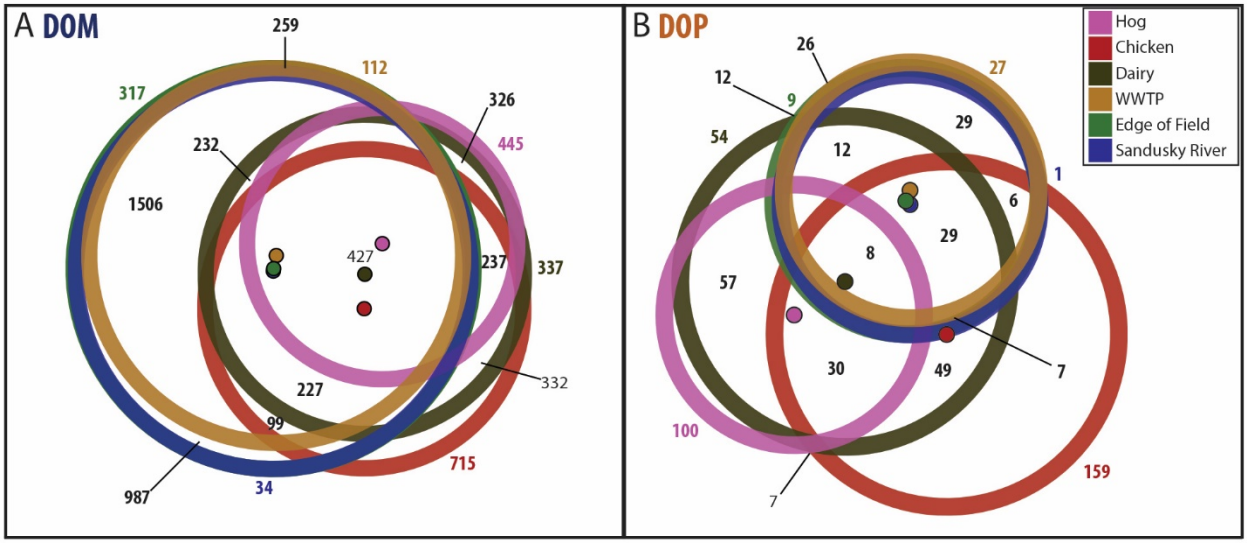


Figure S5. Sample similarity based on presence/absence data. visualized using Euler diagrams for (A) DOM and (B) DOP. The centroid is marked by a small circle with numbers indicating the number of formula shared within an intersection. Not all numbers are indicated but may be found in SI Table S4. The number of unique formula for each sample is color-coded and placed adjacent to that sample's ring.

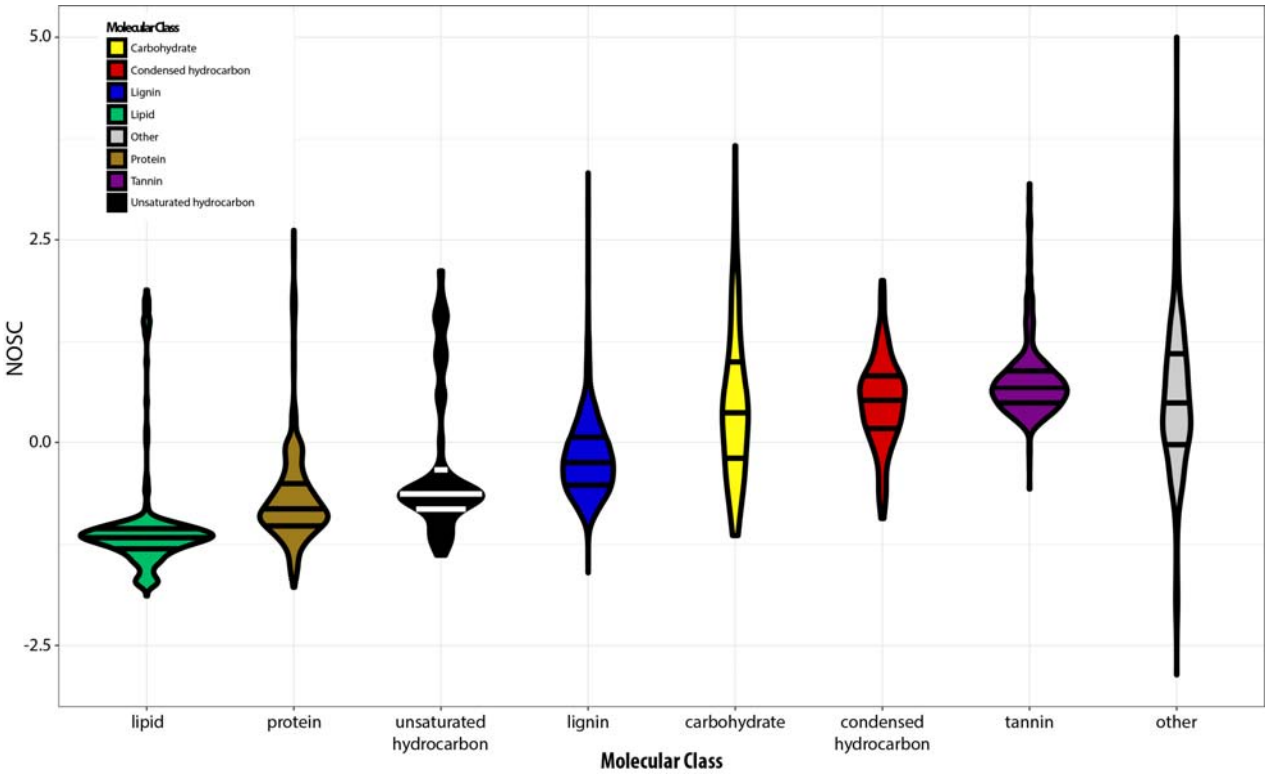


Figure S6. The nominal oxidation state of carbon (NOSC) calculated by molecular class across all samples in this dataset, as visualized in a violin plot. A negative value represents carbon in a reduced state; a positive value represents carbon in an oxidized state; and zero represents carbon which is neutrally charged.

Table S1. Adsorption efficiency across samples using the Bond Elut PAX solid phase extraction resin. Carbon was measured using non-purgeable organic carbon, while nitrogen and phosphorus were measured as the change in concentrations following sample dilution and after passing through the solid phase extraction columns. Values below the limit of quantification (10 µg N L^{-1}, 100 µg P L^{-1}) are reported as estimates. Where effluent values were above influent, values are reported as <0%. BDL indicates that the N or P concentration was below the detection limit.

Sample	Replicate	C	N	P
Chicken	replicate 1	18%	26%	<0%
	replicate 2	19%	28%	15%
Dairy	replicate 1	20%	13%	6.4%
	replicate 2	8%	31%	5.2%
Hog	replicate 1	10%	41%	<0%
	replicate 2	19%	41%	<0%
WWTP Effluent	replicate 1	44%	32%	est. 100%
	replicate 2	12%	32%	est. 97%
Edge of Field	replicate 1	21%	6%	est. 17%
	replicate 2	16%	7%	est. 9.1%
Sandusky River	replicate 1	36%	25%	est. 59%
	replicate 2	28%	33%	est. 76%
SRFA	-	41%	BDL	BDL
PLFA	-	42%	BDL	BDL

Table S2. ESI(-) FT-ICR-MS analysis detected a total of 14637 peaks, spread across the samples and replicates. The data was quality filtered by removing peaks detected in the DI procedural blank; the extraction solvent; singletons (detected in only 1 sample of the entire dataset); and peaks which had no assigned formula. The reproducibility was determined between sample replicates (shared#/mean#).

				Hog		Chicken		Dairy		Wastewater		Edge of field		Sandusky River		NOM	
Processing	Total	DI	Solvent	1	2	1	2	1	2	1	2	1	2	1	2	PLFA	SRFA
All Detected Peaks	14637	3014	534	2497	2352	3452	2493	3519	1626	2388	3602	4449	4702	4169	3154	3412	3707
Remove Peaks in Blank	11633	-	377	2094	2053	3219	2312	3215	1232	2245	3341	4338	4483	3983	3023	3262	3415
Remove Peaks in Solvent	11246	-	-	1983	1931	3098	2200	3070	1128	2171	3270	4260	4396	3895	2939	3146	3325
Removed Singleton Peaks	7438	-	-	1673	1700	2364	2072	2476	995	2070	3096	3979	4254	3853	2815	2630	3027
Assigned Formula	**7250**	**-**	**-**	**1590**	**1625**	**2315**	**2021**	**2444**	**964**	**2046**	**3071**	**3974**	**4220**	**3846**	**2798**	**2622**	**3021**
Reproducibility between sample replicates				88%		88%		68%		81%		90%		85%		-	

Table S3. ESI(-) FT-ICR-MS analysis provided peaks which were assigned formulas with C/H/O/N/P/S elements. The distribution of the *m/z* values detected in each sample were distributed across 8 formula classes. The numbers of formula are printed for each sample replicate with the number in bold indicating the total number detected in the combined samples.

		Hog		Chicken		Dairy		WWTP Effluent		Edge of field		Sandusky River		NOM	
	Total	1	2	1	2	1	2	1	2	1	2	1	2	PLFA	SRFA
CHO	**3981**	772	857	824	801	1139	475	1700	2374	3037	3190	2913	2154	**1749**	**2727**
		906		**908**		**1144**		**2413**		**3356**		**2924**			
CHON	**2198**	502	466	1047	814	923	276	239	479	732	827	769	549	**751**	**119**
		550		**1064**		**927**		**488**		**903**		**811**			
CHOP	**394**	111	99	172	150	203	112	57	126	117	117	109	75	**70**	**78**
		119		**179**		**207**		**129**		**132**		**111**			
CHOS	**254**	62	68	75	69	89	37	7	40	68	35	21	2	**38**	**74**
		72		**82**		**93**		**40**		**75**		**22**			
CHONP	**147**	50	44	81	79	30	23	9	19	7	18	12	6	**3**	**5**
		53		**83**		**35**		**19**		**18**		**13**			
CHONS	**149**	39	46	66	61	37	20	22	19	10	19	15	9	**8**	**12**
		47		**69**		**40**		**28**		**22**		**16**			
CHOPS	**62**	30	24	24	23	18	14	2	5	0	3	2	0	**1**	**2**
		30		**25**		**18**		**5**		**3**		**2**			
CHONPS	**65**	24	21	26	24	5	7	10	9	3	11	5	3	**2**	**4**
		26		**26**		**9**		**14**		**13**		**6**			
Total	**7250**	1590	1625	2315	2021	2444	964	2046	3071	3974	4220	3846	2798	**2622**	**3021**
		1803		**2436**		**2473**		**3136**		**4522**		**3905**			

Table S4. The Venn counts used to produce the Euler diagrams plotted in Figure 4. The samples columns are binary (0 not included; 1 included) with the numbers in the DOM and DOP columns indicating the number of formula for that group of samples.

Hog	Chicken	Dairy	WWTP Effluent	Edge of Field	Sandusky River	DOM	DOP
0	0	0	0	0	1	34	1
0	0	0	0	1	0	317	9
0	0	0	0	1	1	987	12
0	0	0	1	0	0	112	27
0	0	0	1	0	1	26	2
0	0	0	1	1	0	259	26
0	0	0	1	1	1	1506	29
0	0	1	0	0	0	337	54
0	0	1	0	0	1	4	2
0	0	1	0	1	0	17	0
0	0	1	0	1	1	17	1
0	0	1	1	0	0	15	3
0	0	1	1	0	1	1	0
0	0	1	1	1	0	26	1
0	0	1	1	1	1	232	12
0	1	0	0	0	0	715	159
0	1	0	0	0	1	4	0
0	1	0	0	1	0	9	0
0	1	0	0	1	1	59	4
0	1	0	1	0	0	1	1
0	1	0	1	0	1	1	0
0	1	0	1	1	0	6	1
0	1	0	1	1	1	99	6
0	1	1	0	0	0	332	49
0	1	1	0	0	1	4	1
0	1	1	0	1	0	18	0
0	1	1	0	1	1	61	7
0	1	1	1	0	0	5	4
0	1	1	1	0	1	1	0
0	1	1	1	1	0	5	0
0	1	1	1	1	1	227	29
1	0	0	0	0	0	445	100
1	0	0	0	0	1	3	1
1	0	0	0	1	0	8	1
1	0	0	0	1	1	3	0

1	0	0	1	0	0	5	1
1	0	0	1	0	1	0	0
1	0	0	1	1	0	9	3
1	0	0	1	1	1	35	8
1	0	1	0	0	0	326	57
1	0	1	0	0	1	1	0
1	0	1	0	1	0	8	0
1	0	1	0	1	1	1	0
1	0	1	1	0	0	7	1
1	0	1	1	0	1	0	0
1	0	1	1	1	0	11	0
1	0	1	1	1	1	52	4
1	1	0	0	0	0	69	7
1	1	0	0	0	1	1	0
1	1	0	0	1	0	1	1
1	1	0	0	1	1	6	0
1	1	0	1	0	0	1	0
1	1	0	1	0	1	1	0
1	1	0	1	1	0	0	0
1	1	0	1	1	1	45	0
1	1	1	0	0	0	237	30
1	1	1	0	0	1	17	1
1	1	1	0	1	0	14	0
1	1	1	0	1	1	49	4
1	1	1	1	0	0	12	1
1	1	1	1	0	1	1	0
1	1	1	1	1	0	8	0
1	1	1	1	1	1	427	8

Table S5. List of potential marker formulas found in source and Sandusky River samples. The mass to charge (*m/z*) ratios were used to identify a molecular formula. C13 indicates the presence (1) or absence (0) of a single ^{13}C isotope in the formula. The relative peak height for the *m/z* values in the samples is provided, and - signifies that the *m/z* value was not detected for that sample.

m/z	Formula	C13	Hog	Chicken	Dairy	WWTP Effluent	Edge of field	Sandusky River
432.0675772	C20H20O6NPS	-	-	-	2.14E-04	-	-	6.62E-05
464.1477252	C22H28O8NP	-	-	-	3.78E-04	-	-	8.25E-05
277.1433072	C10H23O3N4P	-	-	1.10E-04	4.24E-04	-	-	3.39E-05
276.0724817	C11H17O6P	1	2.54E-04	6.89E-04	8.44E-04	-	-	4.09E-05
408.2239451	C19H37O7P	1	-	-	-	7.22E-05	-	5.76E-05
376.0885573	C15H21O9P	1	-	-	-	-	1.13E-04	6.34E-05
420.0783553	C16H21O11P	1	-	-	-	-	9.02E-05	4.77E-05
420.1147941	C17H25O10P	1	-	-	-	-	8.38E-05	5.16E-05
430.0779524	C21H19O8P	1	-	-	-	-	9.68E-05	5.18E-05
434.0940188	C17H23O11P	1	-	-	-	-	8.76E-05	5.08E-05
502.1565882	C22H31O11P	1	-	-	-	-	8.25E-05	5.05E-05
406.135515	C17H27O9P	1	-	-	7.44E-05	-	8.54E-05	4.73E-05
332.0623426	C13H17O8P	1	-	2.05E-04	-	-	9.30E-05	4.94E-05
392.0834377	C15H21O10P	1	-	2.02E-04	-	-	7.88E-05	4.62E-05
302.0881438	C13H19O6P	1	-	2.59E-04	4.41E-04	-	8.02E-05	4.26E-05
304.0674032	C12H17O7P	1	-	5.06E-04	4.50E-04	-	7.36E-05	4.10E-05
318.0830637	C13H19O7P	1	-	5.92E-04	1.91E-04	-	8.09E-05	4.78E-05
330.0830667	C14H19O7P	1	-	2.01E-04	1.47E-04	-	1.36E-04	1.38E-04
332.0987091	C14H21O7P	1	-	2.57E-04	1.13E-04	-	9.82E-05	4.42E-05
362.1092918	C15H23O8P	1	-	6.35E-05	7.52E-05	-	9.24E-05	4.80E-05
275.0260021	C9H13O4N2PS	-	4.46E-04	1.31E-04	7.34E-04	-	4.33E-05	8.45E-05
294.0619326	C14H15O5P	1	3.39E-04	4.22E-04	6.09E-04	-	9.19E-05	1.27E-04
320.0775812	C16H17O5P	1	3.64E-04	1.78E-04	4.03E-04	-	4.06E-05	4.53E-05
421.236256	C20H39O7P	-	3.57E-04	1.97E-04	1.33E-03	-	3.64E-05	7.27E-05
372.1664215	C18H29O6P	1	-	-	-	3.03E-04	1.33E-04	1.76E-04
386.1820806	C19H31O6P	1	-	-	-	2.22E-04	9.86E-05	1.10E-04
388.1613368	C18H29O7P	1	-	-	-	2.70E-04	1.54E-04	1.69E-04
402.1769532	C19H31O7P	1	-	-	-	2.09E-04	1.05E-04	1.38E-04
403.1165379	C17H25O9P	-	-	-	-	1.00E-04	9.28E-05	4.57E-05
413.100923	C18H23O9P	-	-	-	-	1.10E-04	8.84E-05	3.93E-05
416.1926161	C20H33O7P	1	-	-	-	1.74E-04	1.04E-04	1.08E-04
417.095797	C17H23O10P	-	-	-	-	9.12E-05	4.37E-05	4.11E-05
418.0990897	C17H23O10P	1	-	-	-	7.87E-05	1.75E-04	1.74E-04
418.1718666	C19H31O8P	1	-	-	-	7.17E-05	9.78E-05	5.75E-05
429.1322172	C19H27O9P	-	-	-	-	1.18E-04	7.46E-05	3.66E-05

488.1773569	C22H33O10P	1	-	-	-	5.45E-05	8.69E-05	4.83E-05
344.1351164	C16H25O6P	1	-	-	8.28E-05	2.60E-04	1.41E-04	1.60E-04
370.1507808	C18H27O6P	1	-	-	9.10E-05	3.58E-04	1.57E-04	2.28E-04
372.1300277	C17H25O7P	1	-	-	9.49E-05	3.00E-04	2.15E-04	2.31E-04
382.1507627	C19H27O6P	1	-	-	1.11E-04	3.36E-04	2.01E-04	2.34E-04
384.166397	C19H29O6P	1	-	-	9.51E-05	2.90E-04	1.77E-04	2.08E-04
399.0852169	C17H21O9P	-	-	-	4.55E-04	7.87E-05	8.40E-05	3.89E-05
414.176966	C20H31O7P	1	-	-	7.62E-05	3.70E-04	2.13E-04	2.34E-04
458.1668009	C21H31O9P	1	-	-	7.86E-05	2.57E-04	1.63E-04	1.92E-04
484.1824486	C23H33O9P	1	-	-	7.72E-05	2.19E-04	1.65E-04	2.03E-04
360.130038	C16H25O7P	1	-	7.45E-05	-	2.13E-04	1.59E-04	1.88E-04
456.1147265	C20H25O10P	1	-	7.61E-05	-	2.90E-04	2.52E-04	2.77E-04
342.1194474	C16H23O6P	1	-	1.67E-04	1.17E-04	3.47E-04	1.65E-04	2.11E-04
344.0987134	C15H21O7P	1	-	1.87E-04	1.48E-04	8.20E-05	1.40E-04	1.73E-04
350.0881719	C17H19O6P	1	-	2.26E-04	4.02E-04	2.42E-04	1.37E-04	1.56E-04
354.0830662	C16H19O7P	1	-	3.05E-04	1.23E-04	2.32E-04	2.25E-04	2.50E-04
356.0623315	C15H17O8P	1	-	3.92E-04	1.82E-04	6.87E-05	2.34E-04	2.79E-04
356.1351082	C17H25O6P	1	-	6.45E-05	1.11E-04	2.68E-04	1.47E-04	1.73E-04
358.1143977	C16H23O7P	1	-	6.54E-05	1.13E-04	2.56E-04	2.25E-04	2.31E-04
368.0987153	C17H21O7P	1	-	2.04E-04	1.38E-04	3.00E-04	2.26E-04	2.86E-04
370.1143768	C17H23O7P	1	-	1.06E-04	1.22E-04	2.99E-04	2.25E-04	2.70E-04
380.098741	C18H21O7P	1	-	1.02E-04	3.92E-04	3.54E-04	2.54E-04	3.08E-04
382.1143594	C18H23O7P	1	-	2.00E-04	1.27E-04	3.69E-04	2.99E-04	3.32E-04
396.1300303	C19H25O7P	1	-	2.24E-04	1.46E-04	4.39E-04	3.31E-04	3.48E-04
398.1456726	C19H27O7P	1	-	2.14E-04	1.23E-04	4.31E-04	3.05E-04	3.31E-04
410.109275	C19H23O8P	1	-	1.94E-04	1.33E-04	4.50E-04	3.60E-04	4.13E-04
410.145669	C20H27O7P	1	-	2.18E-04	1.17E-04	4.47E-04	2.96E-04	3.49E-04
412.1613237	C20H29O7P	1	-	1.75E-04	1.18E-04	4.35E-04	2.96E-04	3.38E-04
430.1354991	C19H27O9P	1	-	6.31E-05	1.10E-04	2.77E-04	2.47E-04	2.50E-04
440.119842	C20H25O9P	1	-	1.73E-04	9.72E-05	3.63E-04	3.13E-04	3.41E-04
440.1562514	C21H29O8P	1	-	1.85E-04	1.13E-04	3.42E-04	2.81E-04	3.06E-04
426.067739	C18H19O10P	1	3.78E-04	-	1.20E-03	2.03E-04	2.97E-04	3.55E-04
296.0775853	C14H17O5P	1	2.26E-04	5.66E-04	4.98E-04	5.81E-05	1.02E-04	6.58E-05
322.0932496	C16H19O5P	1	5.06E-04	2.23E-04	8.40E-04	7.76E-05	1.00E-04	1.61E-04
336.1088903	C17H21O5P	1	1.91E-04	2.01E-04	1.31E-04	2.15E-04	9.13E-05	1.17E-04
366.0830953	C17H19O7P	1	6.32E-04	2.19E-04	1.42E-03	2.65E-04	1.95E-04	2.55E-04
384.0936035	C17H21O8P	1	2.21E-04	3.47E-04	8.02E-04	2.89E-04	3.47E-04	3.64E-04

References

1. Brooker, M.R.; Bohrer, G.; Mouser, P.J. Variations in potential CH_4 flux and CO_2 respiration from freshwater wetland sediments that differ by microsite location, depth and temperature. *Ecological Engineering* **2014,** *72* (The Olentangy River Wetland Research Park: Two Decades of Research on Ecosystem Services), 84-94; http://dx.doi.org/10.1016/j.ecoleng.2014.05.028.

2. Southam, A.D.; Payne, T.G.; Cooper, H.J.; Arvanitis, T.N.; Viant, M.R. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Anal. Chem.* **2007,** *79* (12), 4595-4602; http://dx.doi.org/10.1021/ac062446p.

3. Bhatia, M.P.; Das, S.B.; Longnecker, K.; Charette, M.A.; Kujawinski, E.B. Molecular characterization of dissolved organic matter associated with the Greenland ice sheet. *Geochim. Cosmochim. Acta* **2010,** *74* (13), 3768-3784; http://dx.doi.org/10.1016/j.gca.2010.03.035.

4. Mantini, D.; Petrucci, F.; Pieragostino, D.; Del Boccio, P.; Di Nicola, M.; Di Ilio, C.; Federici, G.; Sacchetta, P.; Comani, S.; Urbani, A. LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* **2007,** *8* (1), 1; http://dx.doi.org/10.1186/1471-2105-8-101.

5. Kujawinski, E.B.; Behn, M.D. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **2006,** *78* (13), 4363-4373; https://dx.doi.org/10.1021/ac0600306.

6. Kujawinski, E.B.; Longnecker, K.; Blough, N.V.; Del Vecchio, R.; Finlay, L.; Kitner, J.B.; Giovannoni, S.J. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochim. Cosmochim. Acta* **2009,** *73* (15), 4384-4399; https://dx.doi.org/10.1016/j.gca.2009.04.033.

7. Ohno, T.; Ohno, P.E. Influence of heteroatom pre-selection on the molecular formula assignment of soil organic matter components determined by ultrahigh resolution mass spectrometry. *Analytical and Bioanalytical Chemistry* **2013,** *405* (10), 3299-3306; https://doi.org/10.1007/s00216-013-6734-3.

8. Cooper, W.T.; Llewelyn, J.M.; Bennett, G.L.; Salters, V.J.M. Mass spectrometry of natural organic phosphorus. *Talanta* **2005,** *66* (2), 348-358; http://dx.doi.org/10.1016/j.talanta.2004.12.028.

9. Karl, David M.,Tien, Georgia, MAGIC: A sensitive and precise method for measuring dissolved phosphorus in aquatic environments. *LNO Limnology and Oceanography* **1992,** *37* (1), 105-116; https://doi.org/10.4319/lo.1992.37.1.0105.

10. Minor, E.C.; Steinbring, C.J.; Longnecker, K.; Kujawinski, E.B. Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. *Org. Geochem.* **2012,** *43*, 1-11; https://dx.doi.org/10.1016/j.orggeochem.2011.11.007.

11. Kamga, A.W.; Behar, F.; Hatcher, P.G. Quantitative Analysis of Long Chain Fatty Acids Present in a Type I Kerogen Using Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Compared with BF_3/MeOH Methylation/GC-FID. *Journal of The American Society for Mass Spectrometry* **2014,** *25* (5), 880-890; http://dx.doi.org/10.1007/s13361-014-0851-x.