Abstract

To further comprehend the soil's biogeochemical processes, it is essential to understand the interactions between the soil DOM (dissolved organic matter) chemodiversity and the microbes habiting in it (soil)~~Understanding the dissolved organic matter (DOM) chemodiversity in soil and its interaction with microbes is crucial to comprehend the soil's biogeochemical processes~~. ~~In this study,~~ FT-ICR-MS ~~u~~(ultrahigh-resolution fourier transform ion cyclotron resonance mass spectrometry ~~(FT-ICR-MS~~) was employed to in detail characterize the DOM chemodiversity at a molecular level. ~~Based on four agro-ecological experimental sites, t~~The variation of the DOM chemodiversity ~~variation in~~ from four agro-ecological experimental sites (paddy ~~soils~~fields) subjected to different long-term fertilizations across different distance-points was evaluated. Geographic distance had a greater impact on DOM chemodiversity than what ~~anthropogenic~~ fertilization had. Distance-decay analysis showed that the ~~dissimilarity~~ differences ~~of~~ in the DOM chemodiversity significantly increased with the increase in geographic distance. Long-term organic fertilizations homogenized the DOM chemodiversity of the soils ~~as it made~~with the lipid-like compounds in them becoming more similar regardless ~~of~~ geography. Combining high throughput sequencing ~~and~~ with FT-ICR-MS, we developed a network analysis to visualize the significant interactions between bacterial species and DOM molecules. Our results showed that the bacterial community diversity dictates the soil's DOM chemodiversity and not vice versa;~~. G~~ also, geographic distance indirectly affects the soil's DOM chemodiversity by shaping the bacterial community habiting in it.

Keywords: dissolved organic matter; chemodiversity; bacterial community; geographic distance; FT-ICR-MS

1. Introduction

Dissolved organic matter (DOM) is ~~ubiquitous~~ abundant in all terrestrial ecosystems and plays a vital role in soil's biogeochemical processes, most significantly, in carbon cycle (Khan et al., 2013). Although soil DOM accounts for only a minor part of soil organic matter (SOM)~~,~~ it is the most active and bioavailable organic 'pool' in soil. DOM is of high geochemical and environmental importance for both soil and adjacent water systems (Wu et al., 2020), especially in paddy fields. The concentration of DOM could be easily ~~represented~~~~enumerated~~defined, however, ~~evaluating~~ the evaluation of the DOM's chemical composition remains a challenge due to its extreme heterogeneity. Chemodiversity represents the material condition for life to emerge and exist (Testa et al., 2009). Better characterization of DOM's chemical diversity could help us understand the biogeochemical processes in soil. Thus, ~~grasping~~ understading the DOM chemodiversity ~~of DOM~~ has become a research 'hotspot' in the recent years (Lehmann and Kleber, 2015; Roth et al., 2019; Zherebker et al., 2020).

~~A~~ R~~r~~ecent~~ly~~ advance~~s in~~d mass spectrometry and data analysis ~~method~~, i.e. ultrahigh-resolution fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), ~~is~~ are often employed to in detail characterize the DOM chemical composition at a molecular level (Koch et al., 2007; Nebbioso and Piccolo, 2013). ~~For~~ Up to now, this technique (FT-ICR-MS) is probably the most suitable approach to obtain detailed molecular formulas for naturally complex and heterogeneous media/matter, such as DOM. The approach has been successfully trialed to investigate the DOM chemodiversity ~~in~~ from marine and lake samples ~~abstracted~~ collected from aquatic

environments (Kellerman et al., 2014; Kujawinski et al., 2016; Stegen et al., 2016; Zhao et al., 2019). Studies focusing at soil DOM have also been ~~presented~~ recently carried out ~~in recent years~~ (Ding et al., 2020; Li et al., 2019; Ohno et al., 2016; Roth et al., 2019); however, the overall understanding ~~about~~ of the chemodiversity of such substrates ~~is still~~remains limited, especially for systems exposed to a plethora of anthrpogenic activities (i.e. in paddy fields).

DOM is generated from root exudates and plant material leachates, and is typically consumed by soil microbes. This is then transformed into microbial-derived compounds which may contain intrinsically recalcitrant molecules (Roth et al., 2019). On one hand, DOM acts as a major source of carbon and nutrients, especially for soil microbial communities; on the other hand, soil microbial metabolites and residues act as a major source for DOM. How microbes contribute to the formation and the composition of DOM, and what their interactions are, are the two fundamental questions related to soil carbon dynamics that remain ~~unclear~~unanswered. Until now, there are only limited research studies that explore the link between the ~~variations of~~ microbial ~~communities~~ community variations and the changes in the soil's DOM chemodiversity (Li et al., 2018b, 2019). Similarly, ~~, to demonstrate~~ unanswered remains the question of whether or not ~~that~~ these two can be significantly correlated, as well as~~.~~ ~~However~~, which of these two has a higher impact over the other. ~~is not yet clear.~~

Paddy fields are the largest anthropogenic wetlands on earth (Kögel-Knabner et al., 2010). They play a vital role in global carbon cycle as they ~~were~~ are identified as one of the major ~~sources of~~ greenhouse gas~~es~~ sources (Khan et al., 2013). Thus, studying

and understanding the DOM chemodiversity ~~of DOM~~ in paddy soils is crucial to further comprehend ~~of~~ the soil's biogeochemical processes and subsequently the potential carbon cycle implications. China has 30.3 Mha of paddy soil which is mainly distributed among the subtropical zone of the country (Li et al., 2018b). The differently long-term operated paddy fields in this area provide us with an ideal model to study the spatial variation of DOM chemodiversity ~~in paddy soils~~. Previous work has showed that geographic location ~~dominated~~ is a key influential factor towards the compositional variation of the bacterial community, ~~which demonstrated~~the community becomes less ~~similarities~~ similar when geographic distance ~~increased~~ increases at a regional scale~~; m~~. Meanwhile fertilization ~~was found to~~ decreases the spatial variation of the bacterial diversity in paddy soils (Chen et al., 2017). Whether the DOM chemodiversity displays similar variation patterns is yet unknown. Based on the affinity between soil microbes and DOM, we hypothesize that the chemodiversity of paddy soil DOM could show similar spatial variation patterns and could also be significantly impacted by anthropogenic farming practices. We also hypothesize that the soil DOM chemodiversity ~~of soil DOM~~ is regulated by the developed microbial community. To test these hypotheses, we ~~illustrate~~ focused on the detailed structural features of the DOM (at a molecular level) ~~in~~ from paddy soils subjected to different long-term fertilizations, and describe the corresponding chemodiversity variations across a geographic distance. The relationship between DOM chemodiversity and soil microbial community ~~is~~ was analyzed at molecule and species level respectively. How the microbial community shapes the DOM chemodiversity ~~is~~ was also ~~reasoned~~ analysed

and discussed.

2. Materials and Methods

2.1. Field experiment and soil sampling

~~The study's~~The experimental ~~area~~ site is located in subtropical China and ~~contains~~ includes four agro-ecological experimental sub-sites (Changshu–CS, Jinxian–JX, Wangcheng–WC, and Yingtan–YT) (Fig. S1). ~~Straight line~~The distance between the experimental sub-sites ~~is~~ ranges between 400–1350 km. Detailed description of the experimental sites is shown in Table S1 including mean annual temperature (MAT), mean annual precipitation (MAP), location and climate information. Each site was subjected to similar long-term fertilization treatments, among which, three typical treatments were selected for investigation: i) CK (no fertilizers), ii) NPK (mineral nitrogen, phosphorus and potassium fertilizers), and iii) OM (NPK fertilizers plus organic amendments). Details about the treatments including fertilizer application rates are also listed in Table S1.

A total of 36 samples were collected with three replicates for each treatment from each site after rice harvest (4 sites × 3 treatments × 3 replicates). Five cores were taken from the topsoil (0–15 cm) of each treatment replicate, and then ~~were~~ homogenized to form one uniform soil sample. The uniformed samples were sieved (<2 mm) and were separated into two sub-samples; one sub-sample was stored at -40 °C for DNA extraction, another subsample was air-dried for analysis of chemical properties under standard procedures (Carter and Gregorich, 2007; Lu, 1999). Soil pH was measured by

a pH meter (Mettler-Toledo, ZRH) using a 1:2.5 soil:water suspension. SOM content was determined by the $K_2Cr_2O_7 + H_2SO_4$ oxidation method. Total and available nitrogen (TN and AN) were quantified using Kjeldahl-N method and the alkali hydrolysable method, respectively. Total phosphorus and potassium (TP and TK) were measured by $HF–HClO_4$ digestion and determined by molybdenum blue colorimetric and flame photometry respectively. Available phosphorus (AP) was determined by the Olsen method. Available K (AK) was measured by NH_4OAC extraction. These chemical properties are described in Table S2.

2.2. DOM extraction and FT-ICR-MS analysis

The DOM was extracted from the soil samples (10 g) using ultrapure water ~~by~~ using a soil:water ratio to 1:10, ~~and~~ the samples were then shaken for 12 h in a horizontal shaker at room temperature. The solutions were centrifuged at 2800*g* for 20 min and filtered through a 0.45-μm membrane filter. HPLC methanol (10 mL) and acidified ultrapure water (10 mL, pH = 2) were passed through PPL cartridges (Agilent Technologies, Santa Clara, CA, USA) for clean-up. Then the DOM solution was loaded on to the PPL cartridges via gravity. After that, the DOM was collected from the cartridges using 10 mL of methanol (HPLC grade; Merck, Germany). The DOM elutes were kept at -20 ℃ in dark prior an electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS) measurement.

Deuterated octadecanoic acid was added to the samples as an internal standard with a dosage of 15 μL (5×10^{-7} mol/L) per milliliter of sample. The ESI FT-ICR MS (Bruker, Billerica, MA, USA) was equipped with a 9.4 T actively shielded superconducting

magnet interfaced with negative-ion mode electrospray ionization. Each sample was injected into the ESI source at a speed of 180 μL/h using a syringe pump. The polarization voltage was 4.0 kV. The capillary column introduction and outlet voltage were 4.5 kV and 320 V, respectively. Ions were accumulated in the hexapole for 0.001 s before being transferred to the ICR cell. The m/z range was 150–800 Da. A 4M word size was selected for the time domain signal acquisition. The signal to noise ratio and dynamic range were enhanced through accumulating 128 times domain FT-ICR transients.

Data Analysis software (Bruker Daltonics version 4.2) was used to convert raw spectra to a list of m/z values using FTMS peak picker (S/N threshold of 6; absolute intensity threshold of 100). To reduce cumulative errors, all sample peaks from the entire dataset were aligned to each other to eliminate possible mass shift. Molecular formulas of mass peaks were calculated using a custom software. A van Krevelen diagram was used to cross-plot oxygen to carbon (O/C, x-axis) and hydrogen to carbon (H/C, y-axis) element ratio. The thousands of formulas plotted in the van Krevelen diagram ~~could be~~were categorized into the following groups based on O/C and H/C ratios (Li et al., 2018b; Qiu et al., 2019): (1) lipid-like compounds (O/C: 0–0.3, H/C: 1.5–2.0), (2) protein/amino sugar-like compounds (O/C: 0.3–0.67, H/C: 1.5–2.2), (3) carbohydrate-like compounds (O/C: 0.67–1.2, H/C:1.5–2.4), (4) unsaturated hydrocarbon compounds (O/C: 0–0.1, H/C: 0.7–1), (5) lignin-like compounds (O/C: 0.1–0.67, H/C: 0.7–1.5), (6) tannin-like compounds (O/C: 0.67–1.2, H/C: 0.5–1.5), (7) condensed aromatic-like compounds (O/C: 0–0.67, H/C: 0.2–0.7), (8) others (the

157 compounds that do not belong to the above 7 groups). The peaks that were taken into
158 consideration for further analysis were those that ~~were~~ observed at least twice among
159 the 36 samples. The unsaturated hydrocarbon compounds were limited and not detected
160 in our samples; hence, our compounds were divided into seven groups and presented
161 on a van Krevelen diagram (Fig. 1).

162 *2.3. Statistical Analysis*

163 The chemodiversity of DOM was analyzed based on the differences between the
164 compounds (one molecular formula in FT-ICR-MS refer to one compound) in each
165 sample. The relative abundance (%) of the van Krevelen diagram-derived compounds'
166 groups, as per FT-ICR-MS analysis for the 36 soil samples is shown in Fig. S1.

167 Two-way permutational multivariate analysis of variance (PERMANOVA) based on
168 Euclidean distance using 9999 permutations (Anderson, 2001) was used to ~~show~~
169 demonstrate the dissimilarity ~~of~~ between molecular formulas ~~in the samples among~~
170 from different fertilization treatments and experimental sites (Table 1). PERMANOVA
171 analysis was also conducted between any two fertilization treatments or any two
172 experimental sites. The rate of distance-decay (Deng et al., 2016) of the DOM
173 chemodiversity was calculated as per the slope of a linear least squares regression,
174 performed between molecular formulas similarity (calculated by the Euclidean distance)
175 and geographic distance. The geographic distance was calculated based on the
176 latitudinal and longitudinal coordinates (Table S1). The distance-decay plots were
177 conducted based on all compounds and based on each van Krevelen diagram-derived
178 compounds' group (Fig. 2). The significant difference of slopes between any two

treatments was tested via univariate linear regression analysis (Table S3).

Variable partitioning analysis (VPA) was applied to analyze the contribution of the environmental ~~factors~~and~~,~~ spatial factors, ~~and~~ the bacterial community and their interactions on the variance of DOM chemodiversity. Spatial variables were derived from Principal Coordinates of Neighbor Matrices (PCNM), ~~which was~~a method able to ~~deconvolute~~ deconstruct total spatial variation into a discrete set of explanatory spatial scales. Forward selection procedures were subsequently used to select the respective subsets of the environmental and spatial variables. The forward selection was stopped if the significance level ($P < 0.05$) was reached, or if no improvement of the selection criterion (R^2) ~~was~~ observed after adding more variables. Then, the contributions of selected subsets to the variance of DOM chemodiversity were calculated using the 'varpart' function in R package 'vegan'. The fractions of the explained variance are based on adjusted fractions (R2adj, adjusted coefficient of multiple determination), which accounts for the number of variables and sample sizes. The significance of each component via partitioning was evaluated with a permutation test, except for the interaction term and the residuals (these cannot be tested statistically). The proportional contributions of the environmental factors, spatial factors, DOM chemodiversity, and their interactions on bacterial community were also analyzed following the same procedure.

The high-throughput sequencing results of the bacterial 16S rRNA genes have been reported ~~in~~ at our previous work (Chen et al., 2017). In ~~this~~ the current work presented here, we only use the operational taxonomic units (OTUs) of the gene sequences gained

from the long-term fertilization experiment ~~in~~ of the four agro-ecological experimental sites for further analysis. The specific links between DOM chemodiversity and relative abundance of OTUs were revealed by network analysis. Pairwise correlations were calculated using the 'WGCNA' package to determine the relationships between individual DOM molecule and bacterial OTUs using Spearman's correlations (Langfelder and Horvath, 2008). A credible interaction was considered a robust correlation between DOM chemodiversity and relative abundance of OTUs ~~with~~ when the Spearman's correlation coefficients were above 0.8 and BH-corrected *P*-values below 0.05. Networks were displayed with Spring Embedded Layout and yFiles Organic Layout method and analyzed with network analyzer in Cytoscape.

3. Results

3.1. Difference in DOM chemodiversity

The unique molecular formulae ~~determined~~ provided by the FT-ICR-MS were used to evaluate the chemodiversity of DOM. ~~For~~ From all 36 DOM samples, a core group of 7,949 molecules, with a molecular mass ranging from 182 to 716 Da, ~~had a molecular formulae~~ observed in at least two of the samples~~, having a molecular mass ranging from 182 to 716 Da~~. A van Krevelen diagram (Fig. 1) was used to visualize the compounds distribution. According to the van Krevelen diagram, carbohydrate-like (CH), condensed aromatic-like (CA), lignin-like (LG), lipid-like (LP), protein/amino sugar-like (PS), tannin-like (TN), others compounds (OT) were detected. For these, LG and CA compounds were the most abundant molecules in the paddy soils DOM ~~originated~~

~~from the long-term cultivated paddy soils and~~ accounting for 38.5−46.8% and 17.6−25.9% of all assigned molecules respectively (Fig. S1). The relative abundance of the DOM's biomolecular compound groups were found significantly different among treatments (Fig. S1) (ANOSIM analysis: $R = 0.289$, $P < 0.001$).

A Venn plot (Fig. 2A) showed that CK, NPK and OM ~~shared~~ had 5,934 ~~compounds~~ from the overall 7,949 ~~DOM molecules~~ DOM compounds in common. Although the DOM samples were originated from different sites, the majority of the compounds detected ~~compounds~~ between them ~~was~~ were similar, observation that further supports the similarity and the chemical stability of the DOM ~~in~~ from paddy soils. There were relatively few unique compounds for each treatment, specifically 166, 155, and 105 for CK, NPK, and OM, respectively. Although DOM similarities between soils were evident (with regards stability in the chemical structure), further understanding of the significant differences in the DOM chemodiversity among different treatments is essential to further evaluate ~~the~~ its structure variation. The unique compounds in CK were mainly CA ~~compounds~~, while in NPK the unique compounds were mainly LG and TN (Fig. 2E-F). The unique compounds in OM were mainly concentrated in LG (Fig. 2G), with almost no CA and TN, which is the main difference between OM and CK/NPK treatments. Fig. 2B-D showed that CK and NPK ~~shared~~ had 88% of their total molecules in common, ~~OM shared~~ 80% and 81% of the molecules ~~with~~ between OM and CK, ~~and~~ NPK were also similar ~~respectively~~. That also meant that CK and NPK were more similar with each other, while OM showed less similarity with CK and NPK. OM also had fewer unique compounds as compared with CK and NPK (Fig. 2C-D).

3.2. Comparison of DOM chemodiversity across geographic distances

A PERMANOVA test was used to evaluate whether geographic location or fertilization ~~exhibits~~ has a higher impact on DOM chemodiversity (Table 1). The global test indicated that DOM chemodiversity showed significant dissimilarity ($P < 0.01$) among fertilization treatments and experimental sites, with the latter having a greater impact on the DOM chemodiversity (higher *F* value as compared to fertlization). Pairwise comparison results showed that DOM chemodiversity was significantly different for any two sites except for the comparison between WC and JX. DOM chemodiversity also showed significant dissimilarity ($P < 0.01$) between organic fertilization and other treatments (OM *vs* CK and OM *vs* NPK); while chemical fertilization did not significantly ($P > 0.05$) changed the DOM chemodiversity as compared to the ~~no~~ un-~~fertilization~~ fertilized treatments (CK *vs* NPK).

A distance-decay analysis was used to study the spatial variation of the DOM chemodiversity. The distance-decay plot (Fig. 3A) showed that the dissimilarity of the DOM chemodiversity for all 36 samples significantly ($P < 0.001$, Table S3) increased along with the increase ~~of~~ in geographic distance. Specially, the detected biomolecular compound groups in all ~~the~~ samples showed the same tendency except for ~~the~~ OT ~~compounds~~ (Fig. 3B-H, Table S3), in which LG compounds ~~showed~~ demonstrated a bigger distance-decay slope than any other compound groups (Table S3).

The distance-decay analysis indicated that geographic location significantly affected the DOM chemodiversity, while anthropogenic fertilization also had a great impact to it. In each fertilization treatment, the entire DOM chemodiversity ~~was found~~

significantly increased with the increase in geographic distance (Fig. 3A). The distance-decay slope was 0.005 for the OM treatment (Table S3), significantly smoother than that of the CK (0.013) and NPK (0.012) treatments, while the slopes of CK and NPK were ~~in~~ at the same level ($P > 0.05$). At different treatments, biomolecular compound groups varied with geographic distance. For CK treatment, the chemodiversity dissimilarities of CA, LG, LP, PS, and TN compounds significantly ($P < 0.05$, Table S3) increased with geographic distance. For NPK treatment, the chemodiversity dissimilarities of CH, LG, LP, and PS compounds significantly increased with geographic distance. For OM treatment, the chemodiversity dissimilarities of LG, PS, and OT compounds significantly increased with geographic distance. From the plots (Fig. 3B-H), it is indicated that distance-decay slopes for CA, LG, LP, and TN compounds at the OM treatment were significantly flattened as compared ~~with~~ to CK. The distance-decay slopes of the LP and OT compounds ~~of~~ from the OM treatment were significantly flattened as compared with NPK (Table S3). In OM treatment, chemodiversity of LP compounds did not significantly change with geographic distance, ~~and~~ but significantly differed ~~from that~~ in CK and NPK treatments due to that.

3.3. Co-variation between bacterial community and DOM chemodiversity

As per the ~~inherent~~intrinsic association ~~link~~ between soil microbes and DOM heterogeneity, we believe the variation of DOM chemodiversity is ~~associated~~ linked with the bacterial community. This assumption was confirmed by spearman's analysis, which highlighted that the dissimilarity between DOM molecules and microbial communities between samples ~~were~~ was significantly correlated ($r = 0.297$, $P < 0.001$).

The mutual contributions between DOM chemodiversity and the bacterial community diversity were further explored using variable partitioning analysis (VPA). VPA results suggested that the bacterial community imposes a considerable factor (17.1%) affecting DOM chemodiversity ($P < 0.01$), while environmental factors and geographical location could not significantly affect it (Fig. 4a). The interaction between bacterial community and geographical location also explains, at a considerable degree (14.7%), the DOM chemodiversity. On the contrary, environmental factors and geographical location significantly affected ($P < 0.05$) the bacterial community variation (Fig. 4b), while the DOM chemodiversity could not itself shape the community (1.1%, $P > 0.05$). The combined effect of ~~the~~ factors like the environmental conditions, the geographical location, and the DOM chemodiversity had a large impact (67.3%) on the bacterial community variation.

The link between DOM chemodiversity and bacterial community was further explored using co-occurrence network analysis. The results revealed a significant association between DOM molecules and specific taxa. Direct connections were made from the selected square nodes (OTUs) to the selected circle nodes (molecule) as edges in the networks (Fig. 5). The strong positive ($r \geq 0.8$) and negative correlations ($r \leq -0.8$) were defined as the positive and negative networks respectively. A total of 3016 nodes and 6978 edges formed the positive networks (Fig. 5A-H), while ~~for~~ only 19 nodes and 15 edges ~~the negative networks there were~~compiled the negative ones ~~only 19 nodes and 15 edges~~ (Fig. 5I). The small subnetworks, which were hard to define as OTU-central or molecule-central, and which contained less than six total nodes were not

shown in the figure. The positive networks consisted of complex networks that could be divided into 8 subnetworks, from which ~~there was~~ one was considered exceptionally large and complex ~~subnetwork~~ with 2521 nodes and 6539 edges (Fig. 5A); ~~and other~~ seven less complex ones were detected (Fig. 5B-D). In the positive network, among the top ten nodes in connection number, eight were OTU-central and only two were molecule-central. The negative networks were much simpler, with only one OTU (Acidobacteria) ~~displaying~~ bearing mutiple links with 11 molecules, which refer to LG, LP, and PS groups (Fig. 5E). This indicates that these species may mainly utilize these three kinds of compound groups.

The 1000 most abundant OTUs and the 1000 most abundant DOM molecules in each fertilization treatment were selected for network analysis. The results showed that OTUs and DOM molecules were significantly connected in each treatment~~,~~ indicating the strong soil bacteria-DOM interactions (Fig. S2). A total of 114 nodes and 442 edges formed the networks in CK treatment;~~.~~ ~~There were~~ 138 nodes and 673 edges ~~forming~~ formed the networks in NPK treatment~~.~~ , ~~While~~ while in OM treatment, the networks contained only 50 nodes and 115 edges. That meant that the bacteria−DOM interaction was stronger in CK and NPK than that for the OM treatment.

4. Discussion

4.1. The variation of DOM chemodiversity among different fertilization treatments

Among the molecular formulas gained from the FT-ICR-MS, LG and CA compounds (H/C < 1.5) ~~are~~ appeared to be the two predominant compounds, both typically recalcitrant ~~components~~ (Li et al., 2019). The consistent predominance of

these two ~~compound~~ groups at all samples indicates the chemical stability of the DOM at ~~such~~ all operational conditions. Lignin is the~~a~~ main ~~chemical~~ raw ~~material~~ matter ~~for~~ in the DOM ~~in~~ originated from paddy soils~~,~~ known for its limited biodegradability. The persistent to degradation nature of lignin-based products obviously allowed its accumulation and subsequently provided had ~~with~~ a stronger 'signal' during the test. Hence, LG compounds accounted for the maximum proportion among all assigned molecules (Fig. S1). Although the ~~DOM~~ samples were originated from different sites, ~~different fertilization treatments had~~ 75% of the detected DOM molecules was present~~in common~~ at all treatments, observation that also ~~indicated~~ indicates the ~~similarity and~~ chemical stability of the DOM in paddy soils. Although the unique molecules, among different treatments were relatively ~~less~~low in absolute numbers, they ~~contributed~~ were adequate to significantly differentiate~~to the significant differences of~~ the relative abundance of the DOM's biomolecular compound groups ~~among different treatments~~. The unique compounds found in OM treatment were ~~concentrated~~ mainly related to the~~in~~ LG group, ~~which then dictated the~~fact directly linked to the ~~dissimilarity~~ differences of the DOM chemodiversity between OM ~~treatment~~ and other treatments. In OM treatment, the unique LG compounds may ~~be~~ derive from the ~~extra~~ excess of organic substances, such as straw, provid~~ed~~ing with more ligin materials.

The distance-decay analysis showed that the DOM chemodiversity ~~was~~ significantly increased along with the increase in geographic distance. LG compounds demonstrated greater variability with geographical distance as compare~~d~~ to CA, CH, LP, PS, TN, and OT compounds. In the meantime, LG ~~compounds~~ accounted for the maximum

proportion of all the compounds, hence, LG imposed the greatest contribution to the spatial variability of the overall DOM. In each fertilization treatment, the entire DOM chemodiversity was also significantly increased along with an increase in geographic distance ~~increase~~. However, the compound groups that significantly changed with spatial variation in each treatment were different~~;~~, ~~indicating~~ this indicates the significant effect~~s~~ of the anthropogenic fertilization (with geographic location also having a great impact). Our observation~~s~~ ~~indicated~~ highlights that long-term organic fertilizations result~~ed~~ in the homogenization of DOM chemodiversity in paddy soils across geographic distance on an LP~~--~~compounds~~--~~presence basis. Many of the LP compounds derived from bacteria (Zou et al., 2004), for example, the phospholipids in bacterial cell membrane can contribute to the LP compounds. Thus, we ~~deduced~~presumed that long-term organic fertilizations in paddy soils firstly homogenize the bacterial community across geographic distance (Chen et al., 2017), making the bacteria~~l~~-derived LP compounds more similar, ~~and this way drive~~resulting to the homogenization of the DOM chemodiversity. In summary, LG compounds mainly dictated the distance-decay pattern of the DOM chemodiversity, whilst LP compounds contributed to the difference ~~of~~ in the pattern among different fertilization treatments.

4.2. Co-variation between bacterial community and DOM chemodiversity

Co-occurrence network analysis showed that DOM chemodiversity and bacterial community were significantly correlated, finding consistent with previous research (Li et al., 2019). More interestingly, VPA analysis indicated that the DOM chemodiversity

variation at a regional scale was best explained by the developed bacterial community but not vice versa. In a similar research, paddy soils were sampled at the tillering phase of rice plants, concluding that the variance in the microbial community and the variance in DOM composition were best explained by each other (Li et al., 2018a). At a tillering sampling scenario, both the DOM and the soil microbial community change rapidly due to the root exudates present that contribute to the DOM, and thus, significantly affect the microbial community (the reverse is also likely as microbes metabolize organics to stand for DOM). On the contrary we decided to sample the paddy soils after the harvest. No root exudates or other organics could directly contribute to the DOM, and the latter was at a stable phase, unlikely to significantly change the microbial community. In this way we excluded a ~~key~~ fundamental experimental bias. On the other hand, the microbial metabolic processes ~~that~~ related to the consumptione, transformation, and the ~~or~~ productione of ~~different~~ various DOM molecules, mainly contributed to the DOM turnover. A previous study has shown that the chemical diversity of SOM was mainly driven by distinct microbial communities rather than by the substrates utilized by the microbes (Kallenbach et al., 2016), ~~which~~ finding that also supports the current study's VPA analysis.

The ecological theory "*diversity begets diversity*" is commonly used ~~in terms~~quote ~~of~~ in biodiversity (Palmer and Maurer, 1997). Tanentzap et al. (2019) extends the theory stating that greater microbial diversity results to greater DOM chemodiversity, and greater DOM chemodiversity provides with more opportunities for the microbes to coexist. The diversity used on the above study was an α-diversity of DOM molecules

and microbes (Tanentzap et al., 2019); thus, it was found that chemodiversity begets biodiversity and vice versa, but at α-diversity level only (samples originated from a natural freshwater ecosystem). At β-diversity level (explored in the current study), we found that a greater bacterial community diversity leads to greater DOM chemodiversity (in a paddy field ecosystem), finding also in line with the "*diversity begets diversity*" theory. Differently, Tanentzap et al. found that DOM had a stronger effect on microbes than vice versa (Tanentzap et al., 2019), while we found that DOM chemodiversity could not significantly shape the bacterial diversity. The reason may be that the influential factors at Tanentzap's study were less in absolute number as compared to those in this study. Hence, with DOM being the main carbon source for microbes it is expected that DOM would influence the microbial community (in two small shallow lakes) (Tanentzap et al., 2019). At our study, the influential factors for the bacterial community were much more complex and DOM was not the only carbon substrate for the bacteria. Thus, we shout that DOM chemodiversity alone cannot significantly shape the bacterial community. On the contrary, the bacterial metabolites and their own residues can also become the main sources of DOM in a long-term fertilized paddy soils. Thus, DOM chemodiversity was mainly affected by the bacterial community and ~~their~~ its metabolic by-products. Logue et al. (2016) also ~~referred~~mentioned that more diverse microbial communities could decompose complex organic matter easier and provide with a ~~varying~~ larger variation ~~type~~ of DOM.

In the distance-decay analysis, the geographical location represents the different ~~circumstances~~conditions of the DOM samples, thus, ~~the~~ geographical~~al~~ location can

dictates the variation of DOM chemodiversity. While In the VPA analysis, the geographical location only refers to the latitude and longitude excluding both environmental and biotic factors, and does unable not to significantly affect the DOM chemodiversity. Combining these two analyses, the results indicated that geographical location could directly affect the bacterial community, and thus, indirectly affect DOM chemodiversity. In the molecule-central networks, OTUs are assumed to be "specialists" when related to limited DOM compounds (Zhao et al., 2019); while OTUs with broader and multiple connections with compounds in the network are considered as "generalists" (can produce and/or consume multiple DOM molecules in positive or negative networks). Our study showed that the positive networks were far more than the negative ones, and the OTU-central nodes were more than the DOM molecule-central ones (among those with more connections in the positive networks (Fig. 5A)). This indicates that at paddy field environments, especially after harvesting, the contribution of bacteria to DOM is greater than their DOM consuming habits. That is also consistent with the VPA observation, where bacteria contributed more to DOM than vice versa.

Our previous study has shown that the spatial variation in the bacterial community was the greatest under CK treatment (Chen et al., 2017); in this studystudy, we evaluated a stronger bacteria−DOM interaction in CK. Thus, the spatial variation of DOM chemodiversity in CK treatment is the greatest in the current study. The spatial variation of the bacterial community under the NPK treatment was greater than that for OM treatment (Chen et al., 2017), and the bacteria−DOM interaction was also stronger in NPK than for OM. Thus, the spatial variation of DOM chemodiversity in NPK

treatment ~~was~~ could be also greater than ~~that for OM treatment~~the ON one. This ~~concludes~~ highlights that the minimum spatial variation ~~of~~ in DOM chemodiversity was ~~at~~ observed at OM treatment. More specifically, ~~anthropogenic~~ organic fertilization decreases the variation of the dissolved organic matter chemodiversity in paddy soils by homogenizing the bacterial community composition.

5. Conclusions

This study shows that the ~~paddy soil~~ DOM chemodiversity from paddy soil exhibits both stability and variation in different aspects. The stability refers to the presence of similar molecules ~~between~~ among samples from different sites (DOM samples). The differences/variation in DOM chemodiversity is mainly attributed to geographic distance at a regional scale, whilst anthropogenic organic fertilizations could significantly homogenize the DOM chemodiversity by providing with alike LP compounds. DOM chemodiversity and bacteria community were significantly correlated, geographic distance ~~,~~ indirectly affected the DOM chemodiversity by shaping the bacterial community. The homogenization of the DOM chemodiversity was dictated by the homogenization of the bacterial community composition, ~~that was~~ the latter (community composition) was directly affected by long-term anthropogenic organic fertilization.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number 41430859), National Key Research and Development Program of China (grant number 2016YFD0200309), Natural Science Foundation of Jiangsu Province (grant

464 number BK20181510).

465 Declaration of Competing Interest

466 The authors declare that they have no conflict of interest.

467 Appendix A. Supplementary material

468 The Supporting Information including additional tables and figures discussed in the
469 main text is available at…

470 References

471 Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of
472 variance. Austral. Ecol. 26, 32–46.
473 Carter, M.R., 2007. Gregorich, E. G. Soil Sampling and Methods of Analysis, 2nd ed.;
474 CRC Press.
475 Chen, R.R., Zhong, L.H., Jing, Z.W., Guo, Z.Y., Li, Z.P., Lin, X.G., Feng, Y.Z., 2017.
476 Fertilization decreases compositional variation of paddy bacterial community
477 across geographical gradient. Soil Biol. Biochem. 114, 181–188.
478 Deng, Y., He, Z.L., Xiong, J.B., Yu, H., Xu, M.Y., Hobbie, S.E., Reich, P.B., Schadt,
479 C.W., Kent, A., Pendall, E., 2016. Elevated carbon dioxide accelerates the spatial
480 turnover of soil microbial communities. GCB Bioenergy 22, 957–964.
481 Ding, Y., Shi, Z.Q., Ye, Q.T., Liang, Y.Z., Liu, M.Q., Dang, Z., Wang, Y.J., Liu, C.X.,
482 2020. Chemodiversity of soil dissolved organic matter. Environ. Sci. Technol. 54,
483 6174–6184.
484 Kallenbach, C.M., Frey, S.D., Grandy, A.S., 2016. Direct evidence for microbial-
485 derived soil organic matter formation and its ecophysiological controls. Nat.

Commun. 7, 13630.

Kellerman, A.M., Dittmar, T., Kothawala, D.N., Tranvik, L.J. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. Nat. Commun. 5, 3804.

Khan, S., Chao, C., Waqas, M., Arp, H.P., Zhu, Y.G., 2013. Sewage sludge biochar influence upon rice (Oryza sativa L) yield, metal bioaccumulation and greenhouse gas emissions from acidic paddy soil. Environ. Sci. Technol. 47, 8624−8632.

Koch B.P., Dittmar, T., Witt, M., Kattner, G., 2007. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. Anal. Chem. 79, 1758–1763.

Kögel-Knabner, I., Amelung, W., Cao, Z.H., Fiedler, S., Frenzel, P., Jahn, R., Kalbitz, K., Kölbl, A., Schloter, M. Biogeochemistry of paddy soils. Geoderma 157, 1–14.

Kujawinski, E.B., Longnecker, K., Barott, K.L., Weber, R.J.M., 2016. Soule MCK. Microbial community structure affects marine dissolved organic matter composition. Front. Mar Sci. 3, 45.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Lehmann, J., Kleber M., 2015. The contentious nature of soil organic matter. Nature 528, 60–68.

Li, H.Y., Wang, H., Wang, H.T., Xin, P.Y., Xu, X.H., Ma, Y., Liu, W.P., Teng, C.Y., Jiang, C.L., Lou, L.P., Arnold, W., Cralle, L., Zhu, Y.G., Chu, J.F., Gilbert, J.A., Zhang, Z.J., 2018a. The chemodiversity of paddy soil dissolved organic matter correlates

508 with microbial community at continental scales. Microbiome 6, 187.
509 Li, X.M., Chen, Q.L., He, C., Shi, Q., Chen, S.C., Reid, B.J., Zhu, Y.G., Sun, G.X.,
510 2019. Organic Carbon Amendments Affect the Chemodiversity of Soil Dissolved
511 Organic Matter and Its Associations with Soil Microbial Communities. Environ.
512 Sci. Technol. 53, 50–59.
513 Li, X.M., Sun, G.X., Chen, S.C., Fang, Z., Yuan, H.Y., Shi, Q., Zhu, Y.G., 2018b.
514 Molecular Chemodiversity of Dissolved Organic Matter in Paddy Soils. Environ.
515 Sci. Technol. 52, 963–971.
516 Logue, J.B., Stedmon, C.A., Kellerman, A.M., Nielsen, N.J., Andersson, A.F., Laudon,
517 H., Lindström, E.S., Kritzberg, E.S., 2016. Experimental insights into the
518 importance of aquatic bacterial community composition to the degradation of
519 dissolved organic matter. ISME J. 10, 533–545.
520 Lu, R.K., 1999. Analytical methods of soil and agricultural chemistry. China
521 Agricultural Science and Technology Press, Beijing.
522 Nebbioso, A., Piccolo, A., 2013. Molecular characterization of dissolved organic matter
523 (DOM): a critical review. Anal. Bioanal. Chem. 405, 109–124.
524 Ohno, T., Sleighter, R.L., Hatcher, P.G., 2016. Comparative study of organic matter
525 chemical characterization using negative and positive mode electrospray
526 ionization ultrahigh-resolution mass spectrometry. Anal. Bioanal. Chem. 408,
527 2497.
528 Palmer, M.W., Maurer, T.A., 1997. Does diversity beget diversity? A case study of crops
529 and weeds. J. Veg. Sci. 8, 235–240.

Qiu, C.P., Feng, Y.Z., Wu, M., Liu, M., Li, W.T., Li, Z.P., 2019. NanoFe3O4 accelerates methanogenic straw degradation by improving energy metabolism. Bioresource Technol. 292, 121930.

Roth, V.N., Lange, M., Simon, C., Hertkorn, N., Bucher, S., Goodall, T., Griffiths, R.I., Mellado-Vázquez, P.G., Mommer, L., Oram, N.J., Weigelt, A., Dittmar, T., Gleixner, G., 2019. Persistence of dissolved organic matter explained by molecular changes during its passage through soil. Nat. Geosci. 12, 755–761.

Stegen, J,C., Fredrickson, J.K., Wilkins, M.J., Konopka, A.E., Nelson, W.C., Arntzen, E.V., Chrisler, W.B., Chu, R.K., Danczak, R.E., Fansler, S.J., Kennedy, D.W., Resch, C.T., Tfaily, M., 2016. Groundwater-surface water mixing shifts ecological assembly processes and stimulates organic carbon turnover. Nat. Commun. 7, 11237.

Tanentzap, A.J., Fitch, A., Orland, C., Emilson, E.J.S., Yakimovich, K.M., Osterholz, H., Dittmar, T., 2019. Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. P. NATL. ACAD. SCI. USA 116, 24689–24695.

Testa, B., Vistoli, G., Pedretti, A., Bojarski, A.J., 2009. Atomic diversity, molecular diversity, and chemical diversity: the concept of chemodiversity. Chem. Biodivers. 6, 1145–1151.

W, Q.Y., Zhou, T.H., Du, Y., Ye, B., Wang, W.L., Hu, H.Y., 2020. Characterizing the molecular weight distribution of dissolved organic matter by measuring the contents of electron-donating moieties, UV absorbance, and fluorescence intensity. Environ. Int. 137, 105570.

552 Zhao, Z., Gonsior, M., Schmitt-Kopplin, P., Zhan, Y., Zhang, R., Jiao, N., Chen, F.,
553 2019. Microbial transformation of virus-induced dissolved organic matter from
554 picocyanobacteria: coupling of bacterial diversity and DOM chemodiversity.
555 ISME J. 13, 2551–2565.
556 Zherebker, A., Shirshin, E., Rubekina, A., Kharybin, O., Kononikhin, A., Kulikova, N.
557 A., Zaitsev, K.V., Roznyatovsky, V.A., Grishin, Y.K., Perminova, I.V., Nikolaev,
558 E.N., 2020. Optical properties of soil dissolved organic matter are related to acidic
559 functions of its components as revealed by fractionation, selective
560 deuteromethylation, and ultrahigh resolution mass spectrometry. Environ. Sci.
561 Technol. 54, 2667–2677.
562 Zou, L., Wang, X.C., Callahan, J., Culp, R.A., Chen, R.F., Altabet, M.A., Sun, M.Y.,
563 2004. Bacterial roles in the formation of high-molecular-weight dissolved organic
564 matter in estuarine and coastal waters: Evidence from lipids and the compound
565 specific isotopic ratios. Limnol. Oceanogr. 49, 297–302.
566
567

568 **Table 1** PERMANOVA test based on the DOM chemodiversity among three fertilization practices
569 and among four experimental sites.

	Global Test		Pairwise Comparison		
	F	*P*		*F*	*P*
Fertilization treatment	1.791	0.007	CK *vs* NPK	0.689	0.699
			CK *vs* OM	3.149	0.007
			NPK *vs* OM	2.000	0.041
Experimental site	3.275	0.001	WC *vs* YT	4.486	0.001
			WC *vs* CS	4.847	0.006
			WC *vs* JX	1.371	0.082
			YT *vs* CS	8.268	0.001
			YT *vs* JX	3.794	0.001
			CS *vs* JX	5.113	0.002

570 The global test was conducted with two-way PERMANOVA and the pairwise comparison was
571 conducted with one-way PERMANOVA.
572

Figure captions

Fig. 1. van Krevelen diagrams from the FT-ICR-MS spectra of DOM. Dotted boxes overlain on the plot refer to the major biomolecular compound groups: carbohydrates, condensed aromatic, lignins, lipids, proteins/amino sugars, tannins, and others (the compounds not belonging to the above 6 groups). Darker color indicates a higher molecular density, which means higher occurrence frequency of molecular formulas in specific area of the diagram.

Fig. 2. Venn analysis of the molecular formulas among the three treatments CK, NPK, and OM (**A**), between CK and NPK (**B**), CK and OM (**C**), between NPK and OM (**D**) in four sites. **E-G** are the van Krevelen diagrams of unique compounds in CK, NPK, and OM, respectively. Density indicates the occurrence frequency of molecular formulas in specific area of the diagrams.

Fig. 3. Distance-decay curves of DOM chemodiversity based on all compounds (**A**) and seven biomolecular compound groups (**B-H**, carbohydrates, condensed aromatics, lignins, lipids, protein/amino sugar, tannins, and others, respectively) under three fertilization practices within and among four experimental sites. The y-axis is DOM composition dissimilarity calculated using Bray-Curtis distance of molecular formulas in all compounds and seven compound groups. X-axis is geographic distance in kilometer. Different colours refer to different treatments, Blue: CK, Green: NPK, Purple: OM, Red: including all the three treatments.

Fig. 4. Variable partitioning analysis (VPA) was applied to analyze the effects of environmental factors (E), geographical location (G) and bacterial community (B) and their interactions on the variance of DOM chemodiversity (a) and the effects of (E), (G), and DOM chemodiversity (D) on the variance of bacterial community (b) based all the samples.

Fig. 5. Network plots showing the association between DOM formulaes and bacterial species in all

595 the samples. Compound groups and bacterial species are shown in circle nodes and square nodes,
596 repectively. Different classes are labeled with different colors. Direct connections are made between
597 DOM formulae and OTUs when the Spearmen's rank correlation coefficient is strong $r \geq 0.8$
598 (Positive network, **A-D**) and $r \leq -0.8$ (Negative network, **E**), with $P < 0.05$. Size of the nodes are
599 proportional to the number of connections.
600